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

(51) International Patent Classification⁷: G06F 17/20, G11B 27/10

(21) International Application Number:
PCT/AU2004/001797

(22) International Filing Date:
22 December 2004 (22.12.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2003907124   24 December 2003 (24.12.2003)   AU

(71) Applicant and
(72) Inventor: PAYNE, Rodney [AU/AU]; c/o 79 Warrimoo Ave St Ives, NSW 2075 (AU).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (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: RECORDING AND TRANSCRIPTION SYSTEM
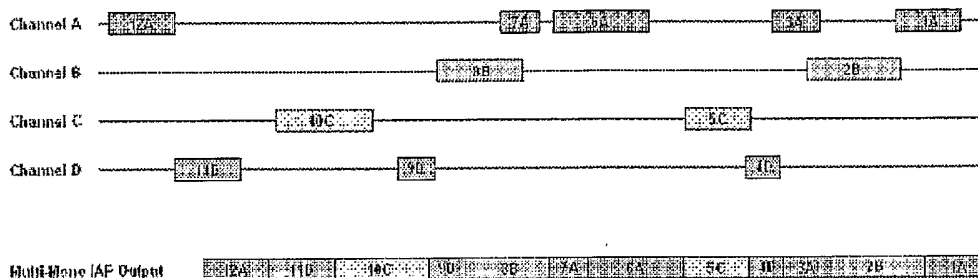
(57) Abstract: A recording and transcription System comprising a recording device adapted to simultaneously receive and record Input from multiple audio recording devices Bach originating from a separate channel; the recording device merging information from the multiple audio input channels onto an intelligent audio File or database playable as a single channel with the original separate channels replayable by a playback device substantially in chronological order and in chonologically indexed non overlapping sequence; one or more input controls operable at the time of recording to facilitate the embedding and linking of information and/or markers unique to each speaker onto recorded material from each channel or discrete sections of recorded material from each channel; the playback device being adapted to utilise the embedded information markers for linking text typed from the intelligent audio file or database to both the relevant section of the audio file or database and the imbedded information therein including speaker identification such that a typist listening to the playback device need not know, comprehend or type speaker identification or other imbedded information but that such information may automatically appear in the finished text file in a predetermined format.

RECORDING AND TRANSCRIPTION SYSTEM

Field of the Invention

The present invention relates to a system for transcribing the spoken word and in particular, to a system which enables a multi-channel audio file such as may result from a court hearing to be efficiently transcribed by a typist, or group of typists without the need for the typist to manually determine which of a multiplicity of speakers relate to any particular portion of the audio file being transcribed at any given time.

Background of the Invention

For many years multi-channel analogue and digital recording and transcription systems have used parallel processes whereby multiple microphones have been used to record multiple channels of audio information at the same time. These systems are for example, common in courtrooms in order that a court transcript may be prepared for transcription. There may, for example, be a microphone for the judge, a separate microphone for the witness stand and a microphone for each of Defence and Prosecution counsel.

When an audio file compiled as abovementioned reaches a typist, the typist must transcribe the recording to produce a document with a text layout, format and syntax appropriate to the particular court concerned and, of course, each line of text in chronological order must be preceded, at the very least, by the name of the person who originally spoke the text. The lastmentioned information is usually obtained by

- 2 -

the typist from a log to which she must have constant reference whilst typing and which has been meticulously prepared by a person in the courtroom. Identifying the person who spoke the text being typed is not always easy, even with a log, as speakers at a particular microphone may change frequently such as when one witness finishes in the witness box and another commences. Furthermore, difficulties often occur when two persons and different microphones speak at the same time. The typist must manually enable and disable the different channels on the audio file which correspond to the separate microphones which originally record the information in an attempt to try to unscramble audio information which may become garbled if two parties speak at the same time during the original recording.

Transcription of such audio files is consequently quite labour intensive and, even transcription of a single channel audio file, may take 3 to 4 times the time taken to record the file whereas multi-channel transcription can often take between 10 and 14 times the length of the original audio file to transcribe. Not only does this result in a time consuming and expensive transcript, but it often precludes a reliable transcript being returned to a court in time for persons involved in a trial process to refer to a reliable textual transcript whilst a trial is in progress or immediately thereafter.

The typist is also often required to keep track of the audio file time line in order to insert chronological information into the transcript.

It is also often difficult to remunerate typists for such work on a fair basis due to the fact that audio files differ greatly in complexity and the number of words typed does

- 3 -

not give a fair indication of the amount of work and time which has gone into the producing the transcript.

It is consequently an object of the present invention to ameliorate one or more of the above disadvantages associated with existing recording and transcription systems or at least, to provide the market with an alternative.

Summary of the invention

According to the present invention there is disclosed a recording system to facilitate transcription of audio input from an event comprising two or more audio input devices inputting to one or more recording devices; an audio packet creation device associated with the recording device adapted to separate discreet packets of audio input from each input device and assign to each packet an identifier unique to the input device through which the packet was created and a code indicating time of input contained within each packet relative to time of other packets created from the same event:

Preferred Mode of Carrying out the Invention

According to one embodiment of the present invention there is disclosed a recording and transcription system comprising digital recording means adapted to simultaneously receive and record input from multiple audio recording devices each originating from a separate channel. Software is provided to merge information from the multiple audio input channels into an Multi-Mono Intelligent Audio File (MMIAF) stream. The Multi-Mono Intelligent Audio File (MMIAF) stream is

- 4 -

playable with each original audio channel voice packet isolated and re-playable in

chronological order with no-overlapping sections, as illustrated in Figure 1.

Microphones interfaced with appropriate software in the recording device are

provided which are operable at the time of recording to facilitate the embedding and

linking of information and/or markers unique to each speaker into recorded material

from each channel or discrete sections of recorded material from each channel as

illustrated in Figure 2.

Software is provided to facilitate utilising the embedded information markers for

linking text typed from the Intelligent Audio File or database to both the relevant

section of the audio file or database and the imbedded information therein including

speaker identification such that the typist need not know, comprehend or type speaker

identification or other imbedded information, but that such information may

automatically appear in the finished text file in a predetermined format.

According to the abovementioned embodiment of the present invention each person

who approaches a microphone or other input device for an intended recording would

wear or carry a magnetically encoded card which, when brought into the proximity of

a microphone with an appropriate card reader, would transmit the identity of the

speaker adjacent to the microphone to a device associated with the recording in order

- 5 -

that audio signals recorded after the card reader sensed the proximity of a card would

be attributed to the person to whom the card had been assigned. In a similar manner,

the system may automatically be advised of the termination of a particular speaker's

input when the system sensed the absence of that speaker's proximity card. This

lastmentioned system in conjunction with a real-time clock associated with the

recording could obviate the necessity for a person in a court to keep a log.

Alternatively, speaker changes on any channel can be manually applied by a court

monitor.

The microphones, being the input devices for the recording to be transcribed, would

also be associated with an audio sensing device such that during periods of silence no

audio is sent to the typist and hence a typist would not be required to waste time

listening to periods of silence. This feature is incorporated in the software of the

recording device itself so that periods of silence are extracted after the initial

recording has been made.

As each speaker inputs audio material into a given microphone the audio material

is recorded in an audio file on a computer. As each audio packet is detected and

extracted the header information listed below is added, to allow file management.

Speaker ID (Speaker Name)	Channel ID
Custom Dictionary	Sample Rate & Compression
Speaker Language (English, French etc)	Transcription Priority
Start time / date of audio packet	Duration of packet

- 6 -

End time / date of audio packet   Chronological Key

Court Location & Court Room No

Return IP address for completed transcript

As audio is recorded into each channel, audio (when detected) is extracted as separate Intelligent Audio File packets which can be sent individually, sent as a multi-mono stream or accumulated into predefined payloads with all audio packets in chronological order so there are no overlapping speakers in the final mono audio file created. Essentially, we have programmatically converted Complex audio (2 or more channels) into a Simple audio (mono) file in real-time. As a result, this audio file may then be used by a typist to generate a final transcript within a dedicated multi-mono text editor or with MS Word via VBA integration between the multi-mono audio stream and MS Word. The resultant multi-mono audio file contains embedded information concerning the identity of the speaker, the start and stop time of the speaker and the matter concerned, all of which is not necessarily displayed to the typist. This information is then used to automatically format the document which the typist is generating according to the preferred template and rules required to be applied by a particular court. The software, via which the typist generates text whilst transcribing the audio, may be designed to provide the typist with the inability to apply any form of text formatting to the document other than plain and formatted text.

All formatting is advantageously automatically applied to the text once the typist has completed the transcript. In doing so significant transcription time is saved whilst at the same time ensuring perfect formatting consistency, regardless of how

- 7 -

many typists do the work on the same transcript. Since the text is fully indexed to the audio a fully formatted transcript can be instantly created utilising a software function similar to that of "mail merge" whereby the text within the database for the indexed audio is automatically formatted based on a particular court template. Speaker names (which are not transcribed by the typist) are also added to the final transcript based upon information embedded in the original Intelligent Audio File.

It will be appreciated that utilising apparatus and a system in accordance with the present invention as above described, will result in a very simple task for a typist who no longer has to enable or disable various channels of a recording in order to hear a recording when two or more speakers speak at once in order to keep track of and determine who is speaking at any particular time, as illustrated in Figure 3.

The typist need furthermore not know the required format in which the transcript must be presented according to the rules of a particular court or tribunal. There is also no requirement for the typist to constantly have reference to a log.

Another feature of an Multi-Mono Intelligent Audio File such as can be generated in accordance with the present invention is that software may link the audio file to the text being typed. This assists proofreading of the final text document as a person reading text in a document may immediately recall the portion of the audio file which was used to generate that portion of the text. This feature may also be

- 8 -

very useful to counsel arguing a case during, for example, cross-examination as when the text transcript is returned to the court (which may now occur very quickly due to the efficiencies of the present invention) and a person queries whether they said what is actually typed in the transcript, the audio file can be replayed immediately by performing a search function on the relevant word or words of the transcript.

As the embedded information referable to each segment of input from each channel effectively indexes the relationship of that segment to the final document which is intended to be generated it is possible to store each segment separately in a database and then to readily reassemble them when required. This is the case whether the segment is in the form of a mere audio file or whether it is in the form of an audio file which has been transcribed and consequently has associated text. This feature enables the use of multiple typists, if required, in order to compile a single transcript quickly.

From the typists perspective, a typist can readily see how many words they have typed and they know that they have not put a lot of time and effort into formatting the document in accordance with a particular template and keeping track of a log. Typists may therefore be remunerated in accordance with number of words which they have typed which is more likely to be perceived as fair by both the typist and their employer than an hourly rate in a situation where transcription has been complex compared to the amount of words in the final typewritten document.

- 9 -

The present invention has the additional advantage that when the file generated by the audio input is transmitted electronically the bandwidth required is kept to a minimum as multiple channels need not be sent in parallel but rather transmission is preferably by way of a single mono stream of Intelligent Audio Packets file separated by speaker identification markers. Furthermore, periods of silence may be deleted prior to transmission.

The Intelligent Audio Files generated by the present invention are well adapted for encryption should security be required.

Additionally, the invention is capable of adaptation to situations where re-voicing is utilised so as to enable the integration of voice recognition capabilities rather than just manual transcription.

The automated feature of the software used by the typist may also be used in order to add unique words which are specific to a particular matter as well as unusual names and addresses specific to a particular court case. These words and information are added prior to transcription and transmitted with the audio file to the typist. If new words are introduced these can automatically be detected from the resulting transcripts and added to a custom dictionary contained in the audio file during the court case. In this respect, the invention can be said to be self learning.

- 10 -

Rather than referring to text via page and paragraph which can cause discrepancies due to the difference in the manner in which some printers operate (resulting in a different number of lines per page), it is possible, in accordance with the present invention, to reference all text via unique paragraph numbers which are added by the program rather than the typist and consequently are independent of printer drivers and font type or size. This has not been previously practical because typists are currently needed to manually enter paragraph numbers. If multiple typists were to work on the same transcript in accordance with prior art systems it is almost impossible, or at the very least, very time consuming, to synchronise paragraphs in the final document.

Speakers using different or varying languages can easily be transcribed, as speaker language is individually assignable for each channel, hence resulting audio packets sent for transcription have an internal language file set within them, ensuring only typists with the corresponding language skills will type them.

Traditionally, using a parallel method of transcription (tape or digital) if a section of audio is inaudible, all work from this point onwards is delayed. Using the Multi-Mono methodology, if a specific packet of audio is received that is beyond the skill of the typist (for whatever reason), they can simply abort this packet, whereby this packet can be automatically forwarded onto a supervisor who will hopefully, have greater skill and experience to transcribe what was said within this packet. Thus, transcription past this point can continue without delay, while the packet causing issues can be resolved in parallel by a supervisor.

It will be appreciated that variations may be made to the invention as above described without departing from the scope and intendment thereof and for example the proximity card readers contemplated above (in order that each new witness may be identified to the system by way of a proximity card worn by the witness in conjunction with a card reader associated with each microphone) may be substituted with other types of card readers or even a manual system for generating speaker information which is to be transmitted as part of the audio file to the software utilised by the typist.

- 12 -

The claims defining the invention are as follows:

1   A recording system to facilitate transcription of audio input from an event comprising two or more audio input devices inputting to one or more recording devices; an audio packet creation device associated with the recording device adapted to separate discreet packets of audio input from each input device and assign to each packet an identifier unique to the input device through which the packet was created and a code indicating time of input contained within each packet relative to time of other packets created from the same event.

2   A recording system in accordance with claim 1 hereof wherein each packet is also indexed with the name of the speaker creating the audio input contained within the packet.

3   A recording device in accordance with any one of the preceding claims wherein the input devices are microphones provided with sensors adapted to read transmissions from or magnetically active ID tags worn by persons who may input to the microphone when the ID tags come within a predetermined distance from the microphone in order that input to the microphone may be identified with a particular ID tag wearer.

4   A recording system in accordance with any one of the preceding claims wherein the time code includes the start time of the audio input in the packet.

5   A recording system in accordance with any one of the preceding claims wherein the time code includes the start and end time of the audio input in the packet.

6   A recording and transcription system in accordance with any one of the preceding claims including a playback and transcription device adapted to facilitate creation

of written text from a packet and to automatically assign a speaker or channel as well as a time to the written text.

7   A recording and transcription system in accordance with any one the preceding claims wherein the identifiers and time codes are visibly or invisibly maintained in the written text to facilitate cross referencing the written text to the audio file and further to facilitate assembly of separately transcribed packets into a coherent chronological transcript of an entire event.
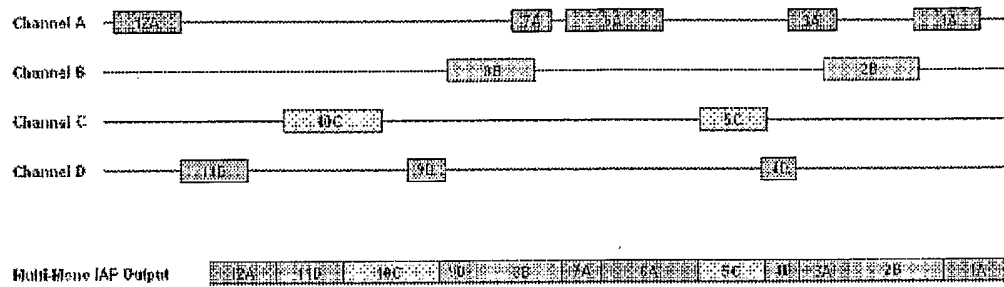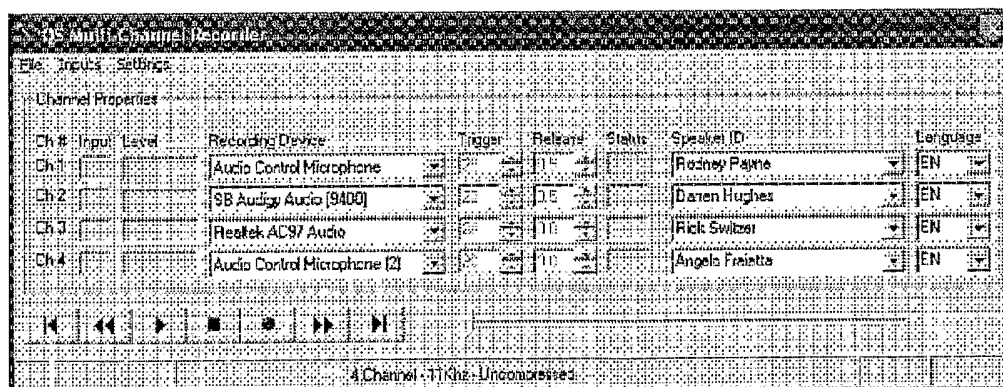
FIGURE 1

FIGURE 2

FIGURE 3

A.   CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]:   G06F 17/20, G11B 27/10

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)
DWPI, IEEE (transcribe, transcription, stenography, court, legal, judicial, hearing, proceeding, conference, etc.)

C.   DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/0169330 A1 (BEN-SHACHER et al), 11 September 2003 the whole document	1-7
X	US 6,477,491 B1 (CHANDLER et al), 5 November 2002 the whole document	1-7
A	EP 1 079 313 A2 (DIGITAKE SOFTWARE SYSTEMS LIMITED), 28 February 2001 the whole document	1-7
A	US 2002/0133513 A1 (TOWNSEND et al), 19 September 2002 the whole document	1-7

| 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
2 March 2005	1 4 MAR 2005

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	MATTHEW HOLLINGWORTH Telephone No : (02) 6283 2024

C (Continuation).   DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002/0049595 A1 (BENNETT et al), 25 April 2002 the whole document	1-7
A, P	C. Schremmer et al, *Audio Meta Data Transcription from Meeting Transcripts for the Continuous Media Web* Proc. 25th Audio Engineering Society Conference on Metadata for Audio, pp. 288-95, June 2004	1-7

This Annex lists the known "A" publication level 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 Cited in Search Report		Patent Family Member					
US	2003169330	US	6674459				
US	6477491	NONE					
EP	1079313	IE	990799	IE	990800		
US	2002133513	CA	2441237	EP	1370960	WO	02075566
US	2002049595	US	5369704	US	5444615	US	5740245
		US	5815392	US	5815639	US	5878186
		US	5884256	US	5926787	US	5940800
		US	5949952	US	5950194	US	5970141
		US	6023675	US	6026395	US	6055531
		US	6282510	US	2002049588	WO	9422246

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

END OF ANNEX