

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
1 July 2004 (01.07.2004)

PCT

(10) International Publication Number
WO 2004/056086 A2

(51) International Patent Classification⁷: **H04N 5/00**

(21) International Application Number:
PCT/IB2003/005912

(22) International Filing Date:
12 December 2003 (12.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/433,722 16 December 2002 (16.12.2002) US

(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): **GUTTA, Srinivas** [US/US]; P.O. Box 3001, Briarcliff Manor, NY 10510-8001 (US).

(74) Common Representative: **KONINKLIJKE PHILIPS ELECTRONICS N.V.**; Intellectual Property & Standards, c/o BELK, Michael E., P.O. Box 3001, Briarcliff Manor, NY 10510-8001 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, 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, 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 (*regional*): ARIPO patent (BW, 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).

Declaration under Rule 4.17:

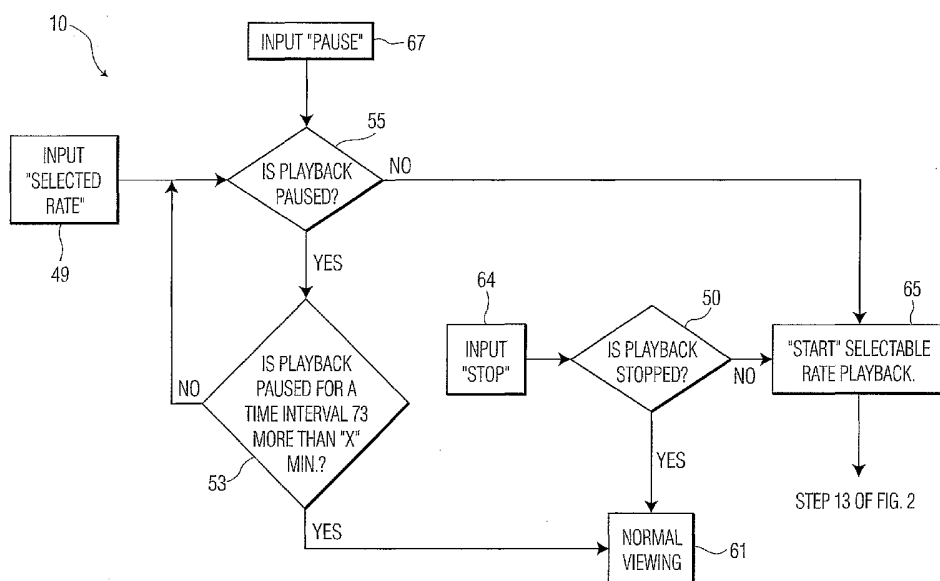
— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for all designations

Published:

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

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

(54) Title: METHOD AND APPARATUS FOR SELECTABLE RATE PLAYBACK WITHOUT SPEECH DISTORTION



(57) Abstract: A method and an apparatus for selectable rate playback of a selected first portion of a separately stored synchronized video and audio content, without distortion of speech due to the selectable rate playback of the playback content and without loss of synchronization of the selected first portion of a separately stored synchronized video and audio content.

WO 2004/056086 A2

METHOD AND APPARATUS FOR SELECTABLE
RATE PLAYBACK WITHOUT SPEECH DISTORTION

5 The present invention relates generally to the field of television. More specifically,
the present invention relates to an apparatus and method for selectable rate playback of
television programs without distorting the audio portion of the programs.

Related Art

Selectable rate playback of the video content from various storage mediums such as
video cassette recorders (VCR) is known. An audio portion of the playback content may
10 be suppressed during selectable rate playback, to avoid distortion of the audio portion.
There is a need for undistorted presentation of the audio portion of the playback content
during selectable rate playback. Hereinafter, "distortion" of the audio portion of the
playback content means a lack of fidelity in reception or reproduction due to a change in a
rate of playback of the audio portion of the playback content compared to the rate of
15 storing the audio portion of the playback content.

The present invention provides a method for playback of playback content at
selectable rates, comprising:

selecting a first portion of separately stored video and audio playback content,
wherein the playback content has been stored at a storing rate, wherein the video and audio
20 are synchronized as stored, and wherein the separately stored synchronized video and
audio content are retrievable for synchronized playback;

selecting a rate of playback of the playback content from the selectable rate,
wherein the selected playback rate is different from the storing rate;

25 tagging speech in the selected first portion of the playback content;
recognizing an at least one phrase in the tagged speech; and

playing said first portion of playback content at said rate of playback, wherein said
playing synchronously retrieves the tagged speech, and wherein playing at said rate does
not result in distortion of speech in the playback content even though said rate is different
than the storing rate, and wherein the video and audio are synchronized at said rate of
30 playback during said playing.

A second embodiment of the present invention discloses an apparatus for selectable
rate playback of playback content, comprising:

a separately stored video and audio playback content, wherein the playback content has been stored at a storing rate;

a selected first portion of the separately stored video and audio playback content in a storage medium, wherein the selected first portion of video and audio content are

5 synchronized and a speech portion of the audio content is tagged;

 a speech recognition device for tagging the speech portion of the audio content;

 a phrase recognition device for determining valid words for phrases from the tagged speech, wherein the valid words are joined into said phrases;

10 a playback device for playback of the selected first portion of the playback content at a rate selected from the selectable rate, wherein the selected rate is different from the storing rate, wherein playback at the selected rate synchronously retrieves the Tagged Speech portion of the audio content, wherein playback at the selected rate does not result in distortion of speech in the playback content even though the selected rate is different from the storing rate, and wherein the video and audio content are synchronized at the selected rate during

15 said playback.

The present invention advantageously provides undistorted presentation of the audio portion of the playback content during selectable rate playback.

FIG. 1 depicts a functionality and logic of an apparatus for starting selectable rate playback of playback content or normal viewing, in accordance with embodiments of the

20 present invention;

FIG. 2 depicts a functionality and logic of an apparatus for selectable rate playback of playback content, in accordance with embodiments of the present invention;

FIG. 3 depicts a playback list, for selecting a first portion of Separately Stored Synchronized Video and Audio Playback Content;

25 FIG. 4 depicts a graphical user interface (GUI), for selecting a first portion of Separately Stored Synchronized Video and Audio Playback Content; and

FIG. 5 depicts a method for selectable rate playback of playback content, in accordance with embodiments of the present invention.

Although certain preferred embodiments of the present invention will be shown and

30 described in detail, it should be understood that various changes and modifications may be made without departing from the scope of the appended claims. The scope of the present invention will in no way be limited to the number of constituting components, the

materials thereof, the shapes thereof, the relative arrangement thereof, etc., and are disclosed simply as an example of the preferred embodiment. The features and advantages of the present invention are illustrated in detail in the accompanying drawings, wherein like reference numerals refer to like elements throughout the drawings. Although the drawings are intended to illustrate the present invention, the drawings are not necessarily drawn to scale.

The present invention relates generally to the field of television. More specifically, the present invention relates to an apparatus and method for selectable rate playback of a selected video and audio playback content, without distortion of the speech due to the selectable rate playback of the playback content.

FIG. 1 is a flowchart illustrating a functionality and a logic description of an apparatus 10 for Selectable Rate Playback of Playback Content, in accordance with embodiments of the present invention and in accordance with a method for selectable rate playback of playback content, as depicted by a flow chart 70 in FIG. 5 and described herein. FIG. 1 illustrates that a user may cause a "start" of selectable rate playback in step 65 or a continuing of normal viewing 61, such as viewing independent of the apparatus 10. "Starting" Selectable Rate Playback 65 of Playback Content depends on three inputs: a "Stop" Selectable Rate Playback 64 input; and a "Pause" Selectable Rate Playback 67 input; and a "Selected Rate" 49 input. A user may choose to provide inputs 64, 67, and 49 from a programmable logic controller (PLC), or alternatively from a central processing unit (CPU), equipped with appropriate software.

In one embodiment, a user may start selectable rate playback in step 65 by providing a "Selected Rate" 49 input, if decision step 55 determines that playback has not been paused and decision step 50 determines that playback has not been stopped. The "Selected Rate" 49 input may be a slower rate or a faster rate of playback than was used to store the playback content. In one embodiment, the "Selected Rate" 49 was a range from about 50% to about 150% of the rate used to store the playback content or for any other reason. However, a user may select any appropriate "Selected Rate" 49 that results in a playback of the Selected Separately Stored Synchronized Video and Audio Playback Content 1 that is more clear or understandable to a viewer or listener of the playback content. Hereinafter, "selectable speed" or "selectable rate" means increasing or decreasing a speed or rate of playback of the Selected Separately Stored Synchronized

Video and Audio Playback Content 1, compared to the speed or rate of storing the Selected Separately Stored Synchronized Video and Audio Playback Content 1 without causing distortion of speech in the playback content, as depicted in FIG. 2 and described infra. Playback may be paused by providing a "pause" input 67 to the decision step 55. Playback may be stopped by providing a "stop" input 64 to the decision step 50. When playback is viewed on an audio and video device, such as, for example, a television, and playback is paused by providing the "pause" input 67 for greater than "x" minutes or when playback is stopped by providing a "stop" input 64, Normal Viewing 61 on the audio and video device may result. Hereinafter, Normal Viewing 61 means, for example television operation or operation of any appropriate audio and video viewing device independent of the selectable rate playback apparatus or method of the present invention. When playback is paused for greater than "x" minutes, the "pause" input 67 is provided to decision step 53, resulting in normal viewing 61. Alternatively, if playback is not paused for greater than "x" minutes, the "pause" input 67 loops back to decision step 55, and then again to decision step 53 until the "pause" input 67 is removed. When the pause input 67 is removed, the apparatus goes to the "start" Selectable Rate Playback step 65. In one embodiment, "x" is less than two (2) minutes. Alternatively, "x" may be a time interval less than five (5) minutes. The value of "x" may be any positive real number that represents a number of minutes a user desires to wait for automatic return to the normal viewing 61 step after the "Pause" input 67 has been provided to the apparatus 10.

Normal Viewing 61 may result, in one embodiment, the decision step 50 determines whether the "stop" input 64 has been provided. If yes, Normal Viewing 61 results. Alternatively, if the "Stop" 64 input has not been passed to the decision block 50, the apparatus 10 moves to "start" selectable rate playback step 65.

FIG. 2 depicts an extension of the apparatus 10 of FIG. 1, after adding: a Selecting and Tagging Portion 9; a Phrase and Tokens Recognizing portion 2; and a Selectable Rate Playback portion 4, in accordance with embodiments of the present invention including in accordance with a method for selectable rate playback of playback content, as depicted by a flow chart 70 in FIG. 5 and described infra.

The Selecting and Tagging Portion 9 includes: a Selecting Engine 13, wherein the "Start" Selectable Rate Playback 65 of FIG. 1 has been provided to the Selecting Engine 13, in accordance with embodiments of the present invention including in accordance with

steps 75 and 90 of flow chart 70 in FIG. 5 and described infra. In addition to receiving the “Start” 65 input, the Selecting Engine 13 may receive inputs from Separately Stored Synchronized Video and Audio Content 1, a Playback List 109, and a Graphical User Interface 16.

5 During retrieval, the Selecting Engine 13 passes the audio content synchronized with the visual content to a speech recognition and tagging system 12 so that the parts of the content 1 that are speech and the parts that are noise are tagged and provided to Tagged Speech 7 storage, and Noise 23 storage. The speech recognition and tagging system 12 also inputs individual words or tokens into Tagged Speech 7. Hereinafter, a “token” is any
10 successive group of non-delimiter characters appearing in a string preceded by a delimiter (or appearing at the beginning of the string), wherein a delimiter may be a space, for example, between words or a form of punctuation such as a comma. Hereinafter, “synchronization” of speech or written words or phrases with visual content means words are uttered or written with corresponding visual content when said visual content is
15 displayed. Audio content synchronized with the visual content is available because the Synchronized Video and Audio Content 1 is stored separately, and the Separately Stored Synchronized Video and Audio Content 1 is retrievable for synchronized playback.

Referring to FIG. 2, the Phrase and Tokens Recognizing portion 2 of the apparatus
10 includes: a decision step 29 for determining Valid Words for Phrases, wherein the decision is based on a Test Acceptable Words For Validity 21 input and a Phrase Database 42 input. Hereinafter, “words” or “speech” mean written or spoken English language, or any other language. The decision 29 provides an output Join Words Into Phrases step 31. The Test Acceptable Words For Validity 21 may receive an Input Pronunciation Rules 39. Here, the Test Acceptable Words For Validity 21 may use pronunciation rules to cause the
25 valid words to be pronounced correctly on playback. Hereinafter, “pronouncing correctly” means correcting speech for pronunciation error due to accents or mispronunciations. Consecutive successive valid words and a Phrase Database 42 are input into the decision step 29, resulting in a determination whether the successive valid words are valid words for phrases. If yes, the consecutive successive valid words for phrases are input to the Join
30 Words Into Phrases 31 step. If no, the consecutive successive valid words for phrases are input into Buffer of stored Playback Content 37 as words not valid for phrases. The decision step 29 may apply a process that may include comparison of consecutive

successive valid words with a database of phrases 42. Valid words that are present in the Phrase Database 42 as phrases may be joined in the Join Words Into Phrases 31 step.

Dictionaries, or Lexicons and the like are examples of the Phrase Database 42.

Some examples of phrases include phrases such as "good morning," whose component

5 words often go together. Since the words of the phrases need to be uttered together, then the words of the phrases are uttered together when the corresponding visual content of the Separately Stored Synchronized Video and Audio Content 1 is played back. The user could also be given the option to input additional words or rules into the Test Acceptable Words For Validity 21 step so that other words not part of an established language could
10 also be joined together in phrases in the step 31.

Referring to FIG. 2, the Selectable Rate Playback portion 4 comprises: a Buffer of Stored Playback Content 37; a Selectable Rate Playback Engine 67 and a Selectable Rate Playback Viewing 73. Phrases may be passed into Buffer of Stored Playback Content 37 from the Join Words Into Phrases step 31. Alternatively, valid words may be provided to
15 the Buffer of Stored playback Content 37 if they are determined by decision step 29 to not be valid words for phrases. Alternatively, Noise 23 may be passed to the Buffer of Stored Playback Content 37. In one embodiment, the Selectable Rate Playback Engine 67 provides the Buffer of Stored Playback Content 37 to the Selectable Rate Playback Engine 67. The Selectable Rate Playback Engine 67 provides input to the Selectable Rate
20 Playback Viewing step 73 for Selectable Rate Playback Viewing 73 of the Selected Separately Stored Synchronized Video and Audio Playback Content 1. One purpose of the Selectable Rate Playback Viewing 73 relates to the user not having understood what was uttered or a scene content in a video program was not being clear. In the example where the uttered words were not understood clearly by the user, the Test Acceptable Word
25 Validity 21 may use a pronunciator device that inputs pronunciation rules 39 and then utters the words or phrases correctly. Thus, words incorrectly spoken by an actor may be correctly pronounced by the pronunciator. The user could be given the option whether the valid words should employ a pronunciator for utterance or if the utterance should be as they are spoken by actors in, for example, the video program.

30 FIG. 3 depicts an example of a List 110 of playback content from the Playback List 109. The Playback List includes a playback "y" minutes list item 120, wherein y represents a time from when the Separately Stored Synchronized Video and Audio Content

1 (see FIG. 2) was stored. The time from when the Separately Stored Synchronized Video and Audio Content 1 was stored depends on a storage capacity of the Buffer of Stored playback Content 37, as depicted in FIG. 2, and described herein. The storage capacity of the Buffer of Stored playback Content 37 may be any appropriate capacity needed to
5 accommodate the Separately Stored Synchronized Video and Audio Content 1. In one embodiment the storage capacity of the Buffer of Stored Playback Content 37 is less than 2 minutes. Alternatively, the storage capacity of the Buffer of Stored Playback Content 37 may be less than 5 minutes. Alternatively, the storage capacity of the Buffer of Stored Playback Content 37 may be the capacity required to store the Separately Stored
10 Synchronized Video and Audio Content 1 of the movie or video program, wherein the video program may be a television program.

The Playback List 109 includes a Keywords or Phrases List Item 130 that may be created by a user based on keywords or phrases that the user remembers from listening or viewing the program or movie, that is included in the Separately Stored Synchronized
15 Video and Audio Content 1.

The Playback List 109 includes a Key Frames List Item 140, wherein each entry of the Key Frames List Item 140 may be selected by subtracting an intensity “z” of each of two consecutive successive frames and if the difference “ Δz ” in the intensity “z” between the consecutive successive frames is greater than a threshold “t” then the frame having the
20 higher intensity is selected as the Key Frame. A user can select list items 120, 130 or 140 manually or via a remote selection device. Selection of the list items 120, 130 or 140 provides an input to the Selecting Engine 13.

FIG. 4 depicts a List of playback content from a Graphical User Interface (GUI) 16, wherein the List includes a playback “y” minutes list item 160, a Keywords or Phrases List
25 Item 170 and a Key Frames List Item 180 created in like manner as the corresponding list items 120, 130, and 140 depicted in FIG. 3 and described supra. The List of Playback Content from the GUI 16 includes a scroll bar 190 that can be used to scroll to 160, 170 or 180. A user can select list items 160, 170 or 180 manually or via a remote selection device. Selection of the list items 160, 170 or 180 provides an input to the Selecting
30 Engine 13 from the GUI 16 (see FIG. 2). The graphical user interface 16 may be provided with a list of key video frames using key frame extraction. Hereinafter, “key frame

extraction” means the key frames having a higher intensity than a threshold intensity are selected into the List of Playback Content from the GUI 16.

FIG. 5 depicts a method 70 for Selectable Rate Playback of Playback Content, comprising steps 75, 85, 90, 95 and 97. In one embodiment, a television program, or alternatively, a movie may be stored on a personal video cassette recorder, a DVD or on any appropriate storage medium such as an optical medium, or a magneto optical medium. The program or movie must be Separately Stored Synchronized Video and Audio Content 1 (see FIG. 2), wherein the video and audio are synchronized as stored, and wherein the Separately Stored Synchronized Video and Audio Content 1 are retrievable for synchronized playback. During a playback of the Separately Stored Synchronized Video and Audio Content 1, a user may encounter a portion of the program that may not be satisfactorily understandable such as because either the video portion is unclear or the audio portion is not understandable. The user first stops the playback. In the step 75, a user selects a first portion 44 of the Separately Stored Synchronized Video and Audio Playback Content 1 for “Selected Rate” 49 of playback, wherein the selected first portion 44 corresponds to an list item 120, 130, or 140 from the Playback List 109 of FIG. 3, or a list item 160, 170, or 180 from the GUI 16 of FIG. 4. The playback content 1 has been stored at a storing rate, wherein the storing rate may be any recording rate for a commercial personal video cassette recorder, a DVD or for any appropriate storage medium such as an optical medium, or a magneto optical medium, and wherein the storing rate is different from the “Selected Rate” 49. The “Selected Rate” 49 may be slower or faster than the storing rate for the the playback content 1 without causing distortion of the speech portion of the audio content of the playback content 1.

In the step 85, speech included in the selected first portion 44 of the Separately Stored Synchronized Video and Audio Playback Content 1 (see FIG. 2) corresponding to the selected list item from the Playback Content from Playback List 109 or the Graphical User Interface 16 is tagged by the Speech Recognition and Tagging System 12. In the step 90, acceptable words 7 are recognized by the speech recognition and tagging system 12 (see FIG. 2).

In the step 95, at least one phrase in the Tagged Speech 7 is recognized by the Phrase and Tokens Recognizing portion 2 (see FIG. 2) of the apparatus 10.

In the step 97, the selected first portion 44 of the Separately Stored Synchronized Video and Audio Content 1 (see FIG. 2) may be retrieved for synchronized playback by the Selecting and Tagging Engine 65 (see FIG. 1), since the video and audio content are synchronized and stored separately, wherein Tagged Speech 7 and corresponding video is presented serially, such that selecting the first portion 44 of the Separately Stored Synchronized Video and Audio Content 1 (see FIG. 2) for playing selects a corresponding Tagged Speech 7 for playing.

Speech may be tagged by the Speech Recognition and Tagging System 12, as depicted in FIG. 2, and described in associated text supra. An at least one phrase in the Tagged Speech 7 may be recognized using, for example, the Speech Recognition System and Tagging System 12, as depicted in FIG. 2, and described in associated text supra. The Speech Recognition and Tagging System 12 may use stemming to remove morphological and inflexional endings from words in English from the playback content 1. Hereinafter, "stemming" may be accomplished by the Porter stemming apparatus (or 'Porter stemmer') that is a process for removing the commoner morphological and inflexional endings from words in English. Its main use is as part of a term normalization process that is usually done when setting up Information Retrieval systems. Hereinafter, "morphological" endings for words in English are verb tenses, such as past, present or future, and "inflexional" endings for words in English are endings of nouns or verbs such as "s", "es", or "ing", or endings such as "er", "ier", "iest" for comparative and superlative forms of adjectives.

The selected first portion 44 of the Separately Stored Synchronized Video and Audio Playback Content 1 (see FIG. 2) corresponding to the selected list item from the Playback Content from Playback List 109 or the Graphical User Interface 16 may be played at selectable rate, wherein said playing synchronously retrieves Tagged Speech 7 such as acceptable words. Playing the selected first portion 44 of the Separately Stored Synchronized Video and Audio Playback Content 1 corresponding to the selected list item from the Playback Content from Playback List 109 or the Graphical User Interface 16 at the selectable rate does not result in distortion of speech in the Playback Content 1 (see FIG. 2). The video and audio are to be synchronized at the selectable rate, in accordance with embodiments of the present invention and in accordance with a method, as depicted by the flow chart 70 in FIG. 5 and described supra.

CLAIMS:

1. A method for playback of playback content at selectable rates, comprising:
selecting a first portion of separately stored video and audio playback content,
wherein the playback content has been stored at a storing rate, wherein the video and audio are synchronized as stored, and wherein the separately stored synchronized video and audio content are retrievable for synchronized playback;
selecting a rate of playback of the playback content from the selectable rate,
wherein the selected playback rate is different from the storing rate;
tagging speech in the selected first portion of the playback content;
recognizing an at least one phrase in the tagged speech; and
playing said first portion of playback content at said rate of playback, wherein said playing synchronously retrieves the tagged speech, and wherein playing at said rate does not result in distortion of speech in the playback content even though said rate is different than the storing rate, and wherein the video and audio are synchronized at said rate of playback during said playing.
2. The method of Claim 1, wherein the first portion of the playback content is selected for playing from a playback list.
3. The method of Claim 1, wherein the first portion of the playback content is selected for playing from a graphical user interface.
4. The method of Claim 3, wherein the graphical user interface includes a list of key video frames provided by key frame extraction.
5. The method of Claim 1, wherein tagging the speech further comprises recognizing a plurality of valid words for phrases in the tagged speech.
6. The method of Claim 1, wherein said rate of playback is less than the storing rate.

7. The method of Claim 1, wherein recognizing the at least one phrase of the tagged speech is accomplished by speech recognition.
8. The method of Claim 1, further comprising removing the commoner morphological and inflexional endings from words in English from the playback content by stemming.
9. The method of Claim 9, wherein the key frames in the list of key video frames have a higher intensity than a threshold intensity.
10. The method of Claim 1, wherein tagged speech and corresponding video is presented serially during storing the playback content and playing the first portion of playback content at the rate of playback.
11. The method of Claim 1, wherein playing said first portion of playback content at the rate of playback further comprises playing on an audio and video device, such that when the playing is stopped by a stop input, normal viewing of the audio and video device results.
12. The method of Claim 1, wherein playing said first portion of playback content at the rate of playback further comprises playing on an audio and video device, such that when the playing is paused by a pause input, wherein playing is paused for greater than x minutes, wherein x is any positive real number, normal viewing of the audio and video device results.
13. An apparatus for selectable rate playback of playback content, comprising:
 - a separately stored video and audio playback content, wherein the playback content has been stored at a storing rate;
 - a selected first portion of the separately stored video and audio playback content in a storage medium, wherein the selected first portion of video and audio content are synchronized and a speech portion of the audio content is tagged;
 - a speech recognition device for tagging the speech portion of the audio content;

a phrase recognition device for determining valid words for phrases from the tagged speech, wherein the valid words are joined into said phrases;

a playback device for playback of the selected first portion of the playback content at a rate selected from the selectable rate, wherein the selected rate is different from the storing rate, wherein playback at the selected rate synchronously retrieves the tagged speech portion of the audio content, wherein playback at the selected rate does not result in distortion of speech in the playback content even though the selected rate is different from the storing rate, and wherein the video and audio content are synchronized at the selected rate during said playback.

14. The apparatus of Claim 13, wherein the playback device for playback of the selected first portion of the playback content at the selected rate further comprises a playback list of the selected first portion of separately stored synchronized video and audio playback content.

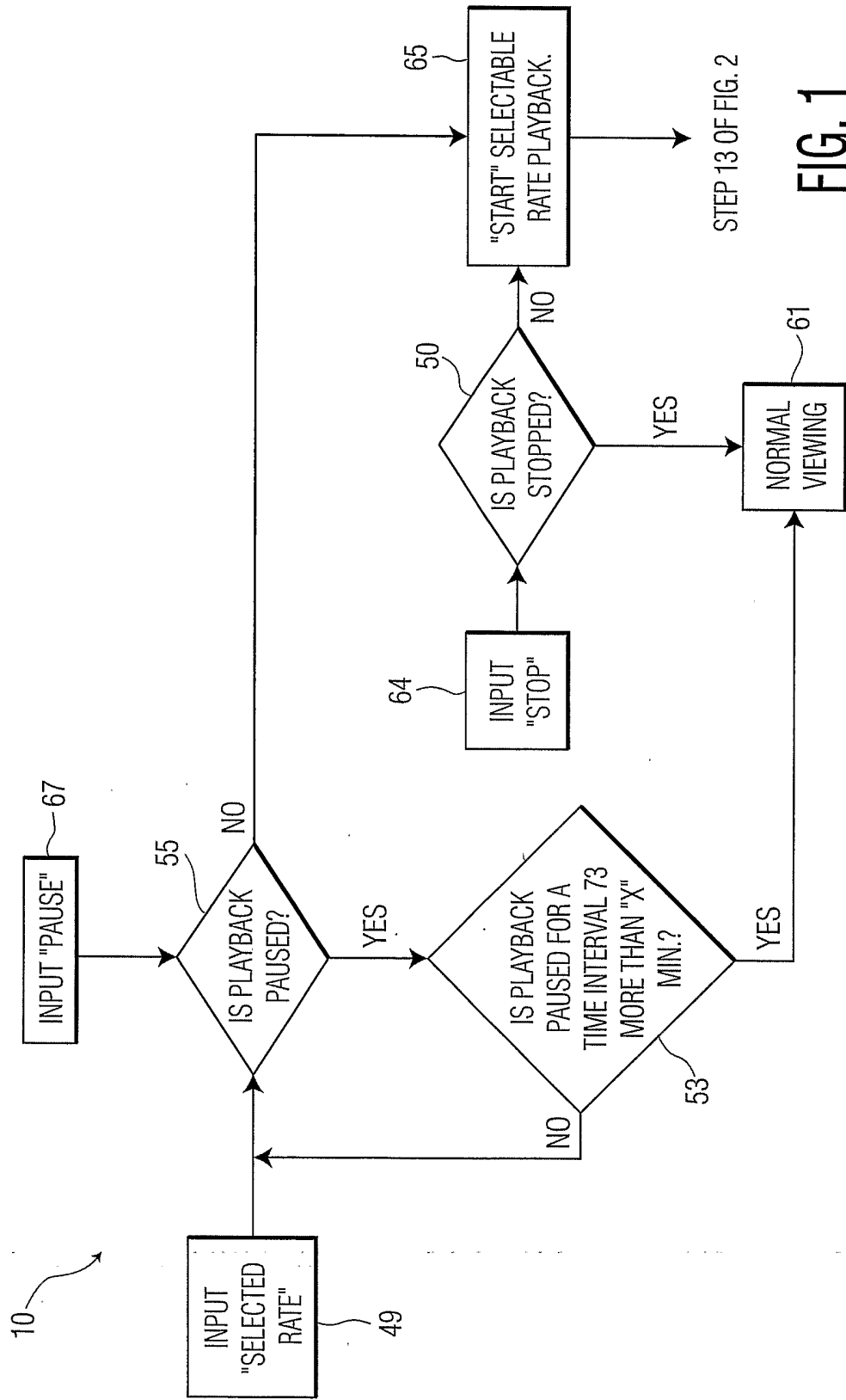
15. The apparatus of Claim 13, wherein the playback device for playback of the selected first portion of the playback content at the selected rate further comprises a graphical user interface of the selected first portion of separately stored synchronized video and audio playback content.

16. The apparatus of Claim 15, wherein the graphical user interface includes a Key Frames List Item, wherein each frame of the Key Frames List Item has an intensity that differs in intensity from an intensity of a consecutive successive frame by more than a threshold value..

17. The apparatus of Claim 13, wherein the phrase recognition device for determining valid words for phrases from the tagged speech includes a join words into phrases step.

18. The apparatus of Claim 13, wherein the phrase recognition device for determining valid words for phrases from the tagged speech includes a pronunciation rules input to cause the valid words to be pronounced correctly on playback.

19. The apparatus of Claim 13, wherein the video content is in video frames.
20. The apparatus of Claim 13, wherein the selected rate of playback is slower than said storing rate, and playing at the selected rate of playback does not result in distortion of speech in the playback content even though the selected rate is slower than the storing rate of the playback content.



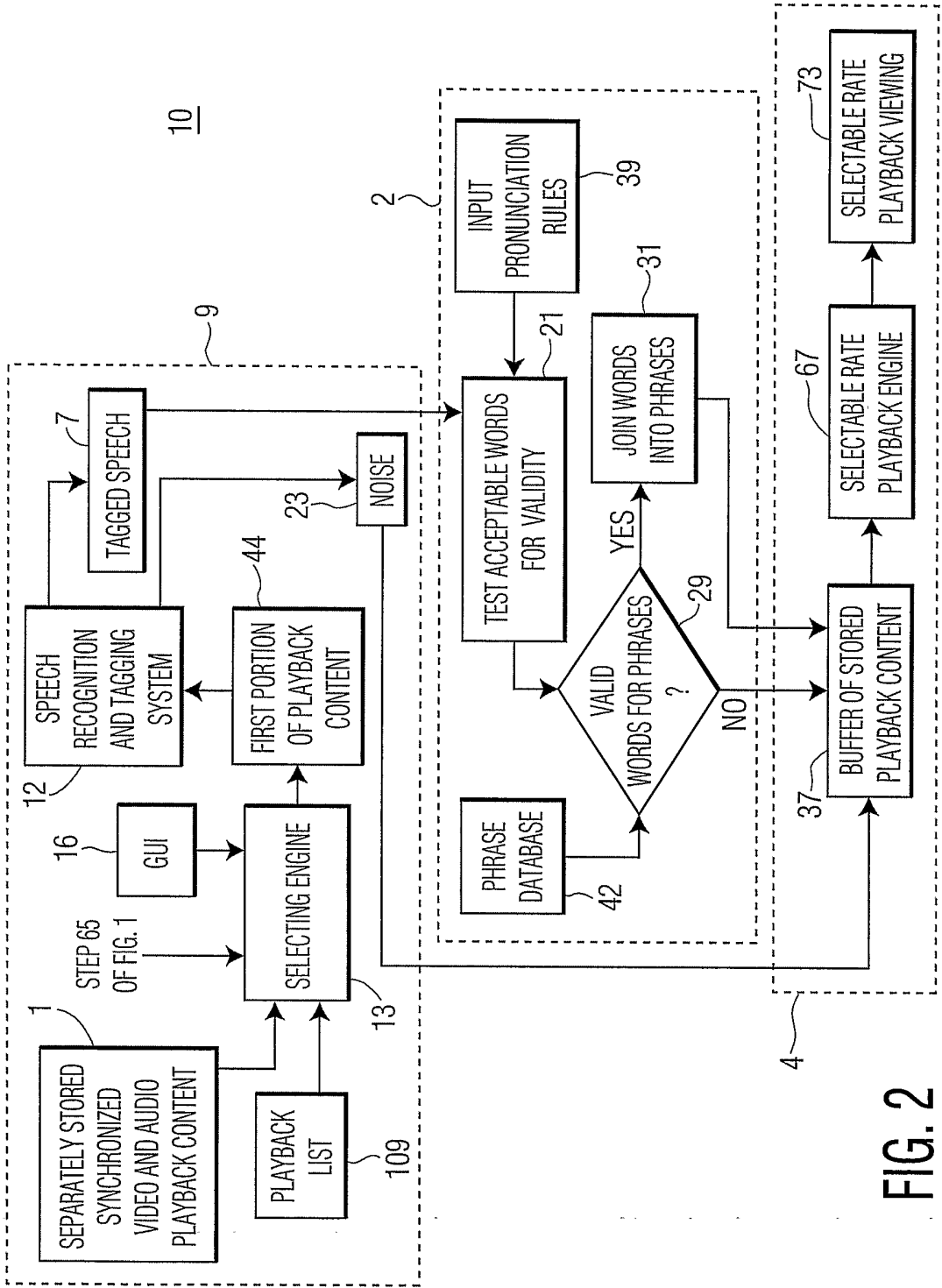


FIG. 2

3/4

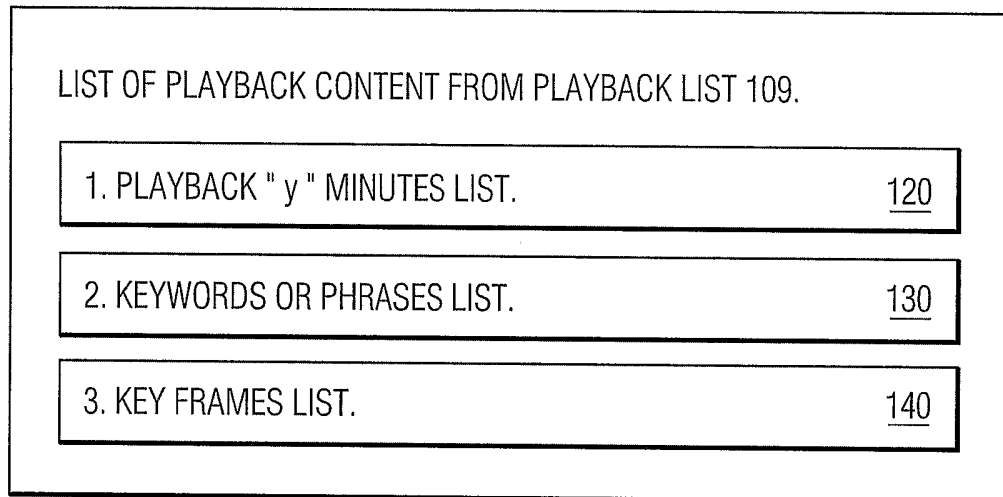


FIG. 3

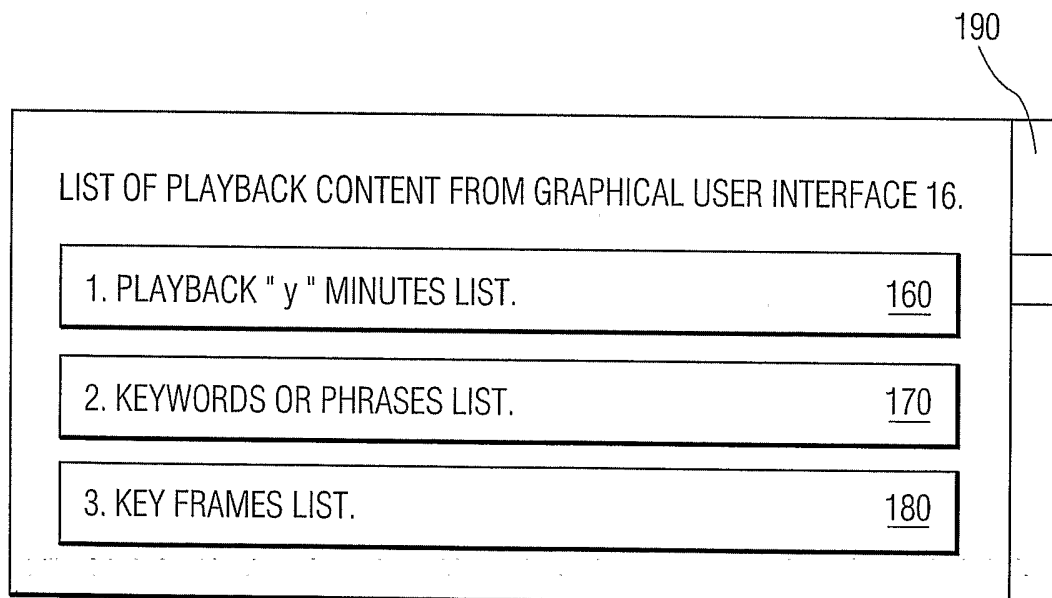


FIG. 4

4/4

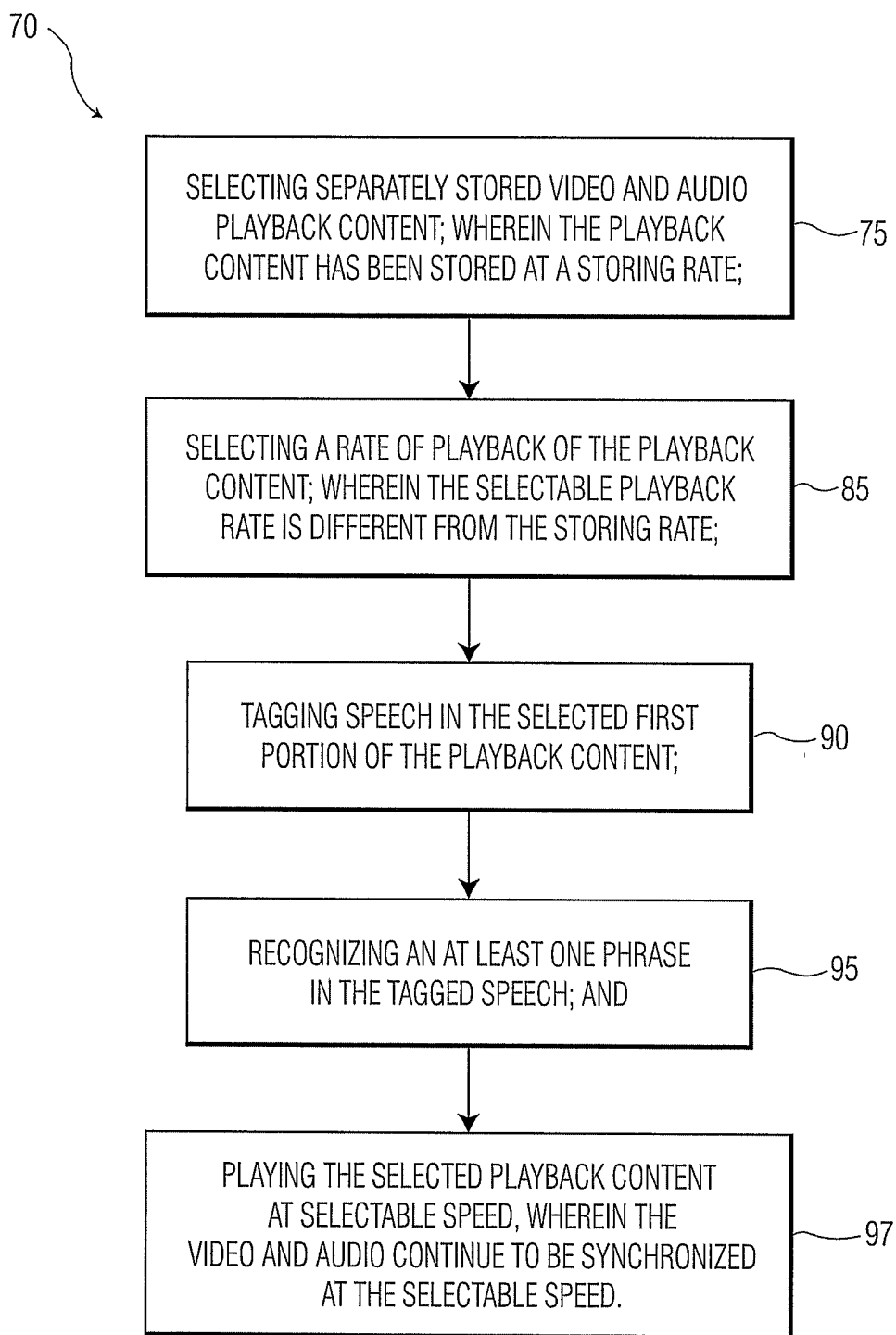


FIG. 5