

## CONDENSED VOICE BUFFERING, TRANSMISSION AND PLAYBACK

### FIELD

[0001] This disclosure relates generally to voice communication and, more particularly, to processing voice information for recording, transmission and playback.

### BACKGROUND

[0002] Communication of voice information using digital techniques generally involves the use of a voice encoder, sometimes referred to as a voice CODEC or vocoder. The voice encoder samples, digitizes and compresses voice information, e.g., speech, for transmission as a series of frames. Many voice encoders provide variable rate encoding. For example, different types of voice information, such as speech, background noise, and pauses can be encoded at different data rates. Compression enables the voice information to be transmitted at a reduced data rate, e.g., over a wired or wireless transmission channel. Voice information may be digitally transmitted, for example, over packet-based networks, such as networks supporting Voice-Over-IP (VOIP).

[0003] Frame-based voice encoding techniques, such as Qualcomm Code Excited Linear Predictive Coding (QCELP), Enhanced Variable Rate Codec (EVRC), and Selectable Mode Vocoder (SMV), encode moments of sound into sequences of bits. The bit sequences represent the sound during the encoded moments, and are commonly referred to as frames. Typically, the encoded frames represent a continuous stream of voice information that is later decoded and synthesized to produce audible output. In particular, the encoded frames may contain parameters that relate to a model of human speech generation. Recognizable speech typically includes pauses following utterances. Accordingly, some of the encoded frames contains the coding of pauses in speech. A decoder uses the parameters received over a transmission channel to resynthesize the speech for audible playback.

### SUMMARY

[0004] This disclosure is directed to techniques for condensed voice buffering, transmission and playback. The condensation techniques may involve identification of encoded voice frames as either speech or a pause, and selective exclusion of frames, for

storage, transmission or playback, based on the identification. In this manner, the techniques are capable of condensing a series of encoded voice frames. Condensation may be effective in reducing the amount of frames stored in memory, transmitted between devices, or decoded and synthesized for playback.

[0005] When variable-rate coding is employed, a pause frame may be identified, for example, based on a threshold comparison for the rate of the encoded frame. Other voice coding techniques may explicitly indicate frames of silence. Some voice coding techniques include noise estimates in the pause frames. In some cases, the techniques may involve excluding only a portion of the identified frames from a consecutive sequence of the identified frames, thereby preserving a minimum number of the identified frames needed for intelligible conversation.

[0006] In one embodiment, a method comprises identifying encoded voice frames representing a pause, and excluding at least some of the identified frames from a series of frames.

[0007] In another embodiment, a device comprises a voice encoder and a processor. The voice encoder generates encoded voice frames. The processor identifies encoded voice frames representing a pause, and excludes at least some of the identified frames from a series of frames.

[0008] In a further embodiment, a machine-readable medium comprises instructions to cause a processor to identify encoded voice frames representing a pause, and exclude at least some of the identified frames from a series of frames.

[0009] In an added embodiment, a machine-readable medium comprises a series of encoded voice frames representing a speech sequence. The series of encoded voice frames omit at least some of the encoded voice frames representing pauses in the speech sequence.

[0010] In another embodiment, a system comprises first and second voice communication devices. The first voice communication device has a voice encoder that generates encoded voice frames, a processor that identifies encoded voice frames representing a pause, and excludes at least some of the identified frames from a series of the frames, and a transmitter that transmits the series of frames. The second voice communication device has a receiver that receives the series of frames transmitted by the first communication device, and a voice decoder that decodes the series of frames for playback.

[0011] Additional details of these and other embodiments are set forth in the accompanying drawings and the description below. Other features will become apparent from the description and drawings, and from the claims.

### **BRIEF DESCRIPTION OF DRAWINGS**

[0012] FIG. 1 is a block diagram illustrating an exemplary voice communication system that employs techniques for condensed voice buffering, transmission and playback.

[0013] FIG. 2 is a block diagram illustrating an exemplary voice communication system in greater detail.

[0014] FIG. 3 is a block diagram of an exemplary voice communication device.

[0015] FIG. 4 is a timing diagram of an exemplary speech sequence.

[0016] FIG. 5 is a timing diagram of the speech sequence of FIG. 4 following encoding to produce a series of encoded voice frames.

[0017] FIG. 6 is a timing diagram of the encoded voice frames of FIG. 5 illustrating identification of pause frames to be excluded from the frame series.

[0018] FIG. 7 is a timing diagram of the encoded voice frames of FIG. 6 following exclusion of the identified pause frames.

[0019] FIG. 8 is a flow diagram illustrating exclusion of pause frames for storage of a series of encoded voice frames in memory.

[0020] FIG. 9 is a flow diagram illustrating exclusion of pause frames for transmission of a series of encoded voice frames.

[0021] FIG. 10 is a flow diagram illustrating exclusion of pause frames for playback of a series of encoded voice frames.

[0022] FIG. 11 is a flow diagram illustrating a technique for identification and selection of pause frames for exclusion from a series of encoded voice frames.

[0023] FIG. 12 is a flow diagram illustrating another technique for identification and selection of pause frames for exclusion from a series of encoded voice frames.

### **DETAILED DESCRIPTION**

[0024] FIG. 1 is a block diagram illustrating a voice communication system 10. As shown in FIG. 1, system 10 may include two or more voice communication devices

12A, 12B (hereinafter 12) that communicate voice information via a network 14. Exemplary voice communication devices 12 may include conventional land-line telephones, IP-equipped telephones, cellular radiotelephones, satellite phones, and computers with IP telephony capabilities.

[0025] In the case of wireless communication, voice communication devices 12 may communicate according to one or more wireless communication standards such as CDMA, GSM, WCDMA, and the like. In addition to voice communication, voice communication devices 12 may be capable of transmitting and receiving data via network 14. Hence, network 14 may represent a packet-based network, a switched telecommunication network, or a combination thereof.

[0026] Voice communication devices 12 may be equipped with variable rate vocoders that compress moments of sound into sequences of bits referred to as encoded voice frames. In accordance with this disclosure, one or more of voice communication devices 12 may implement techniques for condensed voice buffering, transmission and/or playback.

[0027] The techniques implemented by voice communication devices 12 may involve identification of encoded voice frames as representing either speech or a pause, and selective exclusion of frames for storage, transmission or playback based on the identification. In this manner, the techniques are capable of condensing, i.e., shortening, a series of encoded voice frames. Condensation may be effective in reducing the amount of frames stored in memory, transmitted between devices, or decoded and synthesized for playback.

[0028] When variable rate coding is employed, voice communication device 12 may identify a pause frame, for example, based on a threshold comparison for the rate of the encoded frame. In some cases, the condensation techniques implemented by voice communication device 12 may involve excluding only a portion of the identified pause frames from a consecutive sequence of the identified frames, thereby preserving a minimum number of the identified frames needed for intelligible conversation, as some amount of pause may be a necessary component of conversation.

[0029] Condensation may take place within a "sending" voice communication device 12 that encodes frames based on voice input. The voice input may be entered via a microphone associated with the sending voice communication device 12. In this case, the condensation may occur prior to buffering of the frames in memory. In other words,

voice communication device 12 may exclude pause frames produced by the vocoder before the frames are stored in memory. Alternatively, voice communication device 12 may exclude the pause frames upon retrieval from memory, but prior to transmission via network 14.

**[0030]** Condensation also may take place within a "receiving" voice communication device 12 that decodes frames and synthesizes the frame content to produce voice output. Voice output may be produced by a speaker associated with the receiving voice communication device 12. In this case, the encoded voice frames are sent across network 14 and stored in memory at the receiving voice communication device 12. However, the receiving voice communication device 12 does not decode all of the encoded voice frames. Instead, the receiving voice communication device 12 excludes selected pause frames from decoding, synthesis and playback.

**[0031]** Condensing encoded voice frames prior to storage in memory, i.e., in a sending voice communication device 12, can promote more optimal storage within memory without changing the format or coding of the stored information. If QCELP encoding is employed, for example, voice communication device 12 can be configured to selectively exclude pause frames without altering the QCELP coding. Conversely, there is also no need to change the techniques for decoding and synthesizing the stored QCELP frames upon transmission to receiving voice communication device 12. Rather, there are simply less pause frames to decode at the receiving voice communication device 12.

**[0032]** With condensation of frames prior to storage, it may be possible to reduce memory requirements within voice communication device 12. Condensation may be used in combination with additional compression to further improve storage utilization. In addition, by reducing the number of frames associated with a speech sequence, condensation can promote conservation of transmission bandwidth, reduced processing overhead, reduced power consumption, and reduced latency. With respect to latency, in particular, condensation can be used to reduce network delays introduced by channel setup and maintenance.

**[0033]** Similarly, condensing encoded voice frames already stored in memory at the sending voice communication device 12, e.g., prior to transmission to a receiving voice communication device 12, can promote conservation of transmission bandwidth, reduced processing overhead, reduced power consumption, and reduced latency.

Condensing encoded voice frames already stored in memory at the receiving voice communication device 12 can reduce processing overhead and power consumption need for decoding, synthesis and playback. For example, excluding frames from a series of frames for playback reduces the number of frames that need to be decoded and synthesized. Power conservation may be particularly advantageous for mobile, battery-powered voice communication devices.

[0034] FIG. 2 is a block diagram illustrating voice communication system 10 in greater detail. In particular, FIG. 2 illustrates one possible environment for operation of voice communication devices 12 and implementation of a voice condensing techniques as described herein. As shown in FIG. 2, a first voice communication device 12A may take the form of a wireless device that communicates with a base station transceiver 11. A base station controller 13 may provide access to a packet-based network 15 via a packet data serving node 17. Base station 12 also may provide access to telephones or telephony devices coupled to public switched telephone network (PSTN) 19. In this manner, base station controller 12 may route calls between voice communication devices 12 and other remote network equipment or telephony equipment connected to packet-based network 15 or PSTN 19.

[0035] Voice communication device 12A communicates with voice communication device 12B via packet-based network 15, and communicates with voice communication device 12C via PSTN 19. Although voice communication devices 12A, 12B, and 12C are shown in FIG. 2 for purposes of illustration, system 10 may contain a large number of voice communication devices. Voice communication device 12B may receive voice information in the form of IP packets containing encoded voice frames. As described herein, voice communication devices 12A, 12B may employ condensation techniques to selectively exclude pause frames from the encoded voice frames sent and received by the devices.

[0036] FIG. 3 is a block diagram of a voice communication device 12 in greater detail. In the example of FIG. 3, voice communication device 12 takes the form of a wireless communication device such as a cellular radiotelephone. As shown in FIG. 3, voice communication device 12 may include a processor 16, a modem 18, transmit/receive circuitry 20, memory 22 and vocoder 24. Processor 16 controls modem 18 to transmit and receive communications via transmitter/receiver circuitry 20.

Transmit/receive circuitry 20 transmits and receives wireless signals via a radio frequency antenna 21.

[0037] As further shown in FIG. 3, processor 16 also may process user input, including text received from a keypad or other input media (not shown). Vocoder 24 receives voice input received from a microphone 23 via audio circuitry 25. Vocoder 24 encodes and compresses the voice input received from microphone 23 using an encoding technique such as QCELP, EVRC, SMV or the like. In addition, vocoder 24 decodes and synthesizes encoded voice frames received via transmit/receive circuitry 20. Audio circuitry 25 drives speaker circuitry 27 to produce audible voice output based on the results provided by vocoder 24.

[0038] Processor 16 executes instructions stored in memory 22 to control communications and implement voice condensation techniques as described herein. Memory 22 may take the form of random access memory (RAM), read-only memory (ROM), non-volatile random access memory (NVRAM), electrically erasable programmable read-only memory (EEPROM), flash memory, and the like. Memory 22 also may serve as a buffer for encoded voice frames processed by vocoder 24. Alternatively, a dedicated voice buffer may be provided.

[0039] In some embodiments, vocoder 24 may be integrated with processor 16 or modem 18. Alternatively, processor 16, modem 18 and vocoder 24 may be integrated together as a single processing unit. Accordingly, although FIG. 3 depicts processor 16, modem 18 and vocoder 24 as separate units, they may be implemented in a variety of different arrangements using shared hardware. For example, the functions performed by processor 16, modem 18 and vocoder 24 may be programmable features of a microprocessor or DSP, or features implemented in an ASIC, FPGA, discrete logic circuitry or the like. Moreover, in some embodiments, certain functions attributed to processor 16, modem 18 and vocoder 24 may be performed by the other units.

[0040] In operation, processor 16 identifies encoded voice frames, produced by vocoder 24, that represent a pause, and selectively excludes at least some of the identified frames from a series of frames to be stored in memory 22, transmitted via transmit/receive circuitry 20, or retrieved from memory 22 for decoding, synthesis and playback by vocoder 24. In this manner, processor 16 can be configured to promote memory, bandwidth, power, and processing efficiency as well as reduced latency.

[0041] FIG. 4 is a timing diagram of an exemplary speech sequence 26. Although speech sequences vary based on the course of a conversation, they are generally characterized by bursts of speech, or "utterances," separated by periods of no speech, i.e., pauses. Indeed, to be intelligible, speech ordinarily must include pauses between utterances. Hence, upon voice encoding, certain frames will contain the encoding of pauses. As shown in FIG. 4, a particular speech sequence 26 includes a pause period 28, followed by speech period 30, pause period 32, speech period 34 and pause period 36.

[0042] FIG. 5 is a timing diagram of the speech sequence 26 of FIG. 4 following encoding to produce a series of encoded voice frames. Each frame is designated as either a pause (P) frame or a speech (S) frame. Ordinarily, a variable rate vocoder will encode pause frames and speech frames at different rates. Accordingly, pause and speech frames can be readily distinguished by comparing the encoding rate to a threshold rate. In particular, a pause frame typically will be encoded at a lower rate than a frame containing speech.

[0043] FIG. 6 is a timing diagram of the encoded voice frames of FIG. 5 illustrating identification of pause frames to be excluded from the frame series in accordance with the condensation techniques described herein. Because speech sequence 26 is encoded frame-by-frame, the pauses between utterances can be shortened by removing some of the pause frames. As shown in FIG. 6, pause frames corresponding to areas 38 and 40 are eliminated to condense the overall length of speech sequence 26. Area 38 and 40 each correspond to two pause frames, in the example of FIG. 6, that are excluded from the series of frames representing speech sequence 26.

[0044] Notably, not all of the pause frames are excluded in the example of FIG. 6. Rather, in many cases, it will be desirable to exclude only a portion of the pause frames to thereby preserve the intelligibility of speech sequence 26. If all of the pause frames were removed, there would be no separation between speech frames, resulting in speech output that is either unintelligible or difficult to understand. Accordingly, the condensation techniques applied to speech sequence 26 may make use of a minimum pause length threshold to retain a sufficient number of pause frames for intelligibility. Thus, the minimum pause length may be based on the intelligibility needs of the decoded speech.

[0045] In addition to intelligibility, encoded pauses can contain useful information, such as metrics for a background noise level. A receiving device typically uses the background noise level to adjust gain or other playback parameters. To maintain the most up-to-date information, it may be desirable to retain the last frame in a pause, i.e., the last frame in a series of consecutive pause frames. In this case, the pause frames to be excluded can be taken from the beginning or middle of a series of pause frames. At least some of the pause frames are retained in the frame series to permit intelligibility and, optionally, to retain other useful information, such as the background noise level.

[0046] The threshold for pause frame retention may be an absolute number of frames. For example, the condensation process may be configured to exclude only those pause frames in excess of a minimum number of pause frames. Alternatively, the process could be configured to retain a relative pause length. In this case, a minimum percentage of pause frames are retained. Thus, following condensation, a longer pause may retain more frames than a shorter pause. Again, the threshold may work in conjunction with retention of the last frame of a pause, i.e., a last frame rule, for background noise level.

[0047] As an example of the application of a threshold and last-frame rule, FIG. 6 illustrates retention of all of the pause frames associated with pause 32. Whereas pause 28 and pause 36 are modified to exclude a number of pause frames, pause 32 is unchanged due to the effects of the retention threshold and the last frame rule. The results provided in FIG. 6 are for purposes of illustration only. Results may vary according to the particular retention threshold and whether a last frame rule applies.

[0048] FIG. 7 is a timing diagram of the encoded voice frames of FIG. 6 following exclusion of the identified pause frames. As indicated in FIG. 7, the result is a shortened series of encoded voice frames. Upon playback, the pauses between utterances are reduced, but not so much as to adversely affect intelligibility. Over the course of several speech sequences, exclusion of pause frames can result in substantial savings in latency, and reduce bandwidth, power and processing consumption.

[0049] FIG. 8 is a flow diagram illustrating exclusion of pause frames for storage of a series of encoded voice frames in memory. In particular, FIG. 8 represents exclusion of pause frames produced by a vocoder within a sending voice communication device 12 prior to buffering to conserve memory resources. By storing a reduced length speech

sequence, however, bandwidth, latency, processing and power consumption advantages also may result.

[0050] As shown in FIG. 8, the condensation technique may involve obtaining a series of encoded voice frames from a vocoder (42), and identifying encoded voice frames representing a pause (44). The technique further involves excluding either an absolute number or a specified percentage of the identified pause frames from the series of encoded voice frames (46), subject to minimum pause length and last frame rules as discussed above. Upon excluding the pause frames, the technique involves storing the pause-shortened frame series in memory (48), such as memory 22 shown in FIG. 3.

[0051] FIG. 9 is a flow diagram illustrating exclusion of pause frames for transmission of a series of encoded voice frames. In particular, FIG. 9 represents exclusion of pause frames produced by a vocoder within a sending voice communication device 12 prior to transmission of frames representing a speech sequence. In this case, all of the frames produced by the vocoder are stored in memory, but at least some of the pause frames are omitted prior to transmission. By transmitting a reduced length speech sequence, bandwidth, latency, processing and power consumption advantages may result.

[0052] As shown in FIG. 9, the condensation technique may involve retrieving a series of encoded voice frames from memory (50), and identifying encoded voice frames representing a pause (52). The technique further involves excluding either an absolute number or a specified percentage of the identified pause frames from the series of encoded voice frames (54), subject to minimum pause length and last frame rules. Upon excluding the pause frames, the technique involves transmitting the pause-shortened frame series (56), e.g., to a receiving voice communication device 12.

[0053] FIG. 10 is a flow diagram illustrating exclusion of pause frames for playback of a series of encoded voice frames. In particular, FIG. 10 represents exclusion of pause frames retrieved from memory in a receiving voice communication device 12 to reduce the number of frames decoded and synthesized by a vocoder residing in the device prior to playback. In this case, all of the frames received from a sending voice communication device 12 are stored in memory in the receiving voice communication device, but at least some of the pause frames are omitted prior to decoding, synthesis and playback. By decoding a reduced length speech sequence, processing and power consumption advantages may result in the receiving voice communication device 12.

[0054] As shown in FIG. 10, the condensation technique may involve retrieving a series of encoded voice frames from memory (58), and identifying encoded voice frames representing a pause (60). The technique further involves excluding either an absolute number or a specified percentage of the identified pause frames from the series of encoded voice frames (62), subject to minimum pause length and last frame rules. Upon excluding the pause frames, the technique involves decoding and synthesizing the pause-shortened frame series (64) for playback. In some embodiments, exclusion of stored pause frames may be accomplished by skipping forward past the stored pause frames as a frame series is read from memory.

[0055] FIG. 11 is a flow diagram illustrating identification and selection of pause frames for exclusion from a series of encoded voice frames. In particular, FIG. 11 illustrates techniques that may be used for identification and exclusion of pause frames for the condensation techniques described above with respect to FIGS. 8-10. As shown in FIG. 11, upon receipt of the next frame (65) in a series of encoded voice frames, the technique involves determination of the encoding rate associated with the frame (66).

[0056] The encoding rate indicates whether the frame contains a pause or speech. For example, vocoder 24 may encode frames at full rate, half rate, one-quarter rate, or one-eighth rate. Typically, vocoder 24 will encode pauses at one-eighth rate, permitting ready identification of pause frames. If the encoding rate of the frame is above a certain threshold (68), the frame is not a pause frame, and the process continues to consideration of the next frame (65). If the encoding rate is below the threshold (68), however, the frame is a pause frame. In this case, a pause length value is incremented (70). The pause length value represents the running length of a pause, as indicated by the number of consecutive pause frames identified in a speech sequence. Upon identification of a speech frame, the pause length value can be reset.

[0057] Using the pause length value, the technique further involves determining whether the number of pause frames is greater than a minimum number (72). Again, the minimum may be an absolute number of frames, or a dynamically calculated number that represents a minimum percentage of the frames in a pause. If the pause length is not greater than the minimum (72), the present pause frame is not excluded. Instead, the technique proceeds to consideration of the next frame. If the pause length is greater than the minimum (72), however, the technique proceeds to consideration of the next frame (74) for application of a last pause frame rule.

[0058] As discussed above, a last pause frame rule may require retention of the last pause frame in a consecutive series of pause frames to provide a current background noise measurement for decoding. Upon determining the encoding rate of the present frame (76) and comparing the encoding rate to the rate threshold (78), the technique determines whether the frame is a pause frame. If the frame is not a pause frame, as indicated by an encoding rate that is greater than the threshold, the previous frame was the last pause frame and must be retained. In this case, the process proceeds to the next frame.

[0059] If the frame is a pause frame, as indicated by an encoding rate that is greater than the threshold, the previous frame was not the last pause frame. Accordingly, the previous frame is excluded from the series of encoded voice frames (80), and the technique proceeds to increment the pause length value (70). From that point, the technique proceeds to consideration of the present frame in view of the minimum pause length (72) and last pause frame rules, and continues in like fashion for remaining frames in the series of encoded voice frames.

[0060] FIG. 12 is a flow diagram illustrating another technique for identification and selection of pause frames for exclusion from a series of encoded voice frames. FIG. 12 illustrates techniques that may be used for identification and exclusion of pause frames for the condensation techniques described above with respect to FIGS. 8-10. In contrast to the technique of FIG. 11, which generally involves exclusion of pause frames on a frame-by-frame basis, the technique of FIG. 12 illustrates exclusion of a group of pause frames. In particular, upon identifying a consecutive sequence of pause frames, i.e., by identifying the start and end of the pause frame sequence, the technique of FIG. 12 involves excluding a percentage of the pause frames.

[0061] As shown in FIG. 12, upon receipt of the next frame (82) in a series of encoded voice frames, the technique involves determination of the encoding rate associated with the frame (84). Again, the encoding rate indicates whether the frame contains a pause or speech. If the encoding rate of the frame is below a certain threshold (86), the frame is identified as a pause frame (88). The process continues to consideration of the next frame (82). If the encoding rate is above the threshold (86), however, the frame is not identified as a pause frame. In this case, the end of the pause sequence has been reached. In particular, when a non-pause frame is identified

following a sequence of pause frames, the technique detects the end of the pause sequence.

[0062] At this point, a percentage of the identified pause frames are excluded (90) from the series of encoded voice frames. If ten pause frames were identified, for example, and a reduction percentage of 80% were selected, then eight of the ten pause frames would be excluded. The process then continues with consideration of the next encoded voice frame (82). This technique may be accomplished, for example, by working through a sequence of encoded voice frames and buffering intermediate frames so that pause frames can be excluded from a final series of frames to be output, e.g., for buffering, transmission or playback.

[0063] The techniques described herein may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the techniques may be realized by a computer readable medium comprising instructions that, when executed, performs one or more of the techniques described above. In that case, the computer readable medium may comprise random access memory (RAM) such as synchronous dynamic random access memory (SDRAM), read-only memory (ROM), non-volatile random access memory (NVRAM), electrically erasable programmable read-only memory (EEPROM), FLASH memory, magnetic or optical data storage media, and the like.

[0001] The program code may be stored on memory in the form of computer readable instructions. In that case, a processor 16, such as a DSP, provided in a voice communication device 12 may execute instructions stored in memory in order to carry out one or more of the techniques described herein. In some cases, the techniques may be executed by a DSP that invokes various hardware components. In other cases, processor 16, modem 18 or vocoder 24 may be implemented as a microprocessor, one or more application specific integrated circuits (ASICs), one or more field programmable gate arrays (FPGAs), or some other hardware-software combination. Although much of the functionality described herein may be attributed to processor 16 for purposes of illustration, the techniques described herein may be practiced within processor 16, modem 18, vocoder 24, or a combination thereof. In addition, structure and function associated with processor 16, modem 18 and vocoder 24 may be integrated and subject to wide variation in implementation.

[0064] Communication media typically embodies processor readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport medium and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media, such as a wired network or direct-wired connection, and wireless media, such as acoustic, RF, infrared, and other wireless media. Computer readable media may also include combinations of any of the media described above.

[0065] Various embodiments have been described. These and other embodiments are within the scope of the following claims. For example, condensation techniques described herein may be performed within voice communication devices, such as cellular radiotelephones. Alternatively, the condensation techniques may be performed within network equipment responsible for forwarding packets containing the encoded voice frames, particularly for multicasting environments such as point-to-multipoint communication.

## CLAIMS

1. A method comprising:  
identifying encoded voice frames representing a pause; and  
excluding at least some of the identified frames to thereby produce a condensed series of encoded voice frames representative of a speech sequence condensed in playback time, the sequence exclusive of pauses associated with the excluded frames.
2. The method of claim 1, further comprising storing the condensed series of frames in a memory.
3. The method of claim 1, further comprising transmitting the condensed series of frames via a communication medium.
4. The method of claim 1, further comprising retrieving the condensed series of frames from a memory.
5. The method of claim 1, further comprising:  
comparing an encoding rate of the frames to a threshold; and  
identifying the frames representing a pause based on the comparison.
6. The method of claim 1, further comprising excluding only a portion of the identified frames from a consecutive sequence of the identified frames.
7. The method of claim 6, further comprising excluding a percentage of the identified frames from a consecutive sequence of the identified frames.
8. The method of claim 7, further comprising determining the percentage based on a minimum number of the identified frames needed for intelligible conversation.
9. The method of claim 6, further comprising determining a number of the identified frames from a consecutive sequence of the identified frames based on a minimum number of the identified frames needed for intelligible conversation.
10. The method of claim 1, further comprising retaining at least the last frame of a consecutive sequence of the identified frames in the series of frames.

11. A device comprising:
  - a voice encoder that generates encoded voice frames;
  - a processor that:
    - identifies encoded voice frames representing a pause; and
    - excludes at least some of the identified frames to thereby produce a condensed series of encoded voice frames representative of a speech sequence condensed in playback time, the sequence exclusive of pauses associated with the excluded frames.
12. The device of claim 11, further comprising a memory that stores the condensed series of frames.
13. The device of claim 11, further comprising a transmitter that transmits the condensed series of frames via a communication medium.
14. The device of claim 11, further comprising:
  - a memory that stores the condensed series of frames; and
  - a voice decoder that retrieves the condensed series of frames from the memory.
15. The device of claim 11, wherein the processor compares an encoding rate of the frames to a threshold, and identifies the frames representing a pause based on the comparison.
16. The device of claim 11, wherein the processor excludes only a portion of the identified frames from a consecutive sequence of the identified frames.
17. The device of claim 16, wherein the processor excludes a percentage of the identified frames from a consecutive sequence of the identified frames.
18. The device of claim 17, wherein the processor determines the percentage based on a minimum number of the identified frames needed for intelligible conversation.

19. The device of claim 16, wherein the processor determines a number of the identified frames from a consecutive sequence of the identified frames based on a minimum number of the identified frames needed for intelligible conversation.
20. The device of claim 11, wherein the processor retains at least the last frame of a consecutive sequence of the identified frames in the series of frames.
21. A machine-readable medium comprising instructions to cause a processor to:
  - identify encoded voice frames representing a pause; and
  - exclude at least some of the identified frames to thereby produce a condensed series of encoded voice frames representative of a speech sequence condensed in playback time, the sequence exclusive of pauses associated with the excluded frames.
22. The machine-readable medium of claim 21, wherein the instructions cause the processor to store the condensed series of frames in a memory.
23. The machine-readable medium of claim 21, wherein the instructions cause the processor to transmit the condensed series of frames via a communication medium.
24. The machine-readable medium of claim 21, wherein the instructions cause the processor to retrieve the condensed series of frames from a memory.
25. The machine-readable medium of claim 21, wherein the instructions cause the processor to:
  - compare an encoding rate of the frames to a threshold; and
  - identify the frames representing a pause based on the comparison.
26. The machine-readable medium of claim 21, wherein the instructions cause the processor to exclude only a portion of the identified frames from a consecutive sequence of the identified frames.
27. The machine-readable medium of claim 26, wherein the instructions cause the processor to exclude a percentage of the identified frames from a consecutive sequence of the identified frames.

28. The machine-readable medium of claim 27, wherein the instructions cause the processor to determine the percentage based on a minimum number of the identified frames needed for intelligible conversation.
29. The machine-readable medium of claim 26, wherein the instructions cause the processor to determine a number of the identified frames from a consecutive sequence of the identified frames based on a minimum number of the identified frames needed for intelligible conversation.
30. The machine-readable medium of claim 21, wherein the instructions cause the processor to retain at least the last frame of a consecutive sequence of the identified frames in the series of frames.
31. A machine-readable medium comprising a condensed series of encoded voice frames representing a speech sequence condensed in playback time, the condensed series of encoded voice frames omitting at least some encoded voice frames identified as representing pauses in a speech sequence.
32. The machine-readable medium of claim 31, wherein the condensed series of encoded voice frames excludes only a portion of encoded voice frames representing pauses in the speech sequence.
33. The machine-readable medium of claim 31, wherein the condensed series of encoded voice frames excludes a percentage of encoded voice frames from a consecutive sequence of the frames representing pauses in the speech sequence.
34. The machine-readable medium of claim 33, wherein percentage is based on a minimum number of the frames representing pauses that are needed for intelligible conversation.
35. The machine-readable medium of claim 31, wherein the condensed series of encoded voice frames retains at least the last frame of a consecutive sequence of the frames representing pauses.
36. A system comprising:

a first voice communication device having a voice encoder that generates encoded voice frames, a processor that identifies encoded voice frames representing a pause, and excludes at least some of the identified frames to thereby produce a condensed series of encoded voice frames representative of a speech sequence condensed in playback time, the sequence exclusive of pauses associated with the excluded frames, and a transmitter that transmits unexcluded frames; and

a second voice communication device having a receiver that receives the condensed series of frames from the first communication device, and a voice decoder that decodes the unexcluded frames for playback.

37. The system of claim 36, further comprising a memory within the first voice communication device that stores the condensed series of frames.

38. The system of claim 36, wherein the processor compares an encoding rate of the frames to a threshold, and identifies the frames representing a pause based on the comparison.

39. The system of claim 36; wherein the processor excludes only a portion of the identified frames from a consecutive sequence of the identified frames.

40. The system of claim 39, wherein the processor excludes a percentage of the identified frames from a consecutive sequence of the identified frames.

41. The system of claim 40, wherein the processor determines the percentage based on a minimum number of the identified frames needed for intelligible conversation.

42. The device of claim 39, wherein the processor determines a number of the identified frames from a consecutive sequence of the identified frames based on a minimum number of the identified frames needed for intelligible conversation.

43. The system of claim 36, wherein the processor retains at least the last frame of a consecutive sequence of the identified frames in the series of frames.

44. A device comprising:  
    means for generating encoded voice frames;  
    means for identifying encoded voice frames representing a pause; and  
    means for excluding at least some of the identified frames to thereby  
produce a condensed series of encoded voice frames representative of a speech sequence  
condensed in playback time, the sequence exclusive of pauses associated with the  
excluded frames.
45. The device of claim 44 further comprising a memory that stores the condensed  
series of frames.
46. The device of claim 44, further comprising a transmitter that transmits the  
condensed series of frames via a communication medium.
47. The device of claim 44, further comprising:  
    a memory that stores the condensed series of frames; and  
    means for retrieving the condensed series of frames from the memory.
48. The device of claim 44, wherein the identifying means compares an encoding rate  
of the frames to a threshold, and identifies the frames representing a pause based on the  
comparison.