Technical Disclosure Commons

July 18, 2019

Dynamic adjustment of audio playback speed

Jonathan Heinberg

Dynamic adjustment of audio playback speed

ABSTRACT

Many applications used for listening to audio provide mechanisms to adjust the speed of audio playback. Due to variation in speaking speeds of speakers in the audio, users often need to keep adjusting the playback speed during playback to achieve a desired speed or have to tolerate listening to the audio at a speed different from their preferred listening speed. Both situations result in a suboptimal user experience. This disclosure describes techniques to standardize the rate of speech in audio to improve a user's audio listening experience. The standardization of speech can be performed to achieve audio playback based on one or more listening parameters such as words per minute, speech speed multiplier, etc. The appropriate audio playback speed is then determined based on the metrics extracted from analyzing the audio. This enables the audio to be sped up or slowed down automatically during playback to provide the user-preferred audio listening user experience.

KEYWORDS

- Audio playback
- Playback speed
- Playback rate
- Podcast
- Audiobook
- Speech pattern

BACKGROUND

Many applications used for listening to audio provide mechanisms to adjust the speed of audio playback. As different individuals speak at different speeds, the desired adjustments in

audio playback speed for listening to audio can differ across speakers. Even the same individual can speak at different speeds at different times, thus requiring different adjustments to the audio playback speed at various points within the audio playback. Due to variation in speaking speeds of speakers in the audio, users often need to keep adjusting the playback speed during playback to achieve a desired speed or have to tolerate listening to the audio at a speed different from their preferred listening speed. Both situations result in a suboptimal user experience.

DESCRIPTION

This disclosure describes techniques to standardize the rate of speech in audio to improve a user's audio listening experience. The standardization of speech can be performed to achieve audio playback based on one or more listening parameters such as words per minute, speech speed multiplier, etc. To this end, with permission from the involved parties, the audio is analyzed to determine relevant metrics, such as word boundaries, speech speed, etc. The appropriate audio playback speed is then determined based on the metrics extracted from analyzing the audio. This enables the audio to be sped up or slowed down automatically during playback to provide the user-preferred audio listening user experience.

For instance, use of the described techniques to automatically adjust playback speed can result in audio playback being sped up for a segment of the audio that corresponds to a speaker who speaks slower than others while being slowed down for another segment that corresponds to when a speaker that speaks fast. Similarly, playback of the audio of the same person's speech can be slowed down or sped up during different sections to account for the variation in the person's speaking speed during the corresponding sections of the audio.
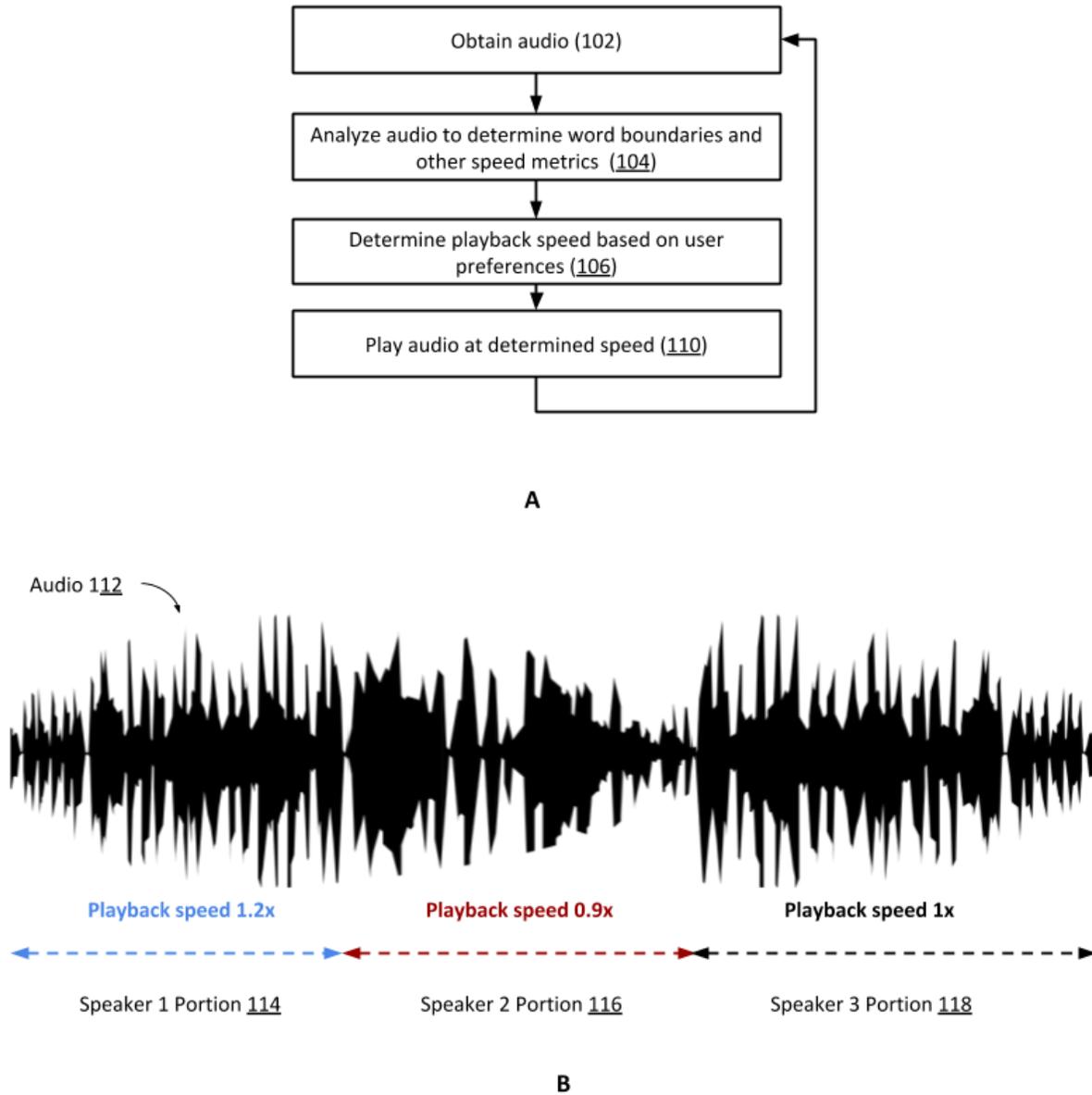
Fig. 1: Determining appropriate adjustments to audio playback speed

Fig. 1 shows an operational implementation of the techniques described in this disclosure with Fig. 1A (top) showing the steps involved and Fig. 1B (bottom) illustrating the application of the steps to a piece of audio (112). As illustrated in Fig. 1A, with permission of the user, the source audio (102) is analyzed to derive various metrics and parameters (104) relevant for determining appropriate audio playback speed. These metrics and parameters along with user-

specified playback and listening preferences, if any, are applied to generate cues that determine the audio playback speed (106) appropriate for various speakers and/or sections within the audio.

The cues are then applied to play the audio at corresponding playback speeds (106). For instance, as shown in Fig. 1B, the resultant cues can result in the audio being played at three different speeds: with portions corresponding to speakers 1, 2, and 3 within the audio (114, 116, and 118, respectively) being played at 1.2 times the default speed, 0.9 times the default speed, and at the default speed, respectively.

With user permission, the described techniques can be implemented by processing the entire audio prior to starting playback to determine the appropriate metrics and parameters required for adjusting the audio playback speed. Alternatively, or in addition, the techniques can be applied dynamically as portions of the audio are received, e.g., for streaming audio. The analysis of the audio can be performed at a server that hosts the audio and/or locally on the user device used for playback.

The playback cues determined based on the metrics and parameters can be transmitted along with the audio and/or embedded within the audio file for appropriate handling by the system used to playback the audio. Alternatively, or additionally, the appropriate playback adjustments corresponding to the determined cues can be applied prior to transmitting the audio.

The metrics and parameters used to determine the appropriate audio playback speed can be specified by the developers and/or can be configurable by the user. If the speakers permit, in cases where the identity of the speakers in the audio is known, appropriate playback adjustment cues can be determined by taking into account information based on previous playback of a speaker's audio and/or corresponding system-specified or user-configured playback settings for

the speaker. With permission of the speakers, speaker identities can be derived from information such as metadata associated with the audio, patterns extracted from voice recognition, etc.

The various operational facets described in this disclosure allow implementation flexibility such that they can be adaptive in various ways depending on the audio and/or the speakers involved. For instance, the audio processing can be performed remotely, locally, or a combination, and the appropriate playback adjustment can be determined based on any combination of metrics and parameters. The operation can be extended by incorporating various heuristics and/or audio algorithms in addition to the determined audio playback adjustment cues. Such operational adaptations can be applied across different audio sources and/or within the same audio file.

The techniques of the disclosure may be applied to a wide variety of audio including but not limited to podcasts, speeches, interviews, audiobooks, etc. Further, the mechanisms can also be applied to audio playback within any files that include an audio component, such as videos.

Application of the techniques described in this disclosure results in automatic and dynamic adjustment to the speed of audio playback, thus obviating the need for frequent manual adjustments to the playback speed. As such, the techniques provide an improved audio playback user experience personalized to the user's listening preferences.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed.

For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to standardize the rate of speech in audio to improve a user's audio listening experience. The standardization of speech can be performed to achieve audio playback based on one or more listening parameters such as words per minute, speech speed multiplier, etc. The appropriate audio playback speed is then determined based on the metrics extracted from analyzing the audio. This enables the audio to be sped up or slowed down automatically during playback to provide the user-preferred audio listening user experience. The described techniques can be used for playback of podcasts, speeches, interviews, audiobooks, etc. as well as audio playback within any content that includes an audio component, such as videos.