May 23, 2018

Seamless transfer of ambient media from environment to mobile device

Dominik Roblek

Matthew Sharifi

Seamless transfer of ambient media from environment to mobile device

ABSTRACT

Some mobile devices are capable of performing always-running, on-device music recognition. This feature enables users to obtain information about music or other audio playing in their environment. Typically, the detected ambient music is transient, e.g., once the song ends or the user leaves the current environment, the results disappear. However, a user may want to continue listening to the ambient music after leaving a location where it is being played. This disclosure enables seamless playback of identified ambient media on a user mobile device such that a user can continue listening or viewing such content after leaving the current environment.

KEYWORDS

- ambient audio
- music recognition
- media recognition
- audio transfer
- media transfer

BACKGROUND

Some mobile devices are capable of performing always-running, on-device music recognition. These features enable users to near-instantly obtain information about music playing in their environment, e.g., by glancing at the device's always-on display. Typically, the detected ambient music is transient, e.g., once the song ends or the user leaves the current environment, the results disappear. However, a user may want to continue listening to the ambient music after leaving a location where it is being played. In a similar manner, a user who

is watching a program on television may wish to continue viewing the program over a mobile device, even after leaving the television room.

DESCRIPTION

This disclosure provides techniques that enable mobile device users to transfer ambient music (or other media) seamlessly from the user's environment onto the mobile device. Transfer of ambient media to mobile device is done with user permission. For ease of interaction, such permission is obtained, e.g., at initial setup, and is modifiable, e.g., can be withdrawn at any time thereafter.

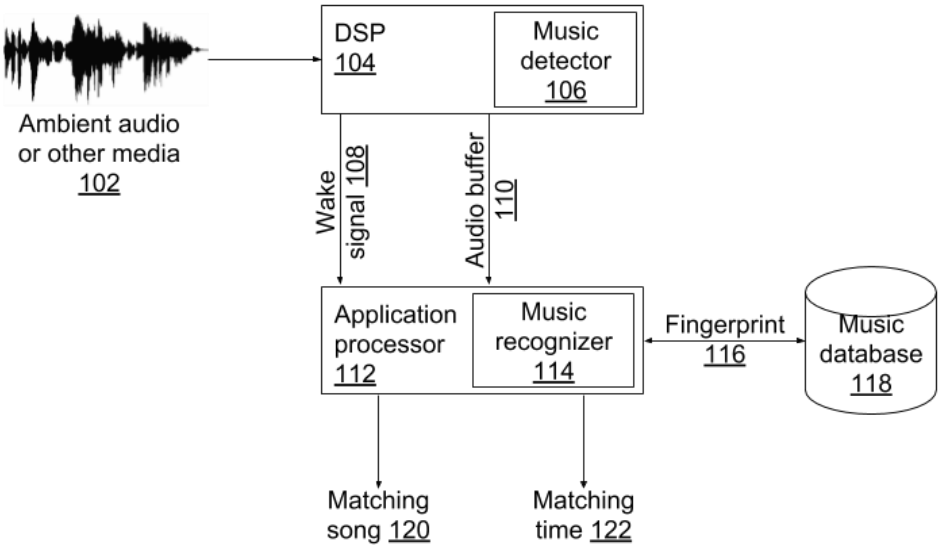

Fig. 1: Detection and recognition of ambient media

Fig. 1 illustrates detection and recognition of ambient audio or other media (102). A music detector (106) runs, e.g., continuously on a low-powered digital signal processor (104) that is part of a mobile device. For example, the music detector can be implemented using a neural network. The music detector detects the presence of ambient music or other media. Once ambient media is detected, the DSP sends a wake signal (108) and provides an audio buffer

(110) that contains audio samples to an application processor (112) or CPU of the mobile device.

A music recognizer (114) running on the application processor performs song recognition. In some implementations, the music recognizer can be implemented as a larger neural network that generates a fingerprint (116) of the ambient audio. The fingerprint is looked up in an on-device database (118). The result of the music recognition process is a matching reference song (120), a matching time (122) within the reference song where the match occurs, and other parameters, e.g., an estimate for rate difference between the ambient audio and the reference song. The reference song and time information is extracted based on a heat map, that captures the similarity between an embedding *i* in the probe (ambient) audio, time-offsets thereof, and embedding *j* in the reference (database-origin) audio, for all pairs *i* and *j*.

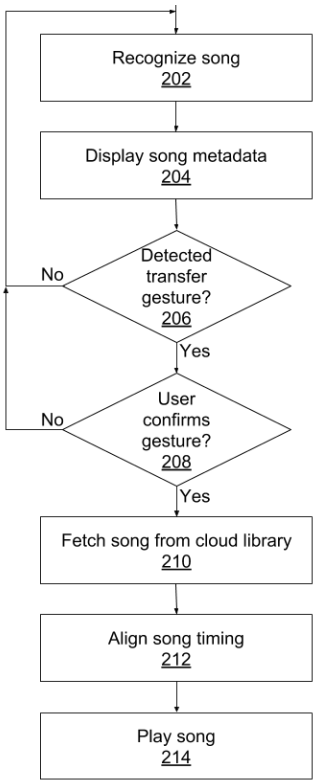

Fig. 2: Process to fetch media content

Fig. 2 illustrates an example process to seamlessly fetch media content from an environment onto a mobile device in a synchronized manner. With user permission to perform detection of ambient media (e.g., using an on-device microphone), an ambient song or musical piece is recognized (202) and song metadata, e.g., name, artist, etc., is displayed on the mobile device (204). User input, e.g., via gesture or other mechanism, is detected (206) that indicates that the displayed song is to be played on the device.

User gestures are detected and recognized with user permission. Example user gestures that indicate such user intent include, e.g., plugging in headphones into the headphone jack, inserting headphones with sensors into ears, squeezing the mobile device, touching a designated icon on the screen, etc. Optionally, a confirmation of the gesture (208) can be requested from the user. Upon receiving the confirmation, the device fetches the song from an available resource, e.g., a cloud library (210).

To account for delays between the ambient and fetched versions of the song, a time-alignment is performed (212) that accounts for the time elapsed between recognition and access from the cloud library. The time-alignment also accounts for any speed deviation between the probe (ambient version) and the reference (fetched version). For example, time alignment can be carried out by computing a binary cross correlation at the signal level. Playback of the song is then started (214) from the point at which it was last heard in the environment, with an appropriate time adjustment as determined in the time-alignment step.

The techniques of this disclosure can be used to detect and continue playback of ambient music or other media, e.g., speech, video, etc. For example, per the techniques, a radio show that a user was listening to while in a car can be continued on a mobile device after the user leaves the car. As another example, video content that was on the TV in a user's home can

be continued on a mobile device once the user leaves the TV room. In this manner, this disclosure provides a seamless way of transferring ambient audio or other media to a mobile device.

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

Some mobile devices are capable of performing always-running, on-device music recognition. This feature enables users to obtain information about music or other audio playing in their environment. Typically, the detected ambient music is transient, e.g., once the song ends or the user leaves the current environment, the results disappear. However, a user may want to continue listening to the ambient music after leaving a location where it is being played. This disclosure enables seamless playback of identified ambient media on a user mobile device

such that a user can continue listening or viewing such content after leaving the current environment.