February 26, 2018

Adaptive audio processing based on scene detection

Leonardo Kusumo

Moonseok Kim

Adaptive audio processing based on scene detection

ABSTRACT

This disclosure describes techniques to automatically adjust audio processing parameters based on acoustic scene detection. A parameter adjustment module utilizes default dynamic range compression settings during initial audio recording using a mobile device. A trained machine learning model is utilized to detect the acoustic scene, based on initial audio signals. Features such as the average loudness, maximum loudness, minimum loudness, acoustic scene, wind noise, and other user context features, as permitted by the user, are provided to the trained model. The model is utilized to adapt audio processing parameters such as the compression profile to the detected scene to improve audio recording quality.

KEYWORDS

- Dynamic range compression
- Audio recording
- Signal processing
- Audio engineering
- Outdoor recording
- Machine learning
- Home video

BACKGROUND

Mobile devices such as mobile phones, tablets, etc. are used for recording audio in a variety of settings. The different settings can include both indoor and outdoor settings, as well as settings with different audio levels (e.g., quiet settings, loud settings). Audio signal processing techniques such as dynamic range compression are employed to process audio signals to account for variations in the acoustic environment. Parameters for audio signal processing can be adjusted manually to improve audio recording quality and subsequent audio reproduction. Manual parameter settings are typically not available to users. Some recording devices provide users options to manually select the recording volume and aggressiveness of wind noise suppression.

It can be cumbersome for a user to manually adjust audio processing parameters when recording audio using a mobile device. Differences in various settings, e.g., indoor vs. outdoor, quiet vs. loud, etc. requires users to set appropriate processing parameters to ensure quality recording. Further, it is important that recordings from a variety of settings be playable on a user device with sufficient quality of audio output without the need for any audio editing or processing.

DESCRIPTION

This disclosure describes techniques to automatically adjust audio processing parameters based on detected acoustic setting(s) when recording audio using a mobile device. Implementation of the described techniques can provide a superior user experience when a mobile device is utilized to record audio in various settings such as an outdoor garden party or an outdoor conversation (quiet settings) as well as in environments such as live concerts (loud setting). Recording settings, e.g., audio processing parameters are automatically adjusted with minimal or no user input.

The techniques are implemented upon specific user permission to detect the acoustic environment, and apply suitable audio processing parameters. The techniques employ trained machine learning models and can be implemented within an audio recording application, other software applications, or in an operating system.

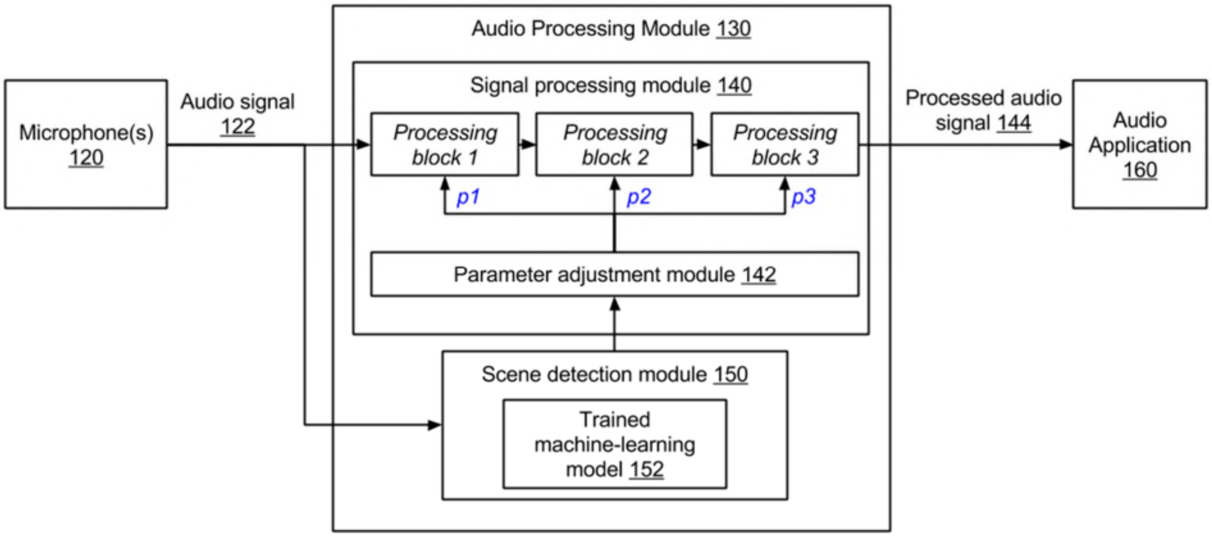

Fig. 1: Scene detection based automatic adjustment of audio processing parameters

Fig. 1 illustrates automatic adjustment of audio processing parameters based on scene detection. Audio signals (122) captured by one or more microphones (120) are provided to an audio processing module (130). The audio processing module includes a signal processing module (140) and a scene detection module (150).

A parameter adjustment module (142) is used for the computation of audio processing parameters (*p1*, *p2*, and *p3*) that are provided to one or more processing blocks (*processing block 1*, *processing block 2*, and *processing block 3*). While the figure illustrates three processing blocks, any number of processing blocks (*p1...px*) can be used in different use cases. The processing blocks can perform signal processing operations such as dynamic range compression (DRC) that can reduce the volume of loud sounds or amplify quiet sounds. The processing blocks can also perform adaptive background noise suppression (including wind noise), and beamforming, with knowledge of a specific location of interest of the capture.

The scene detection module is configured to analyze the audio signal(s), and detect the type of scene (or setting) that the audio signal(s) represent. The scene detection module utilizes a trained machine learning model (152) for acoustic scene detection. The model is trained to detect the start and end of a scene. With user permission, one or more features of audio signal(s) are provided to the trained model. Such features include, e.g., average loudness, maximum loudness, minimum loudness, acoustic scene, wind noise, user context features (e.g., the user is at an outdoor concert or walking in the park, or in a house, etc.). The model can also include a database of audio processing parameters (e.g., tuning or compression profiles) that are suited to a given detected scene. For example, if the audio recording is used for storing narration while video recording, audio beamforming paired with background noise suppression cam enhance the recording for the specific purpose.

The parameter adjustment module initially utilizes default processing settings (e.g., compression settings) that encompass a large dynamic range and that have a conservative profile that works for a majority of scenes. As the scene stabilizes, the trained model (152) detects the scene. Audio processing parameters such as the compression profile are then adapted to the detected scene based on the information provided by the scene detection module to the parameter adjustment module. For example, if it is detected that the scene is that of a concert, the audio processing parameters are tuned to optimize recording audio such that it covers the dynamic range of audio during the concert. If it is detected that the recording is being performed at a quiet scene, the audio processing parameters are tuned to a quiet setting, e.g., that increase the loudness of sounds below a threshold. The processed audio signal (144) is provided to an audio application (160).

Adaptation of the audio processing parameters to the detected scene using the techniques of this disclosure can improve recording quality and provide audio quality in mobile device recordings.

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 automatically adjust audio processing parameters in a mobile device based on acoustic scene detection. A trained machine learning model detects the acoustic scene based on received audio signals. Audio processing parameters such as dynamic range compression profiles are adapted to the detected scene and can provide superior audio recording quality.