May 06, 2019

Automatic adjustment of device volume based on context

Supraja Meedinti

Neil Dhillon

Automatic adjustment of device volume based on context

ABSTRACT

The output volume of a device is usually controlled as a fixed value between zero to a maximum. Unless manually adjusted by a user, the volume of the device output does not vary. In the case of speech output, such constant-volume responses can sound unnatural because human speech typically involves nuanced variations in volume that serve as subtle communicative signals. This disclosure describes techniques to dynamically and automatically adjust the volume of devices such as smartphones, smart speakers, smart appliances, etc. The adjustment is achieved by determining the appropriate level for output volume such that a fixed signal-to-noise (SNR) ratio is maintained. Further, if the user permits, parameters related to the user's context can be used to determine adjustment of the SNR to fit the user's mood. The techniques can be implemented via a trained machine learning model.

KEYWORDS

- Smart speaker
- Volume level
- Dynamic volume
- Volume adjustment
- Virtual assistant
- Voice UI
- Conversational interface
- Ambient noise
- Fixed signal-to-noise ratio (SNR)

BACKGROUND

User requests to a user device often include commands to change the settings of the device. For instance, a user may issue a command to change the volume of a device, such as a smartphone, smart speaker, etc. Volume is an important parameter for devices that provide output via audio. Such devices include smartphones, smart speakers, smart appliances, etc. that implement a virtual assistant. The output volume of the device is typically controlled as a fixed value between zero to a maximum. The value may be represented as an integer between these two extremes or set via a slider or other appropriate user interface (UI) mechanism used to control the volume level. Additionally, for devices that support voice-based interaction, the user can set the desired volume level by specifying it via a spoken command.

Once set, the volume level stays the same until the user makes further adjustments. Therefore, unless manually adjusted by the user, the output volume of the device does not vary when responding to the user. In the case of speech, such constant volume output can sound unnatural since human speech typically involves nuanced variations in volume that serve as subtle signals, such as meaning, emphasis, etc.

DESCRIPTION

This disclosure describes techniques dynamically and automatically adjust the volume of user devices such as smartphones, digital assistants, smart speakers, smart appliances, etc. The adjustment is achieved by determining the appropriate level for output volume such that a fixed signal-to-noise (SNR) ratio is maintained. Since noise levels are dependent on the environment, the approach of maintaining a fixed SNR results in the device volume varying according to external conditions. For instance, the noise in a loud environment is high, thus resulting in the device volume being raised proportionally to maintain a fixed SNR. In contrast, a quiet environment has low levels of noise, which causes the volume to be lowered accordingly for achieving a fixed SNR.

With permission from the user, an ambient noise sensor, such as a microphone, is used to monitor changes in the external noise levels of the environment in which the device is located. Upon changes in noise levels due to external events, e.g., the presence of a barking dog, active doorbell, approaching vehicles, background music, etc., the device volume is automatically and dynamically adjusted to maintain a fixed SNR. The volume adjustments occur even during the delivery of a voice response to the user.

Some devices such as smart speakers and other devices utilize server-based speech processing to analyze and respond to user commands, with user permission. On such devices, when a user interacts with a device via voice, the user's speech input is transmitted to a server that processes the speech. If the user permits, the server that processes the user's speech input can estimate the user's emotional state from aspects such as semantics, auditory properties such as pitch, vocal variation, etc. For instance, a user that speaks in a loud voice and uses words that match anger sentiments can be detected as potentially being in an angry mood. With permission from the user, server inferences regarding the emotional state derived from the user's speech can be provided to the device with which the user interacts.

If the user permits, the parameters related to the user's emotional state can be used to determine whether the SNR being maintained for determining the volume of the device output is to be adjusted to fit the user's mood. For instance, if the user is detected to be angry, SNR can be adjusted to deliver a lower volume to encourage calmness. If the user is detected to be sad, SNR can be adjusted to make the volume louder to convey friendliness and excitement. Such volume adjustments are equivalent to adjusting the tone of the voice response of the device.

The described techniques can be implemented via a trained machine learning model. With permission from the user, the model is provided with information on ambient noise, semantics of the user's speech, user's query history, etc. The output of the model indicates the adjustment to be made to the fixed SNR to determine the dynamic volume level for device audio playback.

The fixed SNR to be maintained for audio output can be determined dynamically, set by the device manufacturer, and/or specified by the user via any appropriate input mechanism, such as device buttons, app settings, voice commands, etc. Further, the fixed SNR can vary by user, such that responses to different users are delivered with different SNRs, each suited to the particular user. Moreover, the user can choose to turn off SNR based dynamic volume adjustment and revert to the default mechanism of manually controlling the volume.

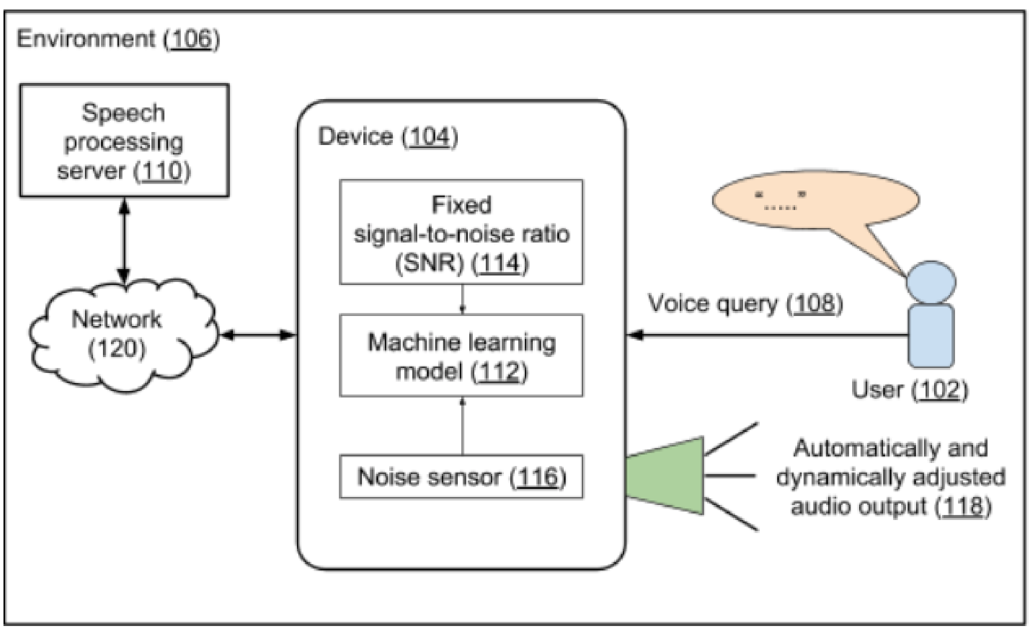

Fig. 1

Fig. 1 shows an operational example of the techniques described in this disclosure. A user (102) interacts via voice with a user device (104) located in an external environment (106). If the user permits, the user's query (108) is relayed by the device to a speech processing server (110) via a network (120). With user permission, the server provides a machine learning model (112) with various properties of the user's speech, including the user's inferred emotional state. The machine learning model is also provided with the level of ambient environmental noise, e.g., obtained via a noise sensor (116) of the device. The noise level is obtained with permission from the user. Based on external noise levels and properties of the user's speech, the machine learning model is used to determine adjustments to be made to the fixed SNR (114). The adjusted SNR is then applied to determine the volume level to deliver automatic and dynamic adjusted audio output (118).

As no additional hardware is needed to implement the described techniques, existing devices can be retrofitted with the functionality presented in this disclosure. Application of the described techniques makes manual adjustments of device volume unnecessary and improves the user experience of interacting with devices, especially devices that involve voice-based user interactions. Moreover, if the user permits, the techniques serve to deliver audio output that takes into account the user context. Such environmental immersion and contextual adjustments to volume levels can result in voice-based interactions that are closer to natural conversations among humans.

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 dynamically and automatically adjust the volume of devices such as smartphones, smart speakers, smart appliances, etc. The adjustment is achieved by determining the appropriate level for output volume such that a fixed signal-to-noise (SNR) ratio is maintained. Further, if the user permits, parameters related to the user's context can be used to determine adjustment of the SNR to fit the user's mood. The techniques can be implemented via a trained machine learning model.