December 19, 2018

Context-dependent output volume for voice-controlled virtual assistant

Victor Cărbune

Thomas Deselaers

Context-dependent output volume for voice-controlled virtual assistant

ABSTRACT

Voice-controlled virtual assistants typically maintain a standard output volume level. In certain circumstances, this leads to a response from the virtual assistant that is either too loud or too low. For example, if the last-used context of the virtual assistant was a party with high ambient noise, then the virtual assistant continues responding too loudly even after the party has ended and the ambient noise level has dropped.

This disclosure describes techniques to automatically adapt the output volume level of a virtual assistant based on current context, user preferences, user feedback, etc. A machine learning model predicts an optimum volume level for the virtual assistant, sets the volume level, and adapts it based on user feedback. With permission, user feedback serves as training data for predicting volume level in different contexts.

KEYWORDS

- virtual assistant
- smart speaker
- smart appliance
- assistant device
- volume control
- ambient volume
- machine learning
- reinforcement learning

BACKGROUND

Voice-controlled virtual assistants that provided output as audio via a speaker, e.g., implemented in consumer devices such as smartphones, wearable devices, smart speakers, home appliances etc., typically maintain a standard output volume level for responses provided to a user. The standard output volume level can sometimes lead to the response from the virtual assistant being too loud or too low. For example, a virtual assistant device such as a smart speaker at a party with a relatively high ambient noise plays music at a correspondingly high volume. However, after the party ends and the ambient noise subsides, a user request for a weather update results in the virtual assistant responding at the previously set and now inappropriately high volume, thereby providing an unsatisfactory user experience.

DESCRIPTION

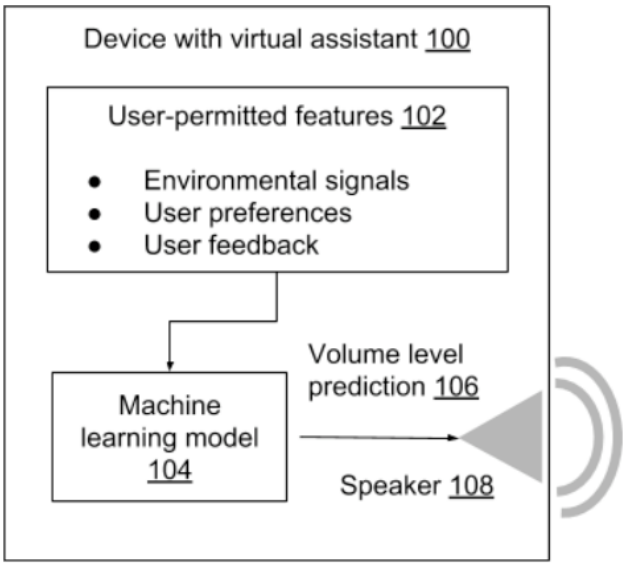

Fig. 1: Adjusting volume based on user context

Fig. 1 illustrates adjusting the volume level of a device with a virtual assistant (100) based on context, per techniques of this disclosure. With user permission, a machine learning

model (104) is provided input features (102) including, e.g., environmental factors; user preferences, e.g., user-preferred volume level; user feedback; etc. Based on the input features, the model provides a volume level prediction (106) which is used for output of the virtual assistant response by a device speaker (108). Users are provided with options to turn off such volume level prediction or to limit the use of automatic volume level determination techniques to contexts when certain conditions are met.

Environmental factors can include, e.g., ambient sound level, day of the week, time of the day, user's calendar, location of the user's house, location of the user within the home, etc. Only such factors are used as permitted by the user. Users are provided with options to selectively enable/disable use of individual factors, or to turn off use of environmental factors entirely. The permitted factors are used specifically for volume level prediction. For example, data from the user's calendar, if permitted, enables the machine learning model to know of times, e.g., party times, quiet hours, etc., during which volume levels deviate from the norm. A factor such as the location of the user's house enables the virtual assistant to conform to local sound regulations and restrictions.

If permitted by the user, the machine learning model takes into account user feedback using, e.g., reinforcement learning (RL). Under RL, the machine learning model regulates the volume automatically. An adjustment of the automatically selected volume by the user is an indication that the predicted volume level was likely inaccurate, and the volume level selected by the user is an indication of the volume level desired by the user. An automatically-set volume that is left undisturbed by the user is taken as a positive reinforcement signal that the predicted volume level matches user needs.

The machine learning model can also be trained using supervised learning, e.g., by taking the user-configured volume as ground truth data. Under supervised learning, sampled data is obtained, with user permission. For example, sample data can be in the form {current contextual features, absolute volume}. A map is generated between contextual features and volume based on the ground truth data. During operation, if a predicted volume level is associated with a high confidence score, the virtual assistant output is automatically provided at the predicted volume level. Alternatively, the difference between predicted and current volume level is suggested to the user as an adjustment.

The machine learning model addresses the regression problem using, e.g., a neural network, a decision tree, etc. Example types of neural networks that can be used include long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc., can also be used. Heuristics can also be used to predict the output volume level.

In this manner, the techniques described herein enable virtual assistant output via audio to be provided at an appropriate volume level based on user context and preferences. User feedback is utilized to train the machine learning model to learn the appropriate volume level. With user permission, the described techniques can be implemented in any user device that provides virtual assistant output via audio, by a server that provides the virtual assistant service, etc.

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 adapt the output volume level of a virtual assistant based on current context, user preferences, user feedback, etc. A machine learning model predicts an optimum volume level for the virtual assistant, sets the volume level, and adapts it based on user feedback. With permission, user feedback serves as training data for predicting volume level in different contexts.