May 2023

Personalizing Audio Content Played While On Hold During a Phone Call

Joseph Johnson Jr

Emmanouil Koukoumidis

Shiblee Hasan

Personalizing Audio Content Played While On Hold During a Phone Call

ABSTRACT

Calls made to a business by a customer, e.g., to request support, are often put into a queue waiting for a human agent to be available. During the hold time, canned music or other audio is played back to the caller. Such audio is low quality owing to the limited capacity of the telephony channel, is not personalized, and repeated multiple times till an agent becomes available, providing an unsatisfactory calling experience. This disclosure describes the use of machine learning techniques to detect canned audio and replace it with high fidelity music or other content. With user permission, the replacement content can be personalized, e.g., based on a user's music playlists/preferences, and context. Machine learning techniques can also be utilized to upscale music on hold experience provided by the business. With user permission, advertising content or helpful content about the business can be delivered during the hold time. The techniques can be integrated into a virtual assistant or device operating system to provide an improved calling experience.

KEYWORDS

- Call center
- Human agent
- Music on hold
- On-hold audio
- Customer service
- Wait time
- Personalized audio

BACKGROUND

When making phone calls to seek services, people are often put on hold while they are in line waiting for a human agent to become available to talk to them. During the wait time, low fidelity music, generic announcements, or advertisements are played on the call. Such content is sometimes played at a high volume and repeats in a loop until a hold ends.

Some businesses attempt to improve the on-hold audio experience by personalizing the audio content. For instance, for users who have provided appropriate permissions and integration with their online music collections, music from their own collections can be streamed while they wait on hold. However, the resolution of the audio content played over a phone call is substantially lower than that of lossless CD-quality audio or other high fidelity audio. Even in cases where high fidelity audio content is played during a period of hold, such audio can be affected by skips and quality degradation, especially when the caller is in motion and switching between cell phone towers.

Alternatively, some phone applications (e.g., [1]) include the ability for users who are put on hold to receive a notification when a human answers at the end of the hold. Until notified of the end of the hold, users can switch to other tasks, including playing their own music locally while waiting for the hold to end. However, this feature may sometimes fail to provide periodic feedback that the call is ongoing, and the user is still on hold. In addition, such a feature is not seamlessly integrated into the on-hold user experience (UX) since activating the feature requires users to take explicit manual action. Moreover, the feature may not be universally available on all devices or phone applications. Further, users who do not wish to switch to another task and wait for a notification may opt not to use the feature and choose to stay with the on-hold audio experience. None of the existing approaches to improve the user experience of listening to on-hold audio content provide a seamless way for users to receive personalized content and/or advertisements at high fidelity.

DESCRIPTION

This disclosure describes the application of machine learning to play personalized audio content at high fidelity while a user is waiting on hold during a voice call. With appropriate permissions from the user, the techniques can be integrated into a virtual assistant by handing the call off to the virtual assistant to play high fidelity personalized audio content during hold times. Users can move around freely without being tied to the phone until notified by the virtual assistant that a human agent has joined the call.

The personalized audio stream can include any suitable content, such as music, spoken content, advertisements, announcements, etc. For instance, audio advertisements played when the user is on hold can be personalized with permission to those that are relevant to the user's tastes. Moreover, such advertisements can be offered in exchange for providing users with relevant services, such as paying for a portion or all of the telephone bill, providing a free service upgrade, offering a trial for a paid subscription, etc.

If users permit, the personalization can be based on detecting emotion or sentiment in their voice during the call. For instance, if a user is detected to be in a bad mood, the on-hold audio can include the types of music that the user prefers when in a bad mood and/or the kind of music that is likely to help bring the user out of the bad mood.

The start and end times of canned audio (boilerplate audio content) played while a call on hold can be detected via any suitable, trained machine learning model, e.g., such as sample-based or video-based deep convolutional neural network (CNN). The input to the classifier can be the waveform of the audio content and/or other relevant features, such as Mel-Frequency Cepstral

Coefficients (MFCCs), as a two-dimensional matrix over time. The output of the classifier indicates whether the input audio is boilerplate content. Boilerplate audio can be considered to have ended once the classifier output indicates a lack of boilerplate audio content for a threshold number of audio frames.

Once the start and end points of boilerplate audio content are determined, the audio can be replaced with any other audio stream, such as personalized music and/or advertisements. Further, suitably trained machine learning models can be implemented on the user device to improve the quality of the received streamed audio. For instance, the streamed audio can be synthesized and stylized locally on the user device to make it sound less compressed and lossy by employing generative adversarial networks (GAN), diffusion models, etc. Alternatively, if users permit, only relevant metadata can be streamed to enable selecting audio content stored locally on the user device (or otherwise accessible via the user device). Such audio can be retrieved and played at high fidelity without the bandwidth limitations of the call audio stream.

If the user permits, the audio content can be personalized to the user based on a user-permitted identifier such as a user device identifier, online account login status, caller ID, voice fingerprint, email address associated with the phone number, etc. The user identification can be used to obtain audio content relevant to the user from services such as cloud-based music catalogs, playlists within music streaming services etc. Audio advertisements can be personalized based on user-permitted data regarding user activities and preferences. If the user permits, advertisements can be additionally personalized based on relevant parameters of the phone call during which the user is on hold. For example, if the user is calling a particular vendor or seeking support for a specific product, the personalized advertisements can be regarding shopping recommendations for that vendor or accessories for that product.

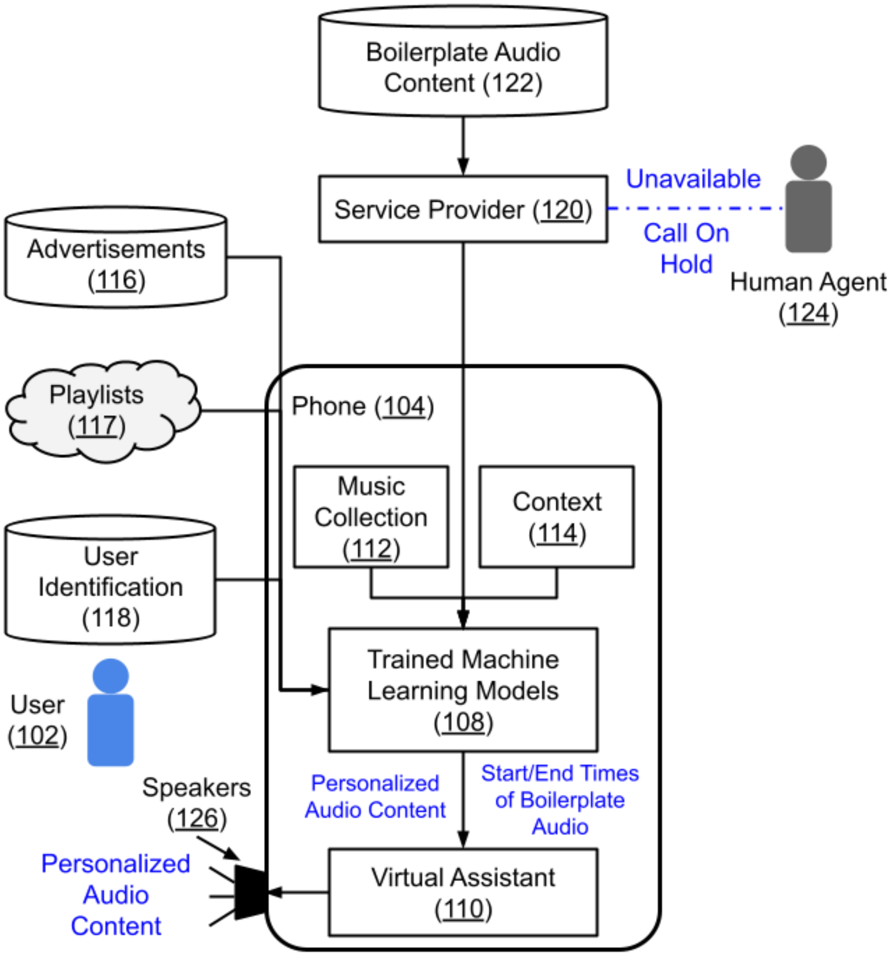

Fig. 1: Replacing boilerplate on-hold audio content with personalized audio content

Fig. 1 shows an example operational implementation of the techniques described in this disclosure. A user (102) makes a voice call to a service provider (120) using a phone (104). Since no human agent (124) is available to talk to the user, the user's call is put on hold and the user is streamed boilerplate audio content (122) during the hold.

With the user's permission, trained machine learning models (108) are employed to process the on-hold audio and detect start and end times of the boilerplate audio. If the user permits, the user can be identified (118) via any appropriate technique and based on user-permitted data. Once the user is identified, the user's online playlists (117) are obtained to generate personalized audio content from the playlists and/or the local music collection (112) on the device. The audio content can additionally include audio advertisements (116) personalized based on the user's activities, preferences, and context (114). The user can choose to hand the call off to a virtual assistant (110) to play the personalized audio content via the device speakers (126) or other output device until a human agent becomes available and ends the hold.

If the on-hold audio content includes long gaps with no audio, the start and end times can be determined based on applying appropriate smoothing by marking the continuous period of time from start to end over a threshold number of seconds. In many cases, the background on-hold audio content is periodically paused or faded to deliver instructions or updates, such as "Please continue to hold. You are currently the fifth caller in line." As appropriate, such repeated messages can be delivered immediately by interrupting the personalized audio content or queued and delivered later at an opportune time, such as the end of a song or advertisement. Alternatively, or in addition, repeated messages can be suppressed by delivering such messages only once or at a low frequency. For instance, instead of repeating a message such as "Please continue to hold" multiple times a minute, it can be played once every two minutes. Further, with user permission, the message can be stylized to be less generic and more pleasant, such as "This is a reminder that you are still on hold." If the periodic messages contain informative external pointers, such as "Visit the returns page on our site for our returns policy," the link to the resource can be sent to the user device and/or saved for generating a text message or notification popup for reviewing after the call.

Personalized advertisements inserted during the hold times can be pre-recorded advertisements specifically for delivery via audio, such as advertisements played on radio.

Alternatively, the advertisements could be the audio stream of video advertisements or audio generated by applying text-to-speech (TTS) to text advertisements.

The techniques described herein can be implemented within any device or application used for making phone calls and integrated into any virtual assistant application and corresponding devices, such as smart speakers that can make calls to businesses. Further, the techniques can be made available as a service to relevant parties, such as call center operators, businesses that provide phone support, etc. If users permit, the service can provide post-call features such as recording the call as an audio file for later reference, generating a text summary of the call and sending it as an email or text message, etc. Further, search features can be provided to retrieve specific content of interest within the audio recordings or text summaries. For instance, users may want to retrieve specific instructions or information provided during a call. Data used to establish user identity is encrypted to ensure that the personalization is performed securely. Appropriate thresholds can be selected to determine the end of the boilerplate audio content.

Implementation of the techniques described in this disclosure can make the experience of being on hold during phone calls seamless and enjoyable by replacing generic, low fidelity on-hold audio content with high fidelity personalized audio content. Further, the personalized audio minimizes interruptions from repeated instructions while ensuring that users still receive adequate call-related information at appropriate times and with appropriate frequency. Moreover, integration into a virtual assistant can enable users to engage in other tasks while waiting on hold, thus minimizing the productivity lost while waiting on hold for a human agent to join the call. Replacing generic advertisements within the on-hold audio with user-permitted personalized advertisements can provide greater business value by better targeting advertisements while keeping the amount of advertising received by users the same. The described techniques can be incorporated into any virtual assistant application, operating system, call center software suite, content streaming services, 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 identity, accounts with music and content streaming services, a user's context, social network, social actions or activities, 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 the use of machine learning techniques to detect canned audio and replace it with high fidelity music or other content. With user permission, the replacement content can be personalized, e.g., based on a user's music playlists/preferences, and context. Machine learning techniques can also be utilized to upscale music on hold experience provided by the business. With user permission, advertising content or helpful content about the business can be delivered during the hold time. The techniques can be integrated into a virtual assistant or device operating system to provide an improved calling experience.

REFERENCES

1. "Use Hold for Me - Google Assistant Help" available online at https://support.google.com/assistant/answer/10071878?hl=en accessed April 30, 2023.
2. "Great telephone on-hold music is great customer service!" available online at https://www.yummy-sounds.com/telephone-on-hold-music/ accessed April 30, 2023.