Technical Disclosure Commons

Defensive Publications Series

July 03, 2017

Dynamic Context-Based Voice Modulation

John D. Lanza
Foley & Lardner LLP

John D. Lanza
Foley & Lardner LLP

DYNAMIC CONTEXT-BASED VOICE MODULATION

ABSTRACT

When audio sponsored content is presented with non-sponsored audio content in a two-way conversation between a user and a respective client computing device, it may be challenging for the user to distinguish the sponsored audio content from the non-sponsored audio content. The client computing device can receive an input audio signal indicative of a respective uttered speech. A data processing system communicatively coupled to the client computing device can process the received audio signal to identify a request corresponding to the uttered speech. The data processing system can generate a response to the identified request, and select an ad based on a context of the identified request. The data processing system can apply one or more sound/audio effects to audio content of the selected ad. The applied sound/audio effects can help audibly discern the ad content from the generated response content. The client computing device can play the audio content of the generated response and the audio content of the ad with the applied sound/audio effects.

DETAILED DESCRIPTION

Voice-based virtual personal assistants provide users with a voice-based interaction modality. While users may still interact with virtual personal assistants via a keyboard, the primary interaction modality with the virtual personal assistants is voice. In particular, the voice-based virtual assistant can engage in two-way audio conversations with users. While these voice capabilities facilitate user interactions with client devices and computer systems, voice interaction can impose technical difficulties related to user experience. In particular, providing

content as audio output can make challenging for users to distinguish between various types or various segments of the output audio content.

When interacting with virtual personal assistant, a user can utter words indicative of a search query to request content related to a product or service category. In responding to the user's request, the virtual personal assistant can provide as audio output a list of search results and one or more ads related to the search query. Also, the user can request a specific service such as a restaurant reservation, plane ticket reservation, a ride service, or some other type of service. Before or while executing the user's requested service, the virtual personal assistant t may present to the user an ad indicative, for example, of an alternative provider of the requested service. When an ad is presented on a webpage or other information resource (e.g., a mobile application, a gaming platform, a social media platform, or geographical map viewport), the user can distinguish the ad from primary content of the webpage or resource based on visual cues of the ad such as the location of the ad within the webpage, the formatting of the ad, or a logo of the corresponding advertiser. However, in a two-way audio conversation with the virtual personal assistant, the user may find it challenging to distinguish sponsored content from primary requested content within audio output. Such challenge may result in confusion on the part of the user. For example, interpreting the sponsored content as a failure on the part of the virtual personal assistant to recognize the user's input speech, the user may unnecessarily repeat his request and may even get irritated. Also, the user may be confused with regard to how to proceed with the conversation.

To improve user experience and facilitate distinction between sponsored audio content and non-sponsored audio content, systems and methods described herein allow for applying audio effects and/or sound effects to sponsored audio content before output to a user. In

particular, the present disclosure is generally directed to a data processing system and one or more client devices for dynamically applying audio effects and/or sound effects to sponsored audio content. The data processing system, or a client device communicatively connected to data processing system, can insert additional audio (e.g., tones, rings or background audio), modify audio content associated with an ad, for example, to generate echo, filtering, equalization, phaser, time stretching, compression, flanger, robotic voice effects.

By applying audio or sound effects to sponsored audio content, the data processing system or the client device can cause sponsored audio content output to user to be more easily discernable from non-sponsored content. Also, the data processing and/or the client devices can provide a user interface (UI) to allow user customization of audio or sound effects to sponsored audio content. For instance, the UI can display the various available options for audio and/or sound effects that the data processing system (or the client device) is capable of applying to audio sponsored content. Each option of audio/sound effect or a combination of audio and/or sound effects option can be represented in the UI via a selectable item. The user of the client device can select one of the effects or combinations of effects listed in the UI. In response to such selection, the data processing system (or the client device of the user) can store the user's selection in a user profile, and apply the selected effect or combination of effects to each audio ad presented to the user by the respective client device. In some implementations, the UI may allow for multiple selections by the user. For example, the user may make a different selection for each type or category of ads or each presentation context. The presentation context can be defined based on the type of request made by the user (e.g., service requests versus search queries), the time of day of presenting the ad to the user, the location of the user (e.g., in a car driving, at work, etc.), or other criteria.

FIG. 1 is an illustration of an example method 100 for dynamically applying audio/sound effects to audio ads presented to a user during a two-way conversation between the user and a corresponding client device. The method 100 can be performed by a data processing system, a client computing device or both. The method 100 can include receiving an input audio signal indicative of speech uttered by a user of a client computing device (step 105). The client computing device can receive the input audio signal via a respective microphone. The client computing device may include a natural language processor component for processing speech signals. The client computing device may transmit the received audio signal to the data processing system for processing by a natural language processor component executed by the data processing system.

At step 110, the method 100 can include the data processing system (or the client computing device) processing the input audio signal. The natural language processor component can machine-translate the audio signal into a corresponding text signal. If the machine translation is performed at the client computing device, the client computing device can transmit the generated text signal to the data processing system. The data processing can parse the text signal to identify one or more trigger keywords defining a specific request by the user of the client computing device. For example, if the machine translated audio signal includes “I need a ride from Taxi Service Company A to go to 1234 Main Street.” The data processing system can identify the trigger keyword(s) “ride” or “need a ride” as indicative of a request for a ride service. The data processing system can also interpret the “Taxi Service Company A” following the keyword “from” as indicative of the desired provider of the requested service, and interpret what follows “go to” as indicative of the desired destination.

The data processing system can determine a type of service or product requested by the user based on the processing of the input audio signal. In some implementations, the data processing system can identify a search query responsive to processing the input audio signal. The data processing system may use trigger keywords like “ride” together with “I need” or command verbs to distinguish a request for a service from a search query.

At step 115, the method 100 can include the data processing system generating a response to the user’s request. For example, if a search query was identified responsive to processing the input audio signal, the data processing system can generate a list of search results corresponding to the search query as a response to be provided by the user. If the end user of the client computing device requested a taxi from Taxi Service Company A, the data processing system can generate a statement requesting the user to confirm that he is asking for a taxi service from Taxi Service Company A to the destination 1234 Main Street. The data processing system can generate the response as audio content. The data processing system may generate a textual response and then convert the textual response to a corresponding audio signal.

At step 120, the method 100 can include the data processing system selecting an ad based on the context of the user’s request. For instance, the data processing system can select the ad based on keywords identified from the input audio signal. For example, the data processing system can identify, based on the trigger keyword “ride,” one or more ads associated with ride service providers. The data processing system can select the ad from a content provider that is different than the service provider requested by the user of the client device in the input audio signal. If the input audio signal is indicative of a search query, the data processing may identify one or more ads relevant to the search query. The data processing may identify only audio ads or ads including audio contents. The data processing system may also identify text ads since the

textual content of such ads can transform into corresponding audio content. The data processing system may run an auction for the identified ads and select one or more ads for presenting to end user of client device based on the auction result(s). If a selected ad is not an audio ad, the data processing system can convert textual content associated with the selected ad to corresponding audio content.

At step 125, the method 100 can include applying one or more audio/sound effects to an audio signal associated with selected ad. The audio/sound effects can be applied at the data processing system or the client computing device. Applying the audio/sound effects can audibly discern the audio signal associated with the selected ad from the audio signal corresponding to the response generated by the data processing system. Applying the audio/sound effects can include inserting one or more tones or rings at the start of and/or at the end of the audio signal associated with the selected ad. As such, the added tone(s) or ring(s) can represent audio indications of the start and/or end of the audio content associated with the ad. Applying the audio/sound effects can include adding specific background audio content (e.g., music, ringtones, background audio noise, sound of water flow, sound of birds, or the like) to the audio signal associated with the selected ad. The added background audio can have a time duration substantially equal to that of the audio signal associated with the ad. The added audio background can have amplitudes relatively smaller than amplitudes of the ad audio signal.

The data processing system (or the client computing device) may modify the audio signal associated with the ad. For example, the data processing system may introduce an echo effect by merging the ad audio signal with a delayed version of the same ad audio signal. The data processing system may induce a flanging effect by mixing two identical versions of the ad audio signal where the second version is delayed by a small and gradually changing period. The data

processing system may induce a phaser effect by mixing two versions of the ad audio signal with different phase values. The data processing may apply an equalization effect to the ad audio signal by attenuating or boosting different frequency bands of the ad audio signal differently. The data processing may apply audio filters (e.g., a low-pass filter, a high-pass filter, a band-pass filter, a band-stop filter, or a combination thereof) to the ad audio signal. The data processing system may apply time stretching to the ad audio signal by changing the play speed of the ad audio signal. The data processing system may apply a compression effect to the ad audio signal by modifying the frequency or amplitude of a carrier signal (e.g., an envelope) of the ad audio signal. The data processing system may apply a robotic voice effect to the ad audio signal so that the latter sounds like a synthesized human voice (e.g., from multiple available synthetic human voices). If the selected ad is a text ad, the data processing system can select a synthetic human voice (or robotic effect) convert the text ad into corresponding audio content according to the selected robotic effect.

Other sound/audio effects can include pitch shift, pitch increase or decrease, signal modulation, or other sound/audio effects known in the art of speech and audio processing. The applied sound/audio effects may be designed to be heard/noticed (by the client computing device user) without destroying the audible characteristics of the modified ad audio signal. That is, the client computing device user can still recognize the words spoken as part of the ad. The data processing system may apply a combination of audio/sound effects. Also, as discussed above, the audio/sound effects applied may be customized by the user of the client computing device through selection from a UI.

At step 130, the method 100 can include the client computing device playing the audio signal corresponding to the response generated by the data processing system and the ad audio

signal with the applied sound/audio effects. Since the sound/audio effects are applied to the ad audio signal only, the user of the client computing device can distinguish, based at least on the sound/audio effects, between audio content associated with generated response and the audio content associated with the ad. In some implementations, different audio/sound effects may be applied to the response and the ad.

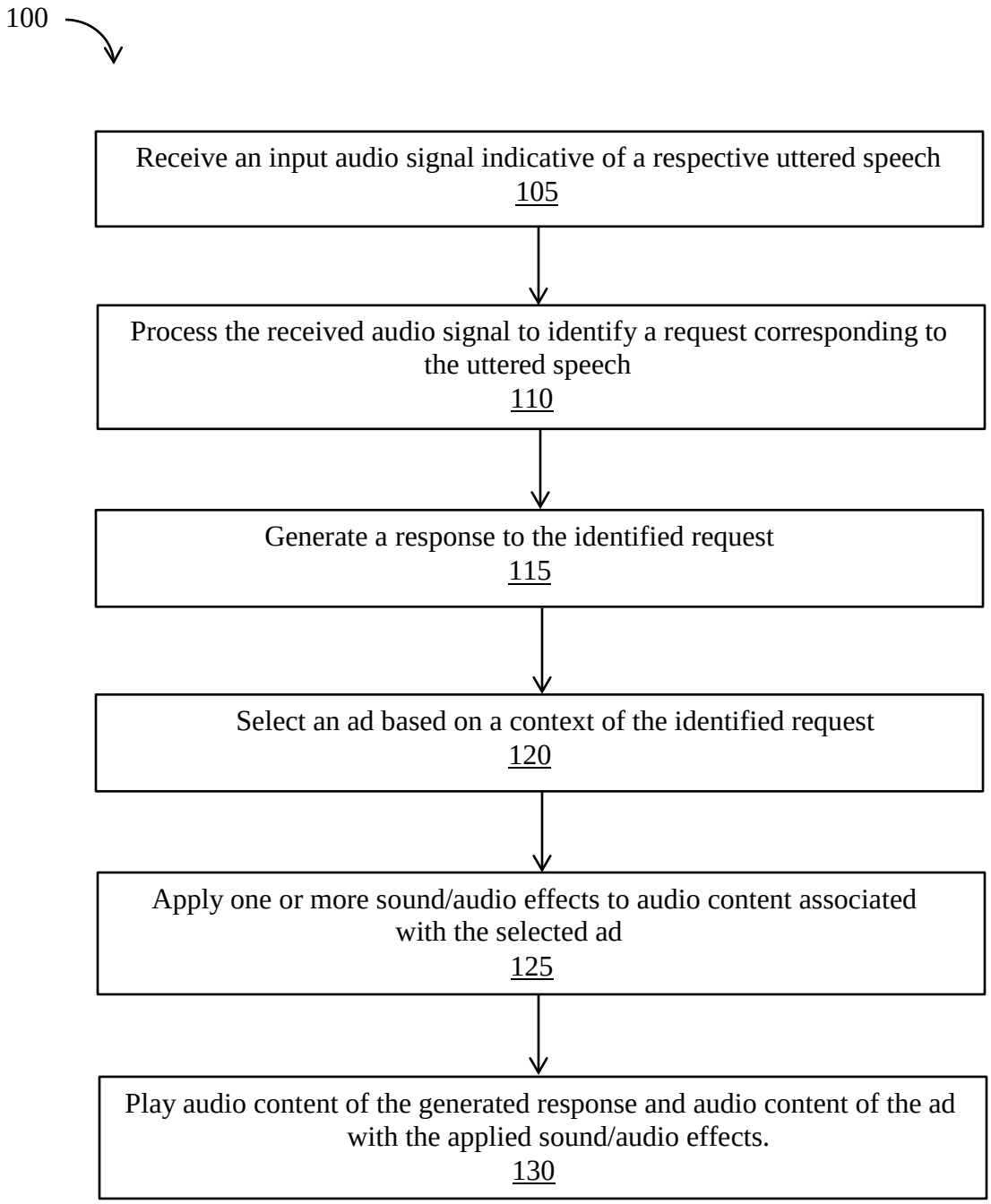
100
Receive an input audio signal indicative of a respective uttered speech
105
Process the received audio signal to identify a request corresponding to the uttered speech
110
Generate a response to the identified request
115
Select an ad based on a context of the identified request
120
Apply one or more sound/audio effects to audio content associated with the selected ad
125
Play audio content of the generated response and audio content of the ad with the applied sound/audio effects.
130

FIG. 1