July 03, 2017

# Cost Per Second (CPS) Ads On Webpages

John D. Lanza
*Foley & Lardner LLP*

John D. Lanza
*Foley & Lardner LLP*

# COST PER SECOND (CPS) ADS ON WEBPAGES

## DETAILED DESCRIPTION

In a voice-based interaction computer environment, users can engage in audio conversations with respective client devices. For instance, a client device implementing a front-end module of a virtual personal assistant can allow for two-way conversations with the respective user. Such conversations usually relate to user requests to be fulfilled by the client device and a data processing system remotely connected to the client device. The context of the conversations (i.e., user requests) and their nature (i.e., voice-based) suggest a relatively high user attention to audio output from the client device. For instance, when the user initiates a conversation to request an online service or an online action to be performed, the user will be paying attention to the response of the client device.

Considering the relatively high user attention to audio output from the client device and the known user intent (e.g., service or action request), the audio conversations between the user and the client device provide potentially high value advertisement opportunities. In particular, during a given conversation, providing an ad that is relevant to the user's request has a high likelihood of engaging the user to interact with the ad and/or execute an ad related transaction (i.e., conversion). The high advertising value of the two-way conversations between users and respective client devices calls for optimization of the selection, presentation and pricing of ads by the virtual personal assistant. In particular, the relevance of a selected ad to the context of the conversation and the timing of the presentation of the selected ad within user conversation with the virtual personal assistant can enhance user experience and increase advertisers' satisfaction. Also, given the relatively high likelihood of user engagement with ads during conversations with the virtual personal assistant, advertisements may be priced per unit time.

In the current disclosure, a data processing system can receives an input audio query, and identify a user request based on the received audio query. Based on the context of the user request, the data processing system can determine a thread of tasks or operations. For example, the thread can indicate a sequence of tasks or operations including a first task followed by a second task, and a third task that is subsequent to the second task. The data processing system can determine or estimate, for each identified task, a respective time segment. The data processing system can select an ad based on at least one of the identified tasks and the corresponding time segments. The selected ad can include an audio signal to be played to the user of the client device. The data processing system can determine a presentation time (or presentation order) of the selected ad based on time segments of the identified tasks.

FIG. 1 is a flowchart illustrating an example method 100 for providing ads in a voice-based interaction computer network environment. At step 105, the method 100 can include identifying a user request in a received audio query. The audio query can be generated by a user of a client device and can be indicative of a request or command made by the user. The user can initiate a conversation with an instance of a virtual personal assistant (e.g., Google Assistant) executing one a respective client device and/or a data processing system remotely connected to the client device. The user can make a request for an online service or action (e.g., an audio search query, online purchase of movie tickets, rendering of an online live stream, scheduling of a cab service, etc.). The client device can receive the audio query (i.e., as an audio signal) via a respective microphone, and transmit the received audio query to the data processing system. A natural language processor (NLP) component of the data processing system can process the audio query to identify the user request. The NLP component can machine-translate the audio query to a corresponding text and parse the generated text to identify one or more keywords. For

example if the user says “OK, I would like to go to go dinner and then a movie tonight,” the NLP component can identify the keywords “go to,” “dinner” and “movie.” Based on the identified keywords, the data processing system can identify the context user’s request to be a search query, or a request for reservation, for a restaurant to have dinner and a movie to watch in a movie theater.

At step 110, the method 100 can also include determining one or more tasks associated with the identified user request. The data processing system can identify a thread of tasks or operations related to the user’s request. The thread can include a sequence of tasks or operations to be executed in order to fulfill the user’s request. For example, based on the audio query “OK, I would like to go to go dinner and then a movie tonight,” the data processing system can determine the tasks of (i) generating a list of restaurants for dinner, (ii) presenting the list of restaurants to the user, (iii) generating a list of movies and/or corresponding movie theaters, and (iv) presenting the list of movies and/or movie theaters to the user. The data processing system may further identify the task of online purchasing a movie ticket and/or the task of online restaurant reservation. The data processing system can identify the thread based on a series of input audio queries such as "Ok, what movies are playing tonight" and "Ok, what restaurants are near the theater."

The data processing system may predict one or more tasks in the thread based on one or more other tasks identified from a received first audio query indicative of a first user request. For example, upon receiving a first audio query, the data processing system can determine a score or a probability value indicative of the likelihood that the first audio query represents the beginning of a thread. The likelihood score can indicate the likelihood of receiving a subsequent second audio query indicative of a second user request that is related to the first request. For

example, "Ok, what movies are playing tonight" can have a high likelihood score of being the start of a thread based on historical data indicating that requests for information related to movies are usually followed by requests for places to eat. The historical data can, for example, take into account the day of the week and/or the time of day at which the user request is made. Based on the same historical data, the query "Ok, what is the weather" can have a relatively low likelihood of being followed by a subsequent second user request.

At step 115, the method 100 can include determining predicted segment durations for one or more identified tasks. For example, the data processing system can determine a first segment duration for presenting a response to first user request and second segment duration for presenting a response to a second user request. The data processing system can predict the presentation durations by analyzing the audio content for each response to the user and determining the time durations of audio segments in the response. The prediction may also be based on an average time previous users engaged with response segments. For example, a response segment can include audio content that is 60 seconds in length, but on average users terminate the playback of the audio segment after 50 seconds. The data processing system may also determine segment durations for the tasks of generating the response to the first request and generating the response to the second request. The data processing system may predict the time durations for generating responses to user requests based on historical data of tasks previously performed by the data processing system. The data processing system may also predict durations of silence time intervals between a presentation task (a segment during which a response to a user requested is presented on the client device) and a subsequent audio query from the user. The silence time intervals can be predicted based on records of historical conversations with the user or with users in general. If a given segment is associated with predicted future user request,

the predicted duration of that segment can be scaled by the probability (or likelihood score) of that future request being made by the user.

At step 120, the method 100 can include identifying candidate ads for presenting during a given segment of the conversation with the user. The data processing system can identify the candidate ads based on one or more criteria including the relevance of the ads to one of the user requests (actual or predicted requests), the duration of the given segment, the duration of the ads, formats (e.g., audio or visual), or a combination thereof. The candidate ads can include visual ads to be displayed on the screen of the client device and/or audio ads to be played on the client device as part of the given segment. The candidate ads may include multiple versions (e.g., audio version, textual version, image versions, and/or video sequence version) of the same ad. A candidate ad can be related to a received user request (e.g., an ad for a specific movie after a user request for movies is received) or a predicted future user request (e.g., a predicted request for a restaurant to eat based on a received request for movies). Each candidate can be associated with a respective bid value. The bid values for various candidate ads can be defined based on a cost per impression (e.g., CPM), a cost per presentation unit time (e.g., cost per second (CPS)), or cost-per-click (CPC).

The given segment can be a segment for presenting a response to a received user request, a silence segment, or segment associated with generating a response to a received request. The data processing system can select candidate ads having a duration smaller than or equal to the duration of the given segment. For audio ads, the data processing system can insert an audio ad within the given segment such that the audio ad does not overlap in time with primary audio content related to a response to a user's request played during the given segment. For instance, the data processing system may place the audio ad at the start of the segment before the primary

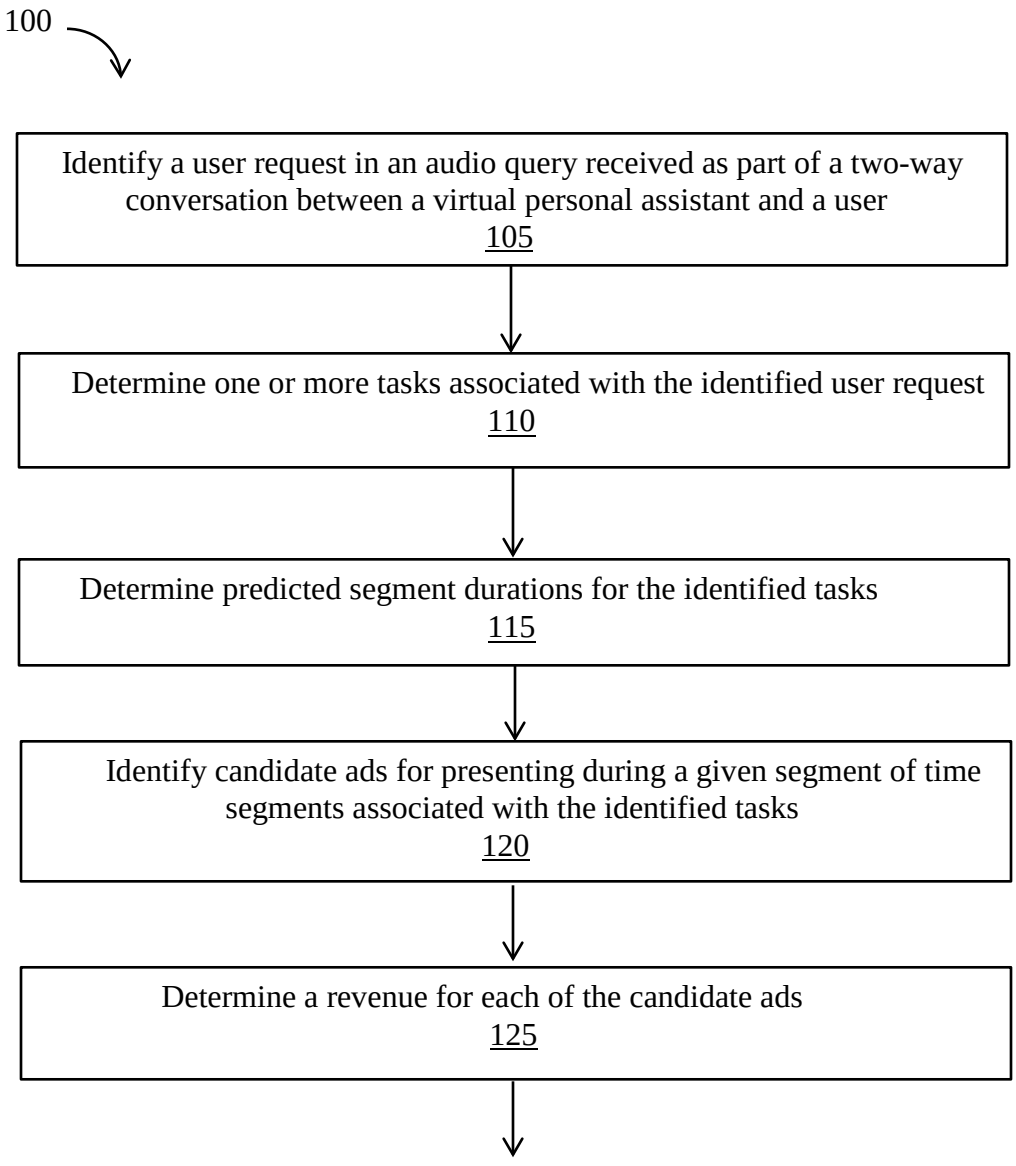
audio content, at the end of the segment after playing the primary audio content, or between audio elements of the primary audio segments.

At step 125, the method 100 can include determining a revenue for each of the candidate ads. The revenue for a candidate ad represents the amount of money the respective advertiser would pay if that ad was presented to the user on the client device. The determination of the revenue for each candidate ad may be based on the bid associated with the ad, the presentation time of the ad, or a combination thereof. For instance, if the bid is defined based on CPM or CPC, then the revenue can be equal to the bid value, however, if the bid is defined based on CPS, then revenue can be equal to the bid value multiplied by the ad's presentation time. If a candidate ad has no predefined presentation time (e.g., a visual ad), then the presentation time is equal to duration of the segment within which the candidate ad would be presented. For example, for a CPS-based bid value equal to $0.01/second and a presentation duration equal to 10 seconds, the corresponding revenue is equal to $0.1 ($0.01/second x 10 seconds). For a candidate ad having a CPM (or CPC) bid value equal to $0.05, the corresponding revenue would be equal to $0.05. However, even for CPM or CPC bids, the bid value can be increased, e.g., by a given amount if the ad presentation time exceeds a given threshold value.

At step 130, the method 100 can include selecting an from the candidate ads based on the determined revenues. The data processing system may select the ad with the highest revenue among the identified candidate ads. The data processing system may select the ad based on a combination of the determined revenues, the relevance of each candidate ad to a received user request and/or a predicted future request, and the formats of the candidate ads. For example, the data processing system can be configured to select an audio ad if, for example, the use is most likely not watching the screen of the client device (e.g., driving or exercising). The data

processing system can transmit the selected ad to the client computing device for presenting to the user. The selected ad can be transmitted as part of an output signal that includes the content segment. In some implementations, the output signal can be transmitted to a second client computing device that is different than the client computing device that transmitted the audio query. If the selected ad is an audio ad, the output signal can cause an application or other component of the client device to drive a speaker to generate an acoustic wave that corresponds to the output signal.

By employing various bid types, the data processing system can optimize monetization of valuable advertising opportunities during conversations with virtual personal assistants. In particular, bids which are monetized as a function of the time (e.g., per second, per minute, etc.) allow for efficient use of any advertising time slots within two-way conversations between users and the virtual personal assistant. Also, such type of bid allows for accommodation of a number of short audio ads.


100
Identify a user request in an audio query received as part of a two-way conversation between a virtual personal assistant and a user
105
Determine one or more tasks associated with the identified user request
110
Determine predicted segment durations for the identified tasks
115
Identify candidate ads for presenting during a given segment of time segments associated with the identified tasks
120
Determine a revenue for each of the candidate ads
125

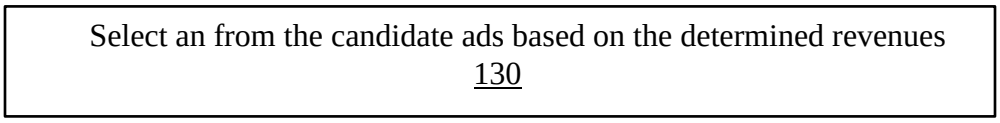

Select an from the candidate ads based on the determined revenues
130

FIG. 1

## ABSTRACT

Systems and methods described herein allow for optimization of selection and monetization of ads for presenting during two-way conversations between users and a virtual personal assistant. A data processing system can receive an audio query from a client device and identify a user request based on the received audio query. The data processing system can determine a thread of tasks associated with the identified user request, and determine predicted segment durations for the identified tasks. The data processing system can identify candidate ads for presenting during a given segment, and determine a revenue value for each candidate ads. The data processing system can then select an ad among the candidate ads based on the determined revenues and the predicted segment durations. The data processing system can then transmit the selected ad to the client device for presentation to a corresponding user.