

## Technical Disclosure Commons

---

Defensive Publications Series

---

June 10, 2019

# AMBIENT AUDIO AUTO-SUGGEST

Tim Wantland

Brandon Barbello

Follow this and additional works at: [https://www.tdcommons.org/dpubs\\_series](https://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Wantland, Tim and Barbello, Brandon, "AMBIENT AUDIO AUTO-SUGGEST", Technical Disclosure Commons, (June 10, 2019)  
[https://www.tdcommons.org/dpubs\\_series/2256](https://www.tdcommons.org/dpubs_series/2256)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

## AMBIENT AUDIO AUTO-SUGGEST

A computing device can analyze ambient audio, identify entities referenced in that audio, and make those entities actionable on a display of the computing device. For example, a mobile computing device can identify that background audio has referenced a famous person and present a selectable notification on a lock screen of the mobile computing device. Should a user select the notification, the mobile computing device can present additional information regarding the famous person (e.g., from a Wikipedia article). The ambient audio can be audio from user conversations or from media content (e.g., a pre-made movie or a live sporting event). Other use cases involve analyzing audio to determine whether a different language is being spoken, and providing an option to translate the audio. Various use cases and details regarding this technology are discussed below.

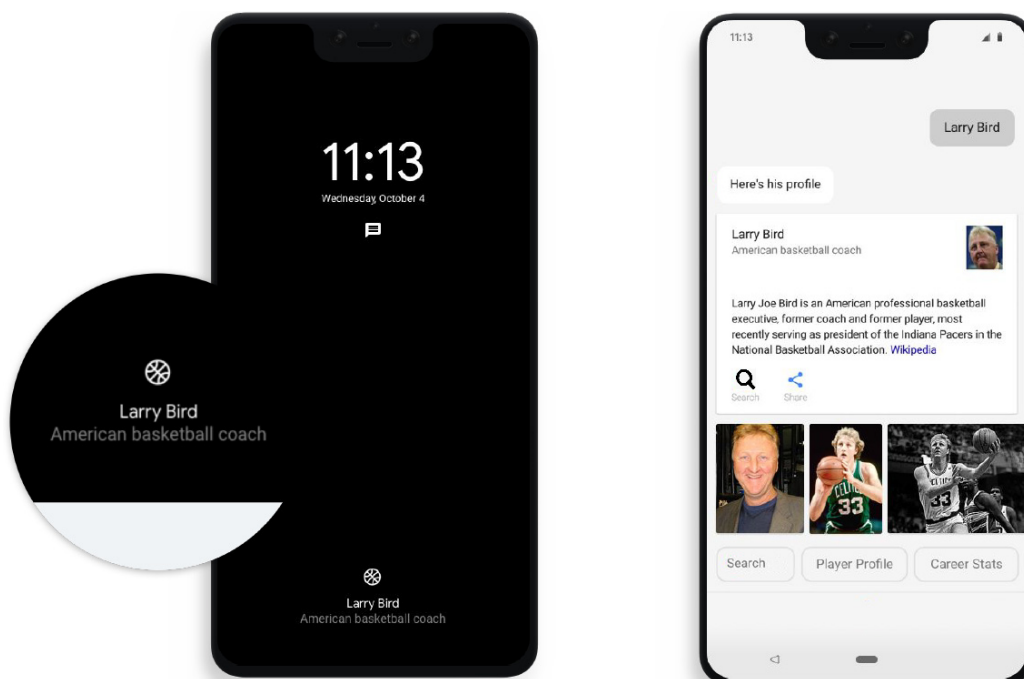
### MEDIA INTELLIGENCE

It is hard to keep up with everything that that is heard during a given day. With this technology, a phone can have a continuous understanding of what a person is listening to. The phone may not just recognize which song is being played in the background (e.g., US 2019/0102458 describes an example of technology that can recognize songs being played in the background), but can also recognize entities being discussed in media, such as movies, TV, radio, podcasts, and audiobooks – the sorts of things that a user may want to look up and learn more about from these different types of media. The phone can present these entities on the lock screen of the phone (e.g., the screen that is always on, or that is first presented upon waking the phone from an off state). Presenting these entities on the lock screen (e.g., as a selectable

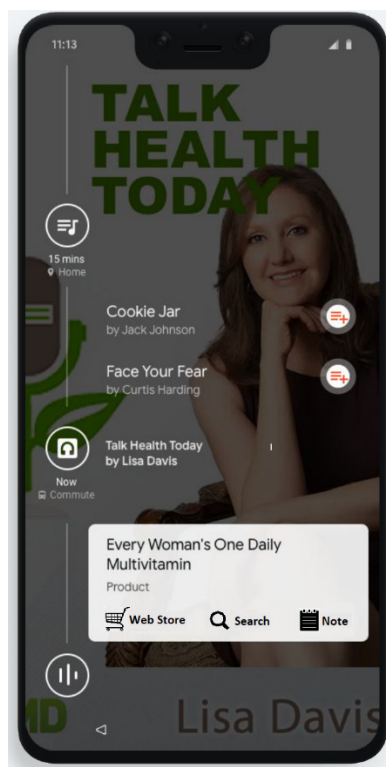
notification) can make those entities immediately actionable and allow a user to view additional information on the entities without having to type out the entity to perform a search on the entity.

The entities that the phone has recognized over a period of time can also be stored by the phone and listed in a history of ambient audio. A user can select a displayed entity on either the lock screen or the historical list of entities to learn more about the entity. For example, in response to a user selection of an entity, such as by single or double tapping on displayed representation of an entity, the phone may perform an internet search based on the entity and/or bring up a Wikipedia entry about the selected entity.

The images below show how a phone recognized that ambient audio referenced the basketball player and coach Larry Bird, and how the phone can present a notification on a lock screen that identifies Larry Bird. Responsive to a user selecting the notification, the phone can present additional information on Larry Bird, such as images of Larry Bird, information from a Wikipedia article about Larry Bird, and options to perform a search engine search for websites that discuss Larry Bird.



For media that is played by the device (e.g., a podcast that is played by the phone), a user interface can show information that relates to the content of the media being played. That information can include selectable elements that allow a user to initiate a search on an entity mentioned by the media being played, create a note about the entity, or even navigate to a web store at which the entity can be purchased (e.g., when the entity is a product). The below image shows a phone that presented information on a type of multivitamin, responsive to the phone determining that a podcast being played by the phone referenced multivitamins.



The entity recognition can even work with live broadcasts, and a TV show can become actionable on a device as that device analyzes audio of the TV show and determines that the TV show is referencing an entity. The device may then show a selectable element that references the entity on a screen of the device. For example, a device may determine that a live broadcast is

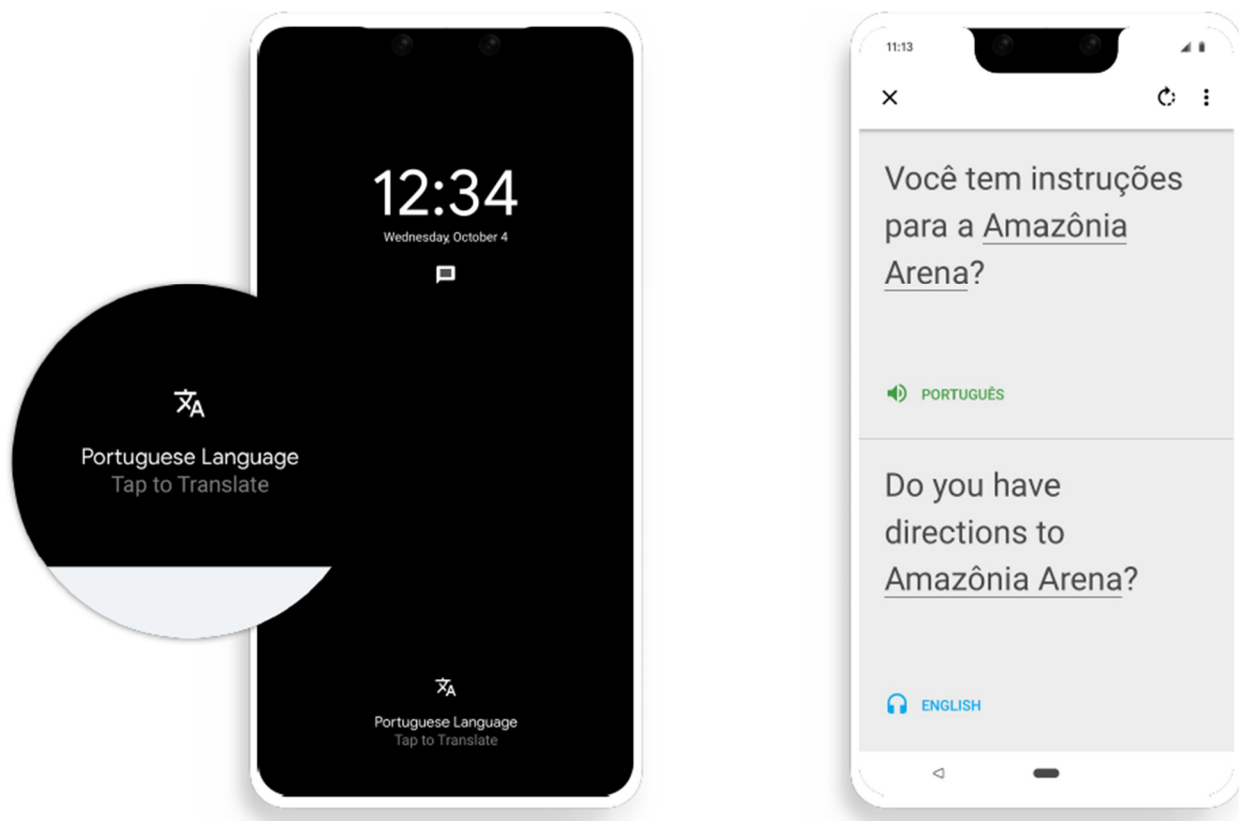
referencing a particular city, and can present a name of that city on a screen of the computing device. The device may also present selectable buttons to view a Wikipedia article that relates to the city, perform a web search using the name of the city as the query, or store the name of the city as a note.

## TRANSLATION INTELLIGENCE

A computing device may also be able to recognize that a non-native language of a user is being spoken, and can offer to translate the non-native language to a native or default language of the user of the device. As with the discussion of recognized entities, a notification that provides a user the option to translate the audio may appear on a lock screen of the computing device (if the phone is locked), or on a notification or control panel that a user can slide out from the side of the display. The ability of a phone to automatically recognize foreign language in audio and present users an immediate option to translate the audio can help users when they travel, and can also help with homework and language learning scenarios. Such a technology can even work in tandem with headphones that generate translated speech of audio spoken in the background in different languages.

This technology can involve analyzing audio and storing that audio in a buffer if it is determined to include a foreign language. The device may then present a notification that a foreign language has been spoken. In response to user input that selects the notification, the device can perform a more-detailed analysis of the audio that has been stored in the buffer and can present a text (and/or audio) translation of the foreign language. The images below show how a phone recognized that ambient audio included words spoken in Portuguese, and presented a notification to a user of the computing device. In response to a user selecting the notification,

the phone analyzed the audio in the buffer and presented a user interface that showed both the original language and the translated language.



## PROCESSING

A computing device can perform the entity recognition using on-device machine learning processing to determine whether portions of audio match a locally-stored database of entities. The entity recognition can involve performing speech recognition of audio to generate a text representation of recorded audio, and then performing entity recognition on the recognized text. All the processing may be performed locally on the computing device, and the device may not send the audio or text generated from the audio to any remote computing devices (at least not until a user interacts with a notification, such as selecting a notification and requesting that a web

search be performed to gather additional information on the entity). The computing device would include setting that allows a user to activate and deactivate audio analysis and entity recognition, and a device would not analyze ambient audio until a user has activated the setting. Moreover, a user may be provided with controls allowing the user to make an election as to both if and when systems described herein may enable collection of user information (e.g., a list of entities recognized). 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. As such, a 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.

#### ABSTRACT

A computing device can analyze ambient audio, recognize entities referenced in the ambient audio, and provide actionable notifications of the entities on a display of the computing device. As an example, a computing device may determine that audio in the background has referenced a famous person and can generate a notification that names the famous person and lists the reason that they were famous (e.g., “Basketball Player”). User selection of the notification may cause the computing device to present a user interface that shows additional information on the person and/or selectable buttons for performing actions related to the individual. These actions can include performing a search engine search using the entity name as the query, creating a note that stores the entity name, and in those circumstances when the entity is a product, navigating to a web store to purchase the product.