February 06, 2019

Automatic translation in voice calls

Sooraj Sasindran

Naveen Kalla

Automatic translation in voice calls

ABSTRACT

Participants in a voice call do not always have a common language. This can pose problems since participants may not be able to understand each other during the call. Some current techniques to address this problem involve incorporating a machine translator that can perform translation. However, for machine translation to be provided to all participants, such techniques require a machine translator mechanism to be provided at each participant's end, which is not always possible.

The techniques of this disclosure incorporates machine translation techniques on a participant device such that the device can perform local translation of speech in a call. Machine translation is performed to automatically translate speech in the call, e.g., between the languages spoken by participants at the near and the far ends. In this manner, automatic translation of voice calls is enabled for all participants in a voice call.

KEYWORDS

- automatic translation
- machine translation
- voice call
- phone call
- language detection

BACKGROUND

Participants in a voice call do not always have a common language. This can pose problems since participants may not be able to understand each other during the call. Some current techniques to address this problem involve incorporating a machine translator that can

perform translation. However, for true machine translation for all participants, the present techniques require a machine translator at each end, which is not always possible.

In certain contexts, the inability of participants to understand each other can have significant consequences, e.g., if an emergency phone call is made by a person who speaks a language that the emergency dispatcher does not understand.
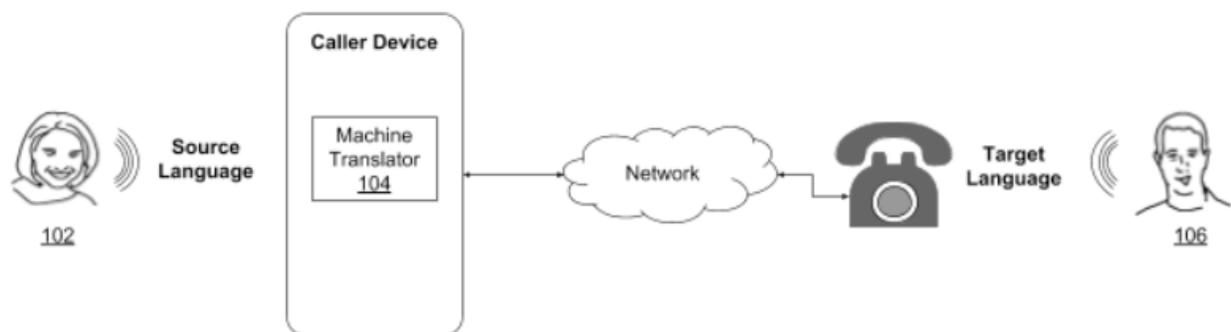
DESCRIPTION

Fig. 1: Automatic machine translation of voice calls

Fig. 1 illustrates an example of automatic two-way machine translation of voice calls, per techniques of this disclosure. The techniques are implemented with permission from participants in the call. Participants are provided with options to turn off translation, e.g., prior to a call, during a call, etc. Further, participants are provided indications that automatic translation and speech rendition (e.g., by a machine-generated voice) is being utilized.

A machine translator (104) is inserted in the audio path of the device that is used to place a call. In this context, the calling device that places the call is referred to as the host. The machine translator translates the speech received from the calling device or host (102) to a language understood by the call recipient (106). The outgoing speech is translated from the host's language to a language understood by the recipient and the incoming speech is translated

in the opposite direction, e.g., from the language of the call recipient to the language of the host. The translated speech is rendered using a machine-generated voice.

For example, the machine translator is configured to detect the language of the speech received from the host and from the recipient. Alternatively, participants can indicate the language used, e.g., via a setting in a user interface for the call. The setting can be modified by the host.

In this manner, the machine translator is built into the host device, e.g., it is directly inserted into the audio path of the host device. For example, a smartphone or other call-capable device can include these features, e.g., as part of the device operating system, a calling application, etc. Settings are provided to control the translator such that the host can choose between an automatic mode, where the translator auto-detects the languages used at both ends, and a manual mode, where the language of the recipient is selected by the host. Incorporating the machine translator in the host device obviates the need for a machine translator at the recipient end since the machine translator at the host end translates bi-directionally. Further, the machine translator can be implemented entirely in software that can run on available hardware of the host device. The techniques can be utilized for voice/video calls, e.g., over plain old telephone service (POTS) lines, over cellular protocols such as 3G, 4G, LTE, etc., over VoIP connections, on over-the-top (OTT) calls, over conferencing (multi-party) calls, 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

The techniques of this disclosure incorporates machine translation techniques on a participant device such that the device can perform local translation of speech in a call. Machine translation is performed to automatically translate speech in the call, e.g., between the languages spoken by participants at the near and the far ends. In this manner, automatic translation of voice calls is enabled for all participants in a voice call.