January 2021

Removing Crutch Phrases from Audio Containing Speech

Ken Schow

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Removing Crutch Phrases from Audio Containing Speech

ABSTRACT

Speakers generally find crutch phrases useful to pause and compose their thoughts as they speak. Listeners can find such utterances distracting and/or confusing. Crutch phases in technology-mediated communication such as audio and/or video calls can affect the quality of interaction. This disclosure describes techniques, implemented with user permission, to process speech audio to detect and remove crutch phrases from speech audio. Phrases in the speech are compared with a standard library to detect crutch phrases in the audio stream and to remove such phrases. With permission, natural language processing and/or machine learning techniques can be used to train a model based on an individual's crutch-word use and speaking style to improve the speed and accuracy of crutch phrase removal. Crutch phrase removal can be implemented within any application, platform, or service that involves speech content, such as audio or video conferencing applications, video streaming, transcription services, etc.

KEYWORDS

- Crutch phrase
- Filler word
- Post-speech filter
- Natural language processing (NLP)
- Audio conferencing
- Video conferencing
- Speech recording
- Streaming broadcast

BACKGROUND

When speaking naturally, people typically tend to use crutch phases, such as "um," "so," etc., as filler words within and/or between sentences. Speakers generally find such phrases useful to pause and compose their thoughts as they speak. However, listeners can find such utterances distracting, especially during formal settings such as meetings, speeches, classroom instruction, etc. Further, crutch phrases can potentially create confusion regarding the speaker's intended message. In addition to in-person conversations, the potential distraction and confusion created by crutch phases can affect video and/or audio-based online interactions, such as online meetings, broadcasts, etc.

Video and audio processing techniques are utilized in online meetings to filter out undesired ambient sounds, such as background noise. In addition, natural language processing (NLP) based models can be used for recognizing spoken phrases within the audiovisual content, e.g., to provide a near real-time transcript. However, these techniques do not currently involve special treatment for crutch phrases encountered within the content being processed.

DESCRIPTION

This disclosure describes techniques to automatically analyze speech audio (accessed with user permission) to detect and remove crutch phrases in the speech audio. For example, the crutch-phase removal can be performed by processing the audio with post-speech filter software designed specifically to detect and remove crutch phrases.

To enable detection, a standard library of known crutch phrases is provided to the post-speech filter software. With user permission, if the filter detects the presence of any of the phrases in the library within the speech being analyzed, the audio corresponding to that phrase is removed from the audio stream.

Further, with user permission, detection and removal of crutch phases can be personalized to individual users. For such personalization, a machine learning model (incorporating natural language processing techniques) can be trained (locally on a user's device) based on the user's speech. Such training can help augment the standard library of known crutch phrases with crutch phrases that may be specific to the user. Additionally, with the user's permission, the personalized model can learn correlations between the user's speech cadence, word use, and sentence construction patterns to predict when the user is likely to use crutch phrases. Such user-permitted personalization can help adapt to individual speaking styles, thus improving the speed and/or quality of detection and removal of crutch phrases.
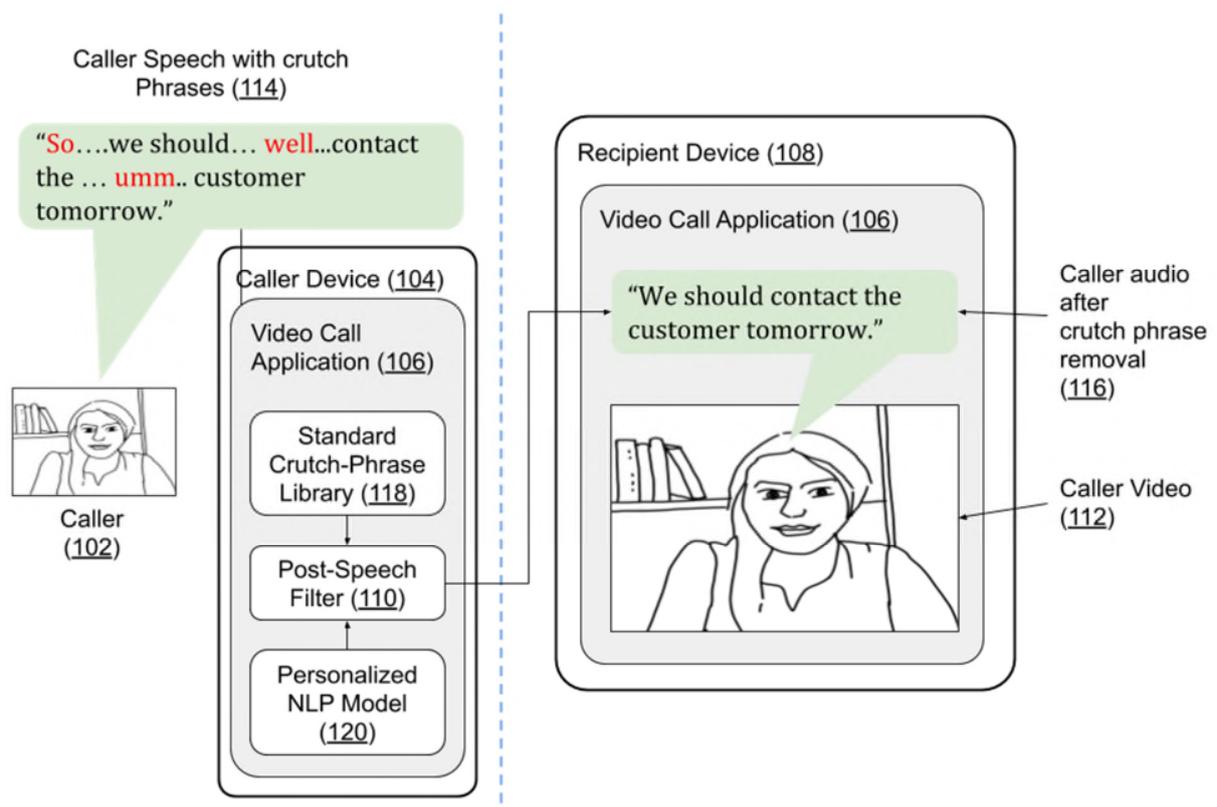
Fig. 1: Removing crutch phrases in speech during a video call

Fig. 1 shows an operational implementation of the techniques described in this disclosure. As shown in Fig. 1, a caller (102) uses a video call application (106) on caller device (104) to make a video call to a recipient that uses a recipient device (108) to receive the call. The caller's speech (114) contains crutch phrases (shown in red in Fig. 1).

With the caller's permission, post-speech filter software (110) is used to process the audio of the caller's speech to remove the crutch phrases. The filter detects crutch phrases in the speech by matching against a standard library of crutch phrases (118). With the caller's permission, the filter can be personalized using NLP models (120) specific to the caller's speaking preferences and style. Caller audio without crutch phrases (116) obtained as output of the post-speech filter is provided to the recipient along with the caller's video (112).

While Fig. 1 shows a video calling scenario, with user permission, crutch phrase removal can be performed for any multimedia content that contains speech audio, such as phone calls, video conferences, streaming broadcasts, instructional recordings, etc. If users permit, apart from the speech audio, the techniques can remove the corresponding crutch phrases from the transcriptions of the audio as well. Such removal can be helpful in various situations where audio transcripts are used, such as closed captions for online videos, subtitles for movies and TV shows, transcripts of video conference meetings, etc.

Crutch phrase removal can be implemented within any application, platform, or service that involves speech content, such as audio or video conferencing applications, video streaming, transcription services, etc. Implementation of the techniques can enhance the quality of online communication by reducing the listener distraction and confusion that can result from a speaker's crutch phase utterances, thus enhancing the user experience (UX).

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

This disclosure describes techniques, implemented with user permission, to process speech audio to detect and remove crutch phrases from speech audio. Phrases in the speech are compared with a standard library to detect crutch phrases in the audio stream and to remove such phrases. With permission, natural language processing and/or machine learning techniques can be used to train a model based on an individual's crutch-word use and speaking style to improve the speed and accuracy of crutch phrase removal. Crutch phrase removal can be implemented within any application, platform, or service that involves speech content, such as audio or video conferencing applications, video streaming, transcription services, etc.