

Technical Disclosure Commons

Defensive Publications Series

May 2020

Health Diagnostics Using User Utterances

Aleksandar Kracun

Ignacio Lopez Moreno

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

Recommended Citation

Kracun, Aleksandar and Moreno, Ignacio Lopez, "Health Diagnostics Using User Utterances", Technical Disclosure Commons, (May 07, 2020)

https://www.tdcommons.org/dpubs_series/3220



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.

Health Diagnostics Using User Utterances

ABSTRACT

Respiratory illnesses can be hard to track and diagnose. Obtaining useful clinical data on these illnesses is difficult because it requires physical interaction, e.g., via nasal or sinus swab. It is known that respiratory illness can impact speech pathways. To this end, this disclosure describes techniques to use readily accessible software to obtain and classify potentially useful data. With user permission, utterances of the user, e.g., activation of a speech-activated device via a hotword, are analyzed to form speaker-ID models. These models are evaluated against additional utterances of the user in a sequential manner. The evaluation scores, along with the timestamps and details of the models, are aggregated to determine if the user has an interval of time where their speaker-ID models are unstable, inconsistent, or lacking self-similarity. This signal can be used as a proxy for detection or as a motivating factor for clinical investigation.

KEYWORDS

- Respiratory illness
- Speaker identification
- Hotword
- Health diagnostics
- Speech patterns
- Speaker-ID model

BACKGROUND

Respiratory illnesses can be hard to track and diagnose. Obtaining useful clinical data on these illnesses is difficult because it requires physical interaction, e.g., via nasal or sinus swab.

Patients that have respiratory illness may experience subtle or slow-onset symptoms or may be asymptomatic and thus, not contribute to clinical data.

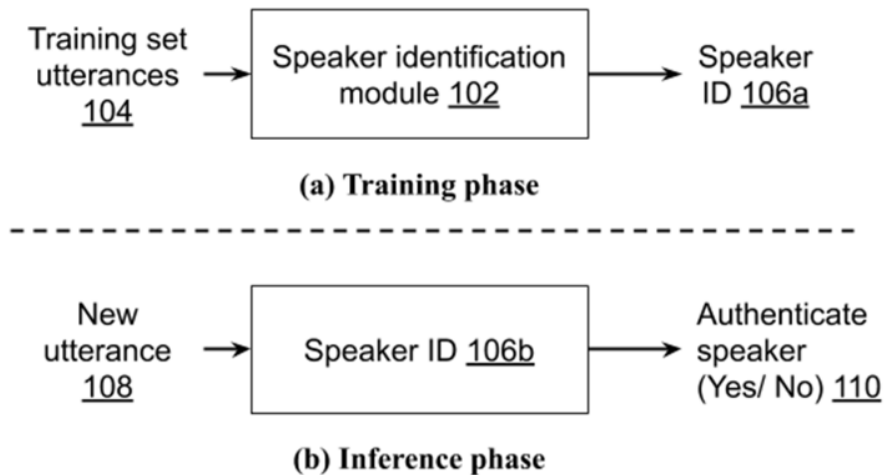


Fig. 1: Speaker identification

Fig. 1 illustrates a typical speaker identification operation. In a training phase, illustrated in Fig. 1(a), the user utters a known set of training-set phrases (104), e.g., “hey,” “okay,” etc. For example, such phrases may include hotwords used for device activation, e.g., to put the device in a listening mode where it listens and responds to user commands. A speaker identification module (102), which can be a machine-learning model, builds a speaker-ID model (106a) (also called fingerprint) based on the user’s utterances. A speaker-ID model S that is built out of an utterance X is mathematically denoted by $S = \text{SID}(X)$.

In an inference phase, illustrated in Fig. 1(b), the speaker-ID model (106b) tests a new utterance (108) for authenticity and determines if the speaker of the new utterance is the known user associated with the device (110). Mathematically, the evaluation of an utterance X for authenticity by a speaker-ID model S is denoted as $E = \text{eval}(S, X)$. An utterance X is declared to be authentic if the evaluation score E is less than a particular threshold.

DESCRIPTION

This disclosure leverages the observation that respiratory illnesses can impact speech pathways. With user permission, utterances of the user, e.g., activation of a speech-activated device via a hotword, are analyzed to determine if there is a period of user input where their speaker-ID models are unstable, inconsistent, or lacking self-similarity. In such a case, the device that ordinarily receives spoken input from the user can raise an alert that the user is not sounding like their usual self.

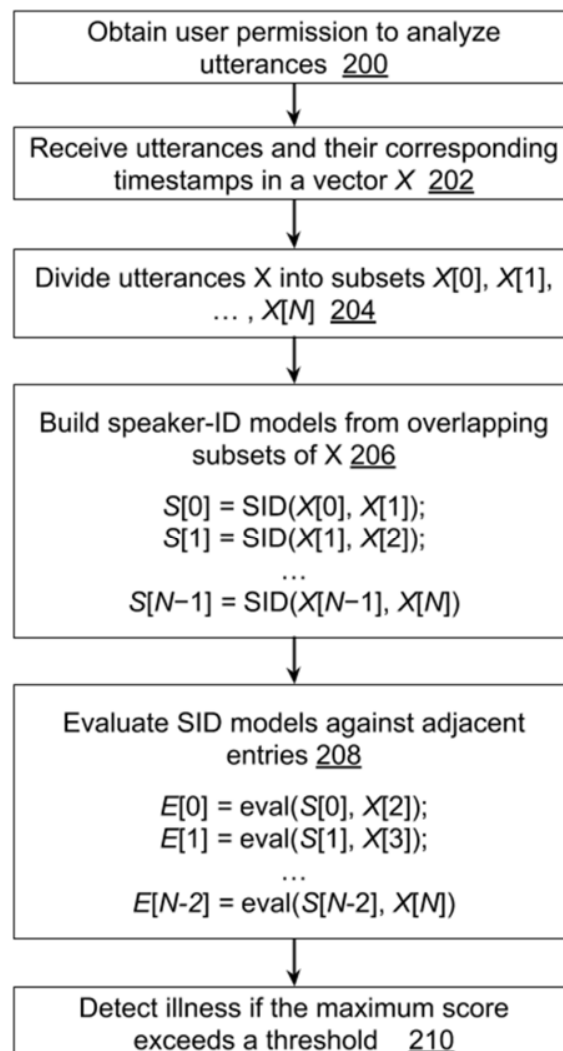


Fig. 2: Health diagnostics using speaker identification

Fig. 2 illustrates an example process to generate health diagnostic information using speaker identification, per the techniques of this disclosure. User permission is obtained to analyze utterances provided by the user for the purpose of determining abnormalities, e.g., likelihood of illness or other condition that impacts speech pathways (200). If the user does not provide such permission, user's utterances are used for the specific purpose of listening to and interpreting user commands. If the user provides permission, steps 202-210 are performed to determine if there is a change in the user's speech that may be indicative of an illness. The user can also choose to turn off speech recognition and/or speaker identification entirely.

Upon receiving user permission, the user is registered and authenticated with a speaker-identification model. For this user, recent utterances and their timestamps $\langle u, t \rangle$ are converted into a vector X (202). The vector X is divided into subsets $X[0], X[1], \dots, X[N]$ (204). Speaker-ID models are built from non-overlapping subsets of X (206):

$$\begin{aligned} S[0] &= \text{SID}(X[0], X[1]); \\ S[1] &= \text{SID}(X[1], X[2]); \\ &\dots \\ S[N-1] &= \text{SID}(X[N-1], X[N]). \end{aligned}$$

With user permission, the speaker-ID models $S[0], S[1], \dots, S[N-1]$ are stored in a vector S . The speaker-ID models are evaluated against adjacent entries of X (208), resulting in a vector of evaluation scores:

$$\begin{aligned} E[0] &= \text{eval}(S[0], X[2]); \\ E[1] &= \text{eval}(S[1], X[3]); \\ &\dots \\ E[N-2] &= \text{eval}(S[N-2], X[N]). \end{aligned}$$

The scores $E[0], E[1], \dots, E[N-2]$ are stored in a vector E . A property of E is found that is associated with abnormal speech input, e.g., caused by respiratory illness. For example, an

increasing sequence of scores can be indicative of a change in the user's speech patterns, or possibly of respiratory illness. As another example, illness, or at least a substantive change in user's speech patterns, can be declared if the maximum element of E exceeds a threshold, *illness_constant* (210):

$$\text{find}(\max(E) > \textit{illness_constant}).$$

The threshold *illness_constant* can be determined via statistical experiments, e.g., from users who volunteer diagnostic data, or by finding a maximum across a subset of users.

Alternatively, additional factors can be used to determine *illness_constant*. For example, with the user's permission, data indicating that the user visited a healthcare facility or had recent contact with other individuals with respiratory illnesses, can be used to adjust, e.g., reduce *illness_constant*.

Filtering logic can be applied to the utterances or users to determine the speech samples that are used (e.g., at blocks 206 and 208), in order to avoid noise or misattributed utterances. Alternative window sizes or strides can be used at one or more of blocks 206, 208, and 210.

In this manner, with user permission, a hotword or speaker-ID technique can be adapted to build a detector for respiratory illness or other conditions affecting speech. With user permission, the described techniques make use of a speaker's regular triggering of the speaker-ID model to effectively obtain clinical data derived from their utterances.

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 speech input and utterances, 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 to use readily accessible software to obtain and classify potentially useful data. With user permission, utterances of the user, e.g., activation of a speech-activated device via a hotword, are analyzed to form speaker-ID models. These models are evaluated against additional utterances of the user in a sequential manner. The evaluation scores, along with the timestamps and details of the models, are aggregated to determine if the user has an interval of time where their speaker-ID models are unstable, inconsistent, or lacking self-similarity. This signal can be used as a proxy for detection or as a motivating factor for clinical investigation.

REFERENCES

1. [Coughona](#) accessed Apr. 30, 2020.
2. [COVID Voice Detector](#) accessed Apr. 30, 2020