

Emotion Recognition from Voice in the Wild

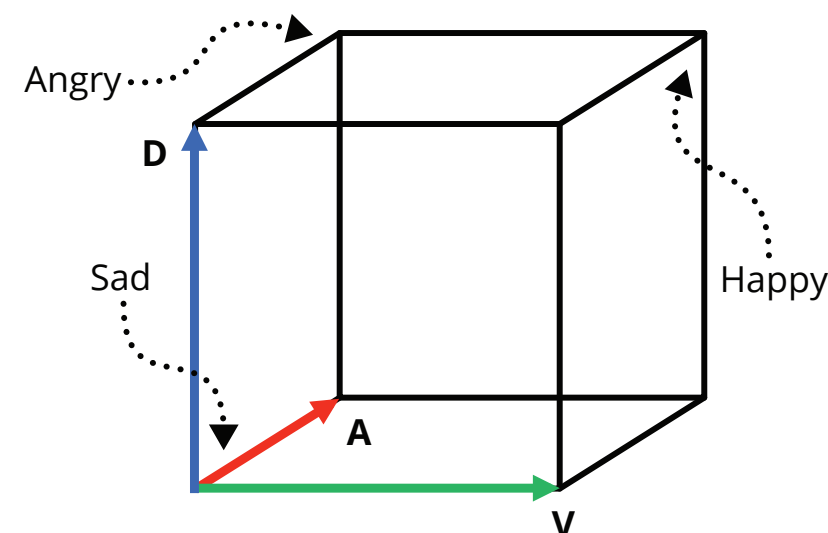
Introduction

Accurately recognizing emotion from voice is important in defense applications such as speaker profiling and human-machine teaming, but is currently infeasible. We introduce a new, continuous speech emotion recognition database, CMU-SER, and a set of micro-articulatory techniques that can capture finer nuances than the current state of the art.

The CMU-SER Database

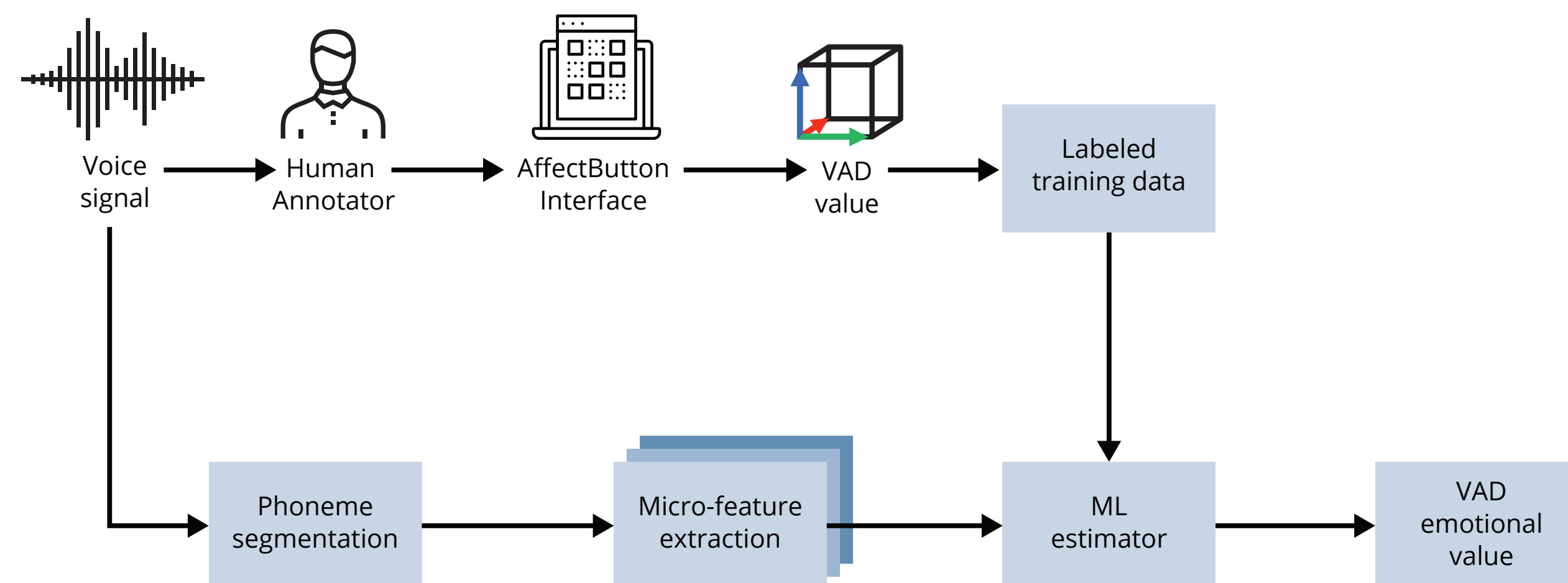
The CMU-SER database, to be released at the end of 2019, is the largest ever speech emotion recognition database. In-the-wild audio clips come from podcasts, radio, and television, and are annotated via crowdsourcing. Annotators used an interface based on the VAD emotional state model that allows users to pick from a continuum of emotions instead of discrete labels. Several annotators label each audio clip. CMU-SER features:

- Over **29,000** annotated audio clips, totaling over **54** hours of voice recordings
- Over **324,000** unique annotations



The VAD model: Valence, arousal, and dominance characterize affect in three dimensions.

We introduce a new in-the-wild **speech emotion recognition database**, and **novel extraction techniques** built with machine learning and micro-articulatory.



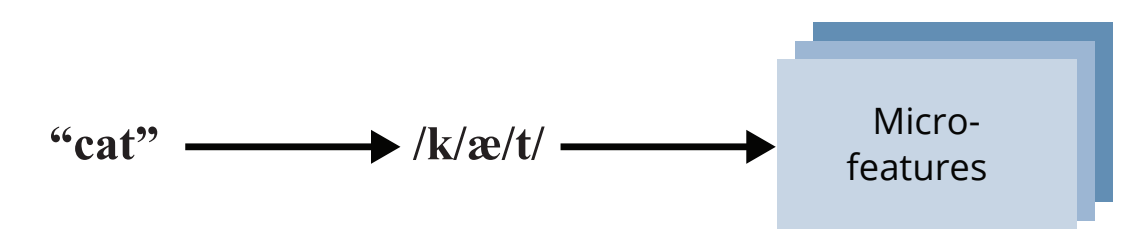
System diagram. Voice data is labeled by crowdsourcing participants using the AffectButton interface. This labeled data is then used to train classifiers and predict emotion.



The AffectButton: An interface first introduced for self-reporting human affect. The facial icon changes based on mouse movement, and each expression corresponds to a unique VAD score.

Micro-Articulometry and Machine Learning

Prior work in speech emotion recognition typically operates at the utterance level—the level of the spoken word or statement. Instead, our approach operates at the phoneme level—the level of the constituent units of speech—using micro-articulatory techniques. We use fine-grain voice features, such as formant position, in conjunction with deep learning to predict emotional state. Results will be published at the end of 2019.



Micro-articulometry: The measurement and modeling of articulatory properties at the phoneme level.

Example Voice Features:

- Diplophonicity
- Flutter
- Formant bandwidth
- Formant position
- Vocal fry
- Glottalization
- Nasality
- Raspiness
- Shimmer
- Tremor
- Voicing onset time
- Wobble

Micro-articulatory voice features are used to train classifiers and predict emotion. Each voice feature requires its own set of signal processing algorithms to extract and measure.

Copyright 2019 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

* These restrictions do not apply to U.S. government entities.

Carnegie Mellon* is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM19-1034