

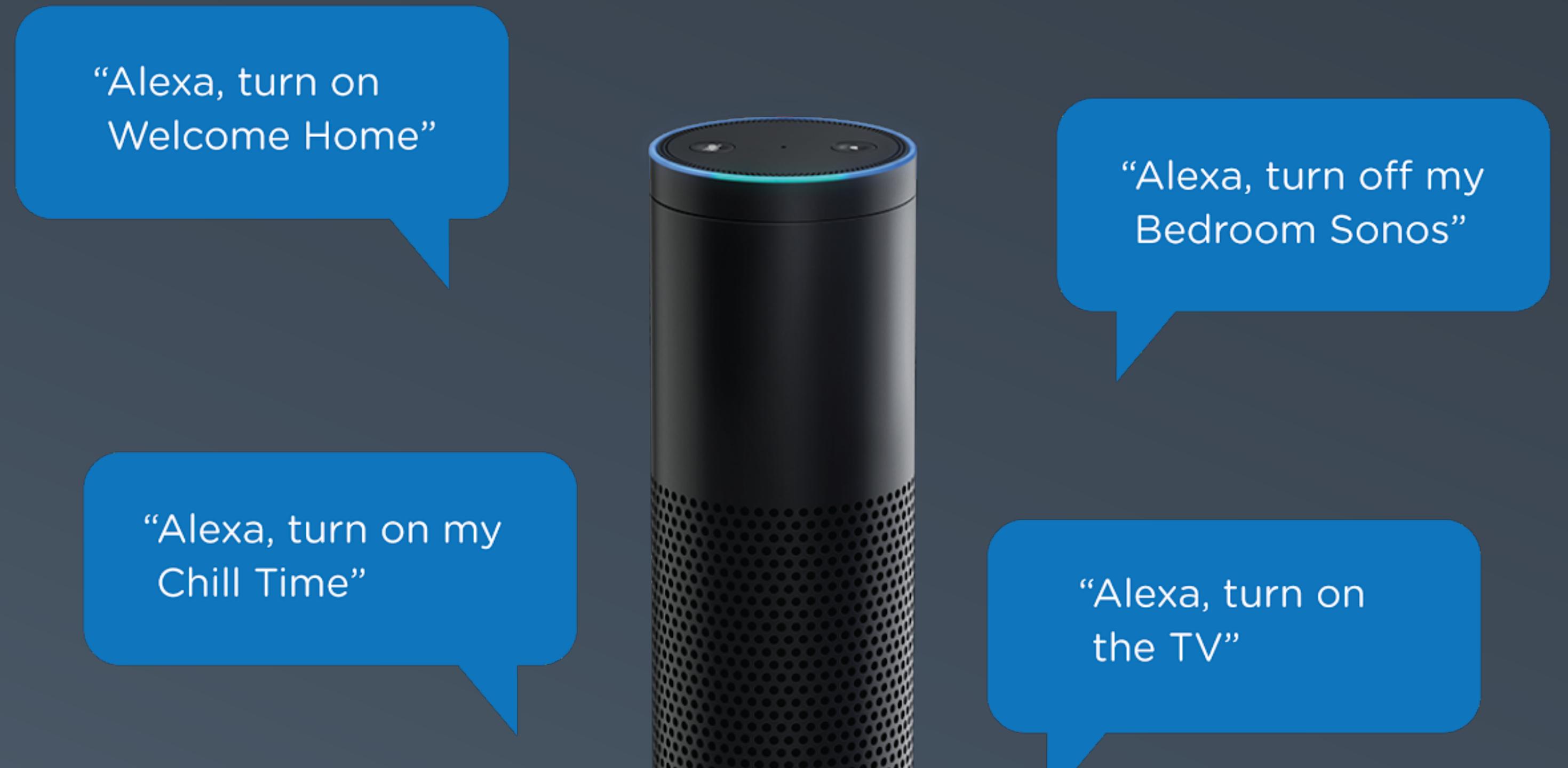
Enabling Speech-to-Meaning with Acoustic Language Processing



Nicolas Perony | AMLD 2019

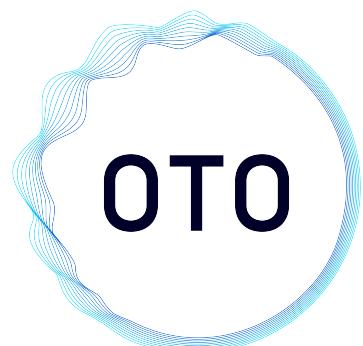
WHAT DO WE TALK ABOUT WHEN WE TALK ABOUT **LANGUAGE**

There is nothing **Natural**
about (much of) the **Language**
we are **Processing**

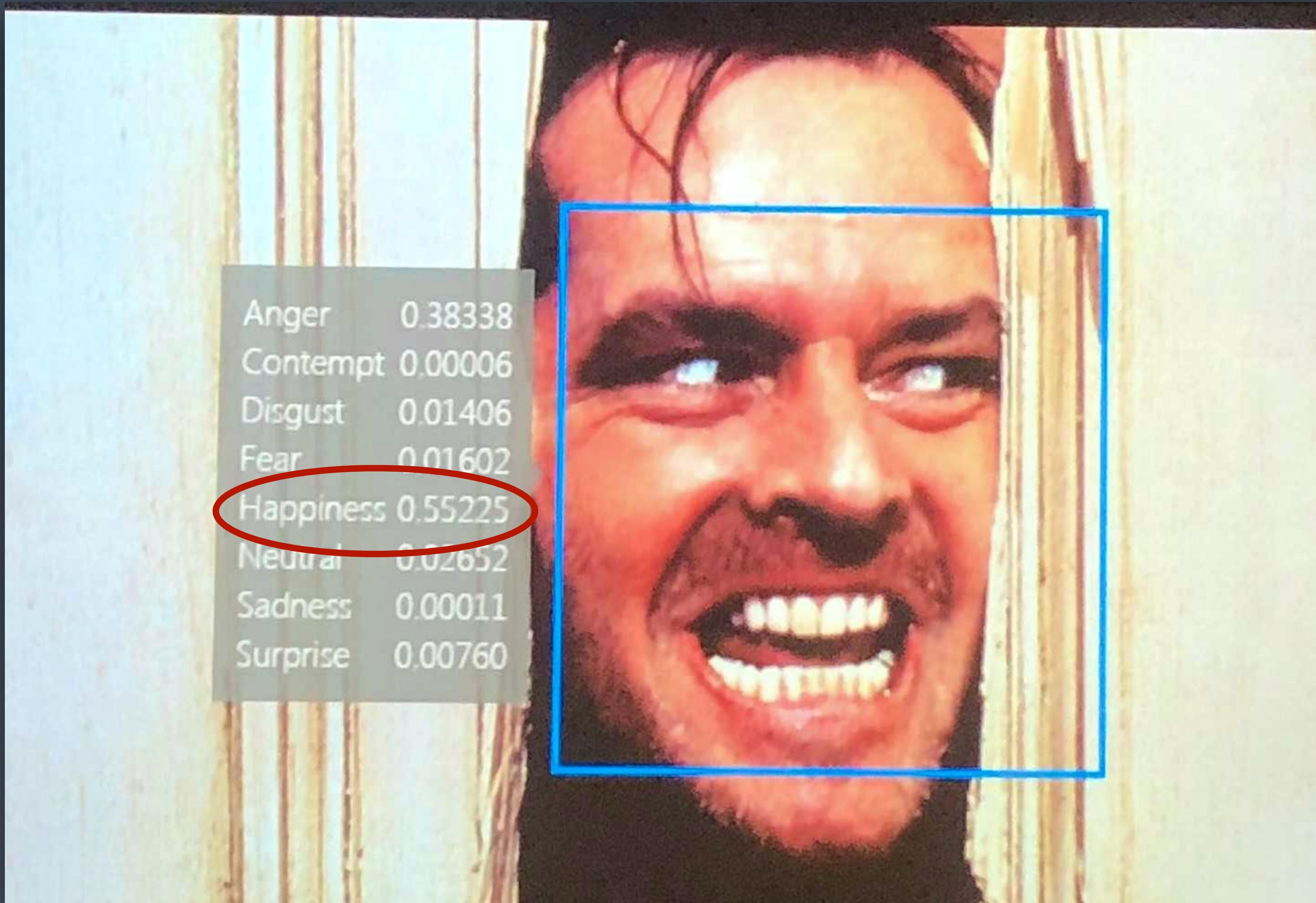


*Our conversations with machines are **uni-dimensional**.*

Speech understanding in 2019



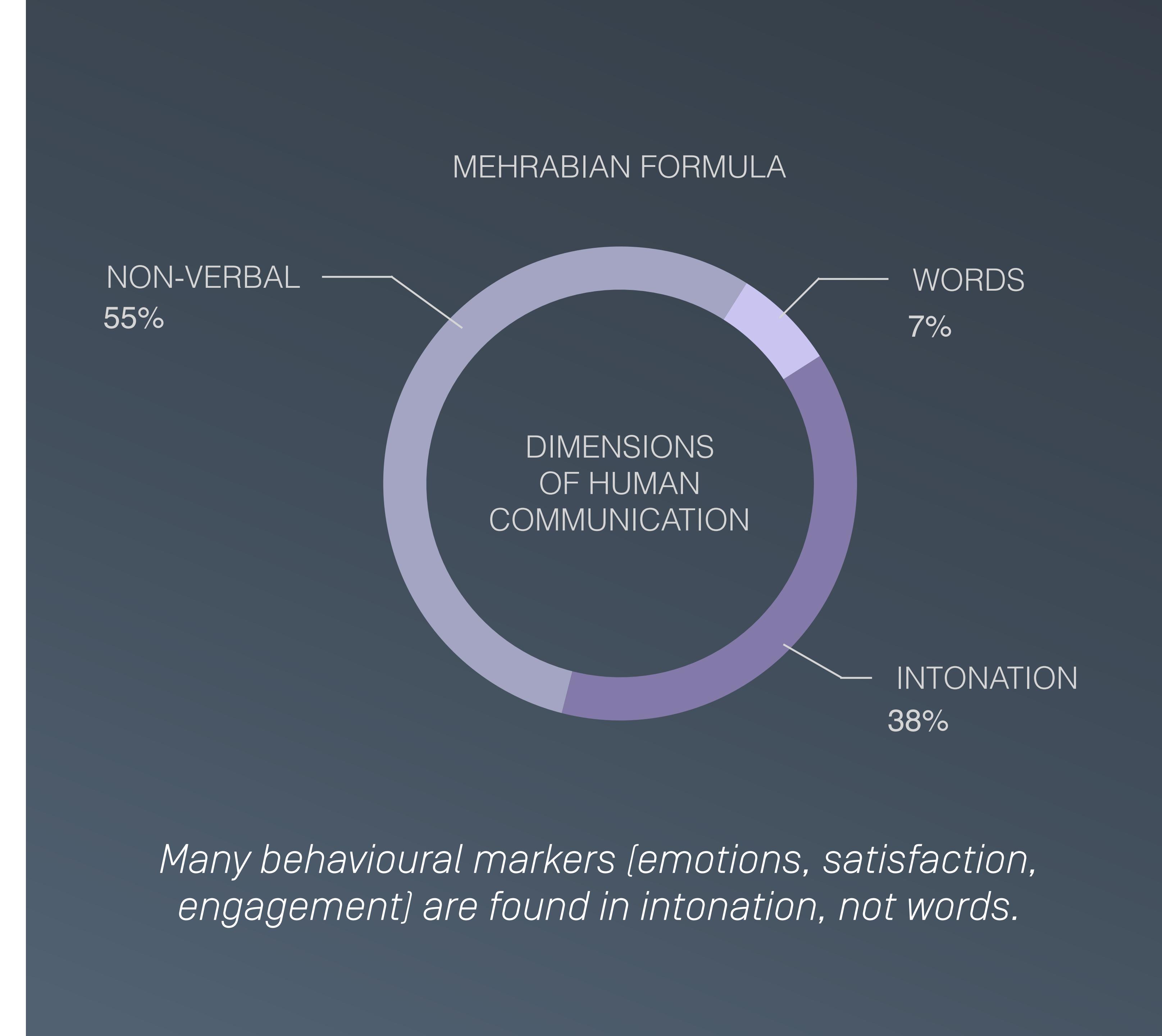
Emotion recognition is (really) hard



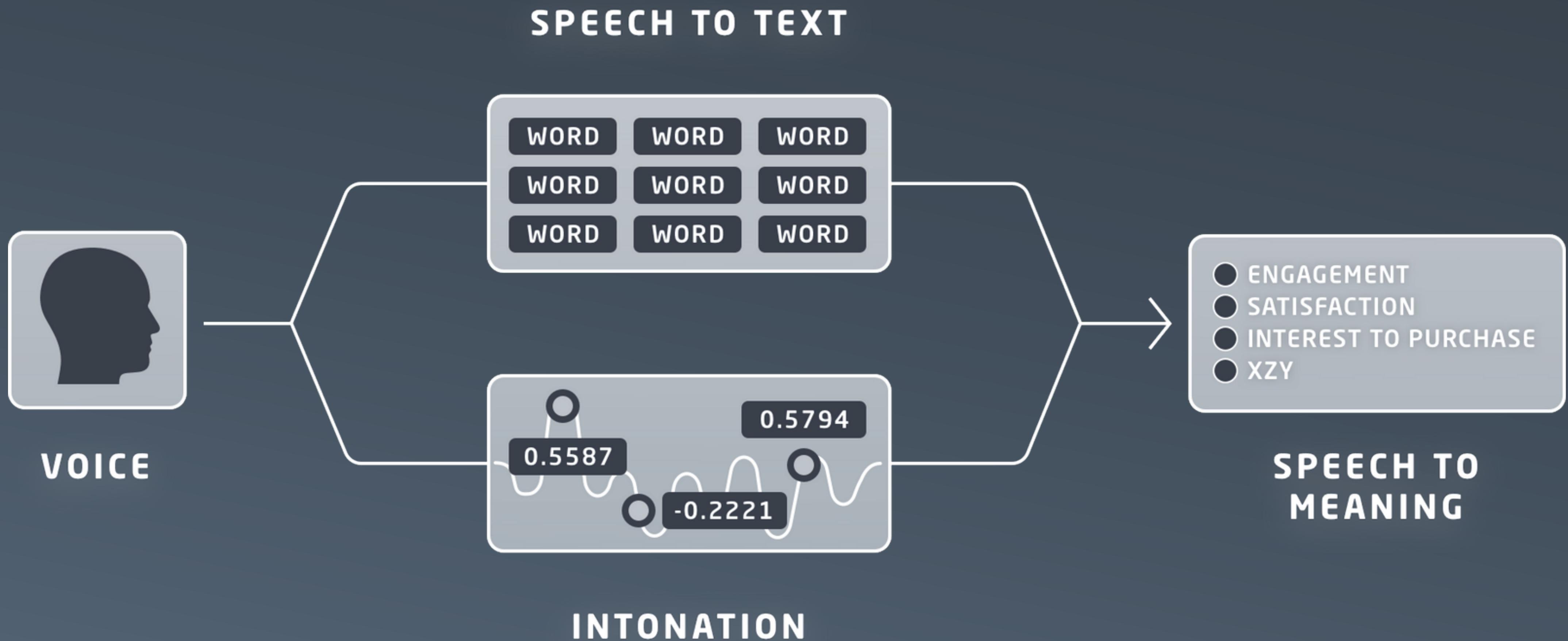
- Cultural differences
- Inter-annotator agreement
- Difficult to get labelled data

Intonation is an untapped gold mine

Contact centres: 75B calls/year



Acoustic Language Processing

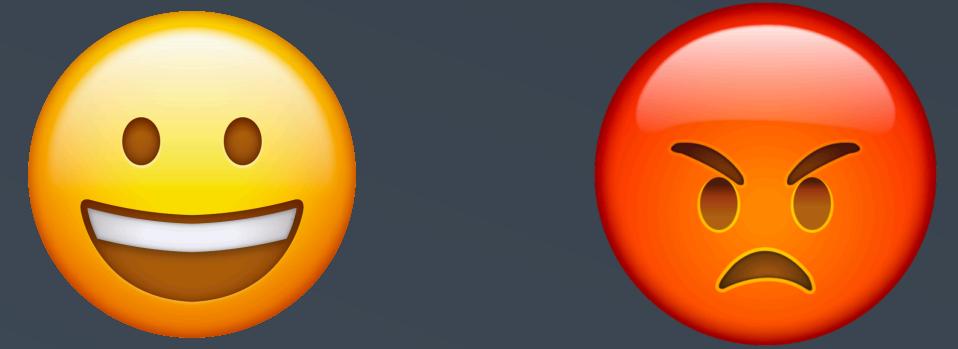


Custom feature extraction frontend + CNN/LSTM backend



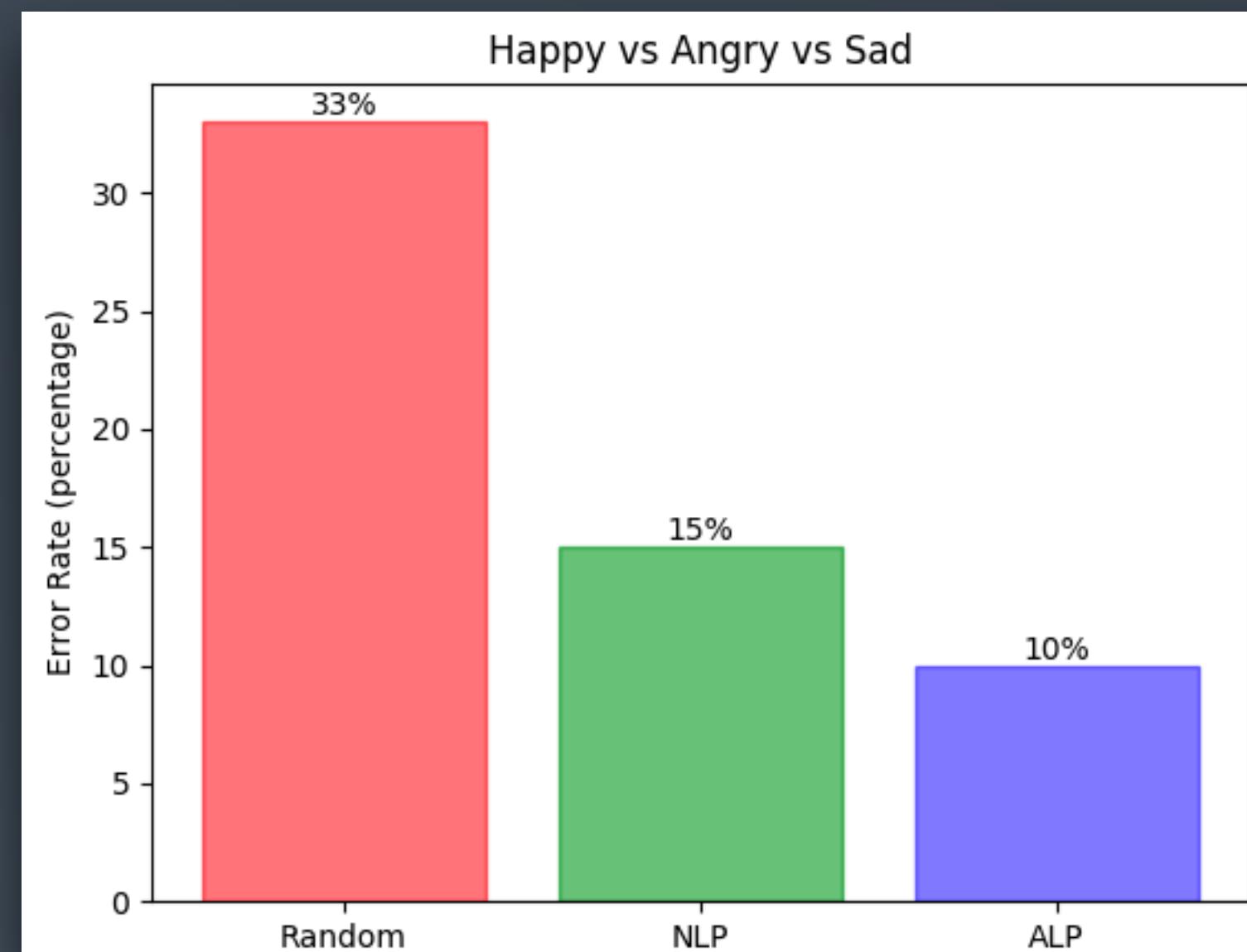
OTO Emotions data set

- 100k acted utterances
- 3k speakers
- 10 emotions



ALP results

- Including acoustics: 30-50% reduction in error rate
- 25% higher accuracy than SoTA features (OpenSMILE)
- Real-time prediction



Real-world applications

- Sales prediction: **94% accuracy** on acoustics alone
- Unsupervised **representation learning** on millions of calls
- **Transfer learning**: adaptation to use cases with little data
- Robust to differences in **languages, accents, contexts**

**Pushing the state
of the art with
more data &
better models**

OTO's live tone coach

The screenshot shows a software application window titled "WXYZ Certification". On the left, there is a sidebar with various prompts and instructions:

- "SimplePledge" section: "Volunteer pledges made simple" with a dropdown menu.
- "Agent Script" section: "Thank you for calling to support WXYZ this is Dolores."
- "May I have the spelling of your name, starting with your first name?"
- "Do you have a Middle Initial?"
- "Can I have the spelling of your Last Name?"
- "PLEASE TAKE ALL SUSTAINER AND PAYMENT UPDATES!!"

The main area is titled "Caller Information" and contains fields for First Name, MI, Last Name, Title, and Suffix. It also includes a checkbox for "Second name if provided" and a note about caller requested additional names. A large green circular button with a microphone icon is prominently displayed in the center-right of the main area. Below this is a "Search Members Module" section with fields for Zip Code, Phone Number, Member last name, and Member first name, along with "Search" and "Clear all fields" buttons. At the bottom, there is a table with columns for First Name, Last Name, First Name(second), Last Name(second), Phone Number, Address, Zip Code, Pledge amt, Pledge date, and Sustainer. A red status bar at the bottom left indicates "Phone Status: FAQs (WXYZ Certification) (248)" and "DoloresL". The footer of the window says "Version SP_02_17_2018 DoloresL; sP 64".

OTO Systems Confidential

OTO

Randomised controlled trial

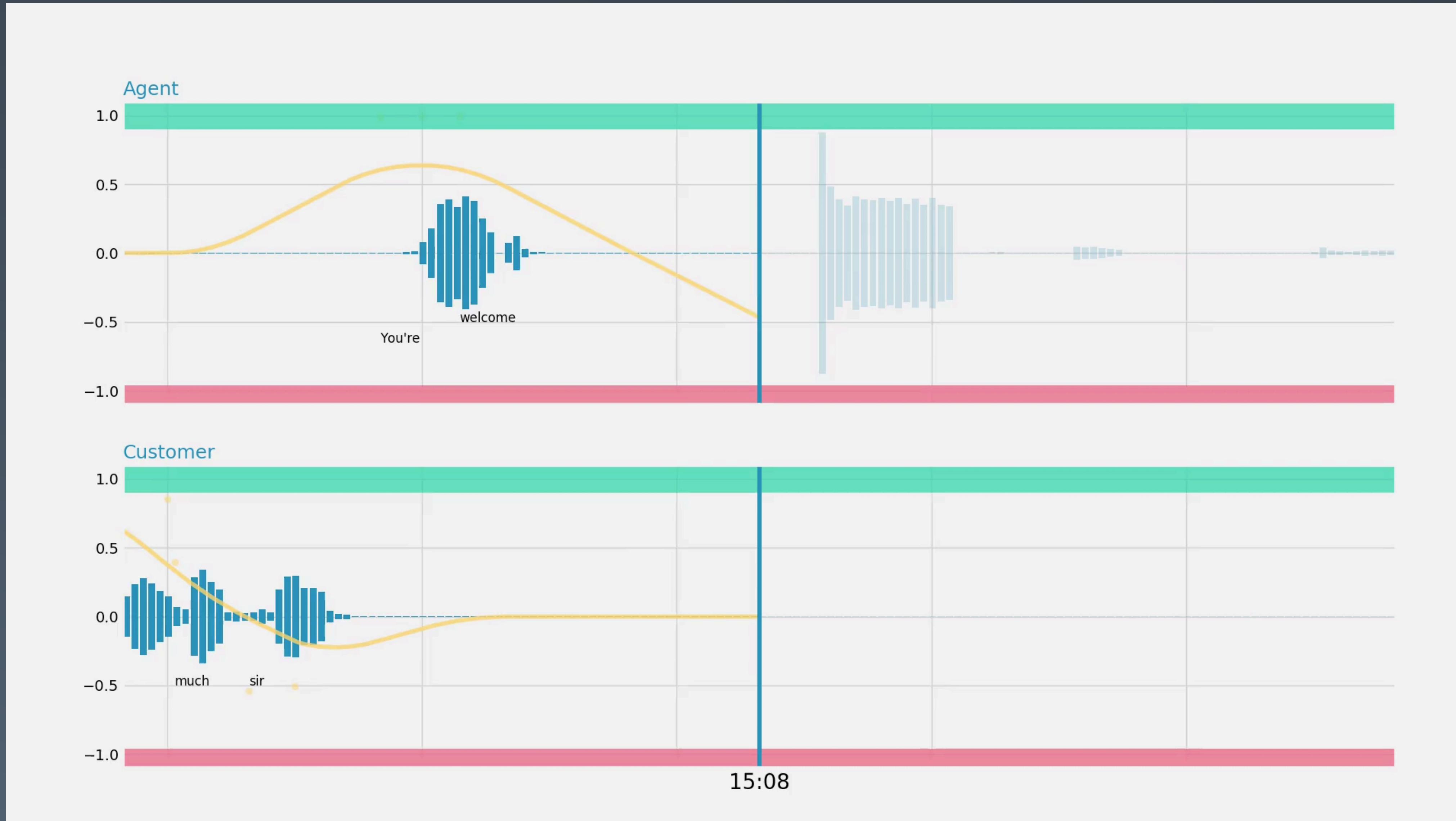
- 20 agents with/without
- 1 month, 20k calls
- +20% engagement
- +5% conversion rate

Enterprise deployment

- 100s of agents
- 200k calls
- +30-50% engagement
- +6% conversion rate
- 50% utilisation rate

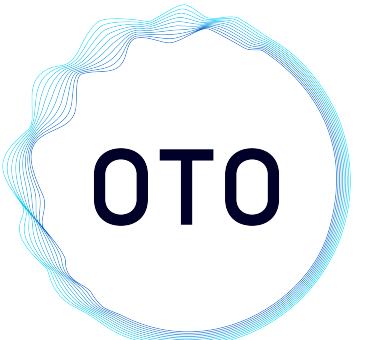


Deploying ALP at scale

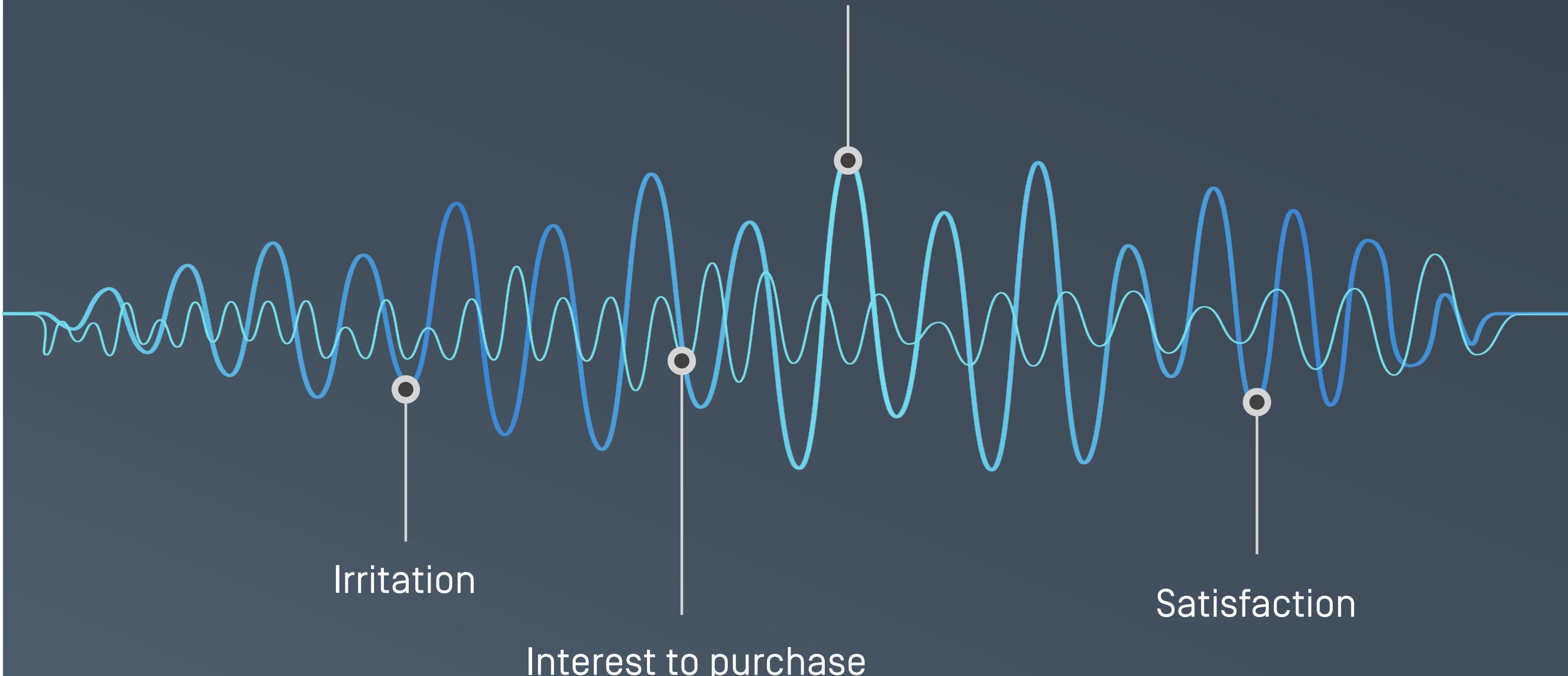


TL;DL

1. Including intonation raises the bar for natural speech understanding
2. Combining custom feature extractor + DL backend: new SoTA on emotion classification
3. Learned representations are transferrable to business data
4. ALP framework scales to real-time prediction on 1000s of streams
5. Interested? We're hiring!



Speech-to-Meaning





Decoding human conversations

