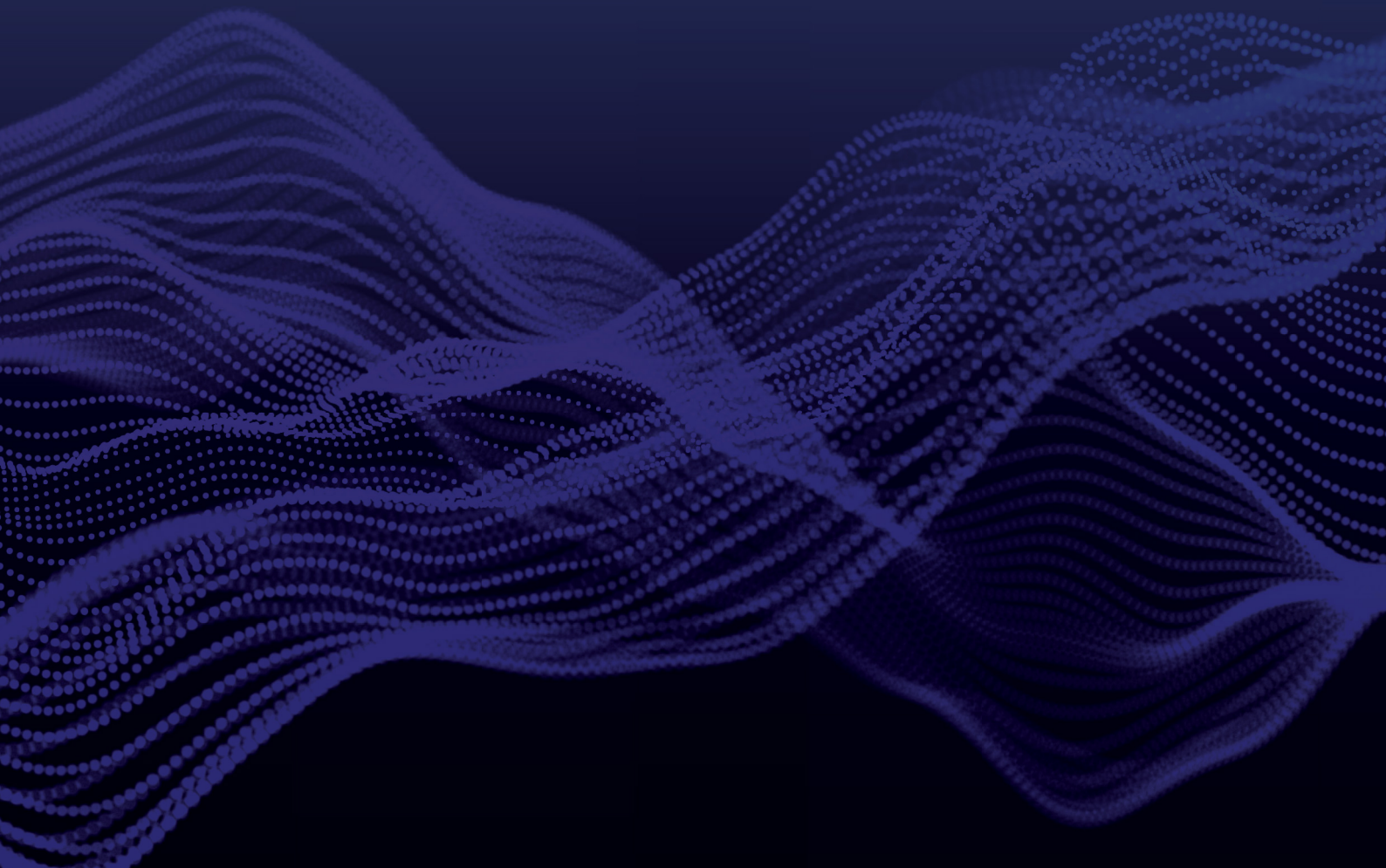
WHITE PAPER

voicesense

Technology Overview



Overview

The unique technology of Voicesense offers predictive behavioral analytics based upon acoustic speech analysis.

The core technology, developed after years of research, links an individual's speech patterns to typical behavioral patterns. By analyzing a person's vocal patterns in real-time, the system can determine the person's typical behavioral tendencies and predict the person's probability for future behaviors.

The technology is patented worldwide (granted). Its main elements include unique acoustic signal processing (speech prosody) and algorithmic machine learning (artificial intelligence).

The operational system provides automatic, remote, language independent and real-time analysis, and supports a combination of cloud, on-premise and mobile architectures. The system is compliant with various regulative standards including ISO, PCI, HIPAA and GDPR.

The commercial uses of the technology include three main industries: Healthcare—remote tracking and screening of depression and other mental health states; Fin-tech—predicting future risk for loan default, collections default and insurance claiming; Enterprise—predicting probability for consumer future behaviors (online buying, buying style, churn) and for employee future behaviors (fit to job, performance, burnout, attrition).

The technology was successfully validated in clinical trials to reflect mental-health states and in various enterprise case studies to show proof of concept in predicting consumer, employee and financial behaviors (e.g. financial default, fraud, employee burnout & attrition, personality & working profile, well-being, customer dissatisfaction and more).

Voicesense Ltd.
www.voicesense.com

Behavioral speech analytics – scientific background and novel approach

Voicesense introduces a new speech analysis approach–Behavioral Speech Analytics. This approach is focused on identifying the speech patterns that are associated with typical behavioral tendencies of individuals, across different languages, cultures or speech content.

The two most researched approaches to speech analysis include: 1) speech recognition and 2) speech prosody. Speech recognition focuses on understanding the content of speech–what is being said–i.e. words, sentences, and meaning. By definition, speech recognition is language dependent and focuses on the current content being spoken.

Speech prosody, however, focuses on the acoustics of the sounds–intonation, rhythm, emphasis, pronunciation, accent, attitude and emotions, rather than the actual spoken content. Prosodic speech analysis is indeed language independent to some extent; nevertheless, its traditional use still has a major limitation. It usually focuses on the current moment–how does the speaker sound right now. Behavioral speech analytics aims to go much further and offers generic behavioral personalization through speech analysis. The idea is that people tend to use typical speech patterns that reflect typical behavioral patterns. Such behavioral patterns may reflect personality, behavioral tendencies, as well as speech patterns that are typical to certain situations or certain individuals.

These characteristic speech patterns can be generalized in two ways: the first generalization exists at the personal level, namely, a certain person would tend to use similar speech patterns over different situations, reflecting his/her personal behavioral tendencies. The second generalization goes beyond the individual level, to the behavioral level, namely, typical speech patterns that reflect common behavioral tendencies, beyond a specific individual or a specific situation. Behavioral speech analysis can be applied to both speech recognition and speech prosody. In speech recognition it means finding typical patterns of using specific words, word-combinations, sentence length or sentence structure, that characterize specific individuals or specific behavioral patterns. Within speech prosody such behavioral speech analysis means finding typical intonation, rhythm, pace, emphasis patterns that characterize specific individuals or specific behavioral patterns. Given its language independence, and its physiologic nature (related to the human speech production physiological system), speech prosody seems to offer wider generalization potential for such behavioral speech analysis.

Bottom line: Behavioral speech analysis offers personalization, and as such, it also offers predictive analytics. If we can assess the typical behavioral patterns characterizing a certain person, we can anticipate, fairly accurately, the person's behavior in various life aspects—work, social, health, consuming, entertainment, etc. since our daily decisions in all these areas lean strongly on our behavioral tendencies.

Consequently, this concept of Behavioral Speech Analytics brings significant value to the world of big data and general predictive analytics methods. Existing predictive analytics methods still rely heavily on demographics and user history. Adding the behavioral angle by behavioral speech analysis can drastically improve predictions. For example, banks have a lot of data about their customers—financial strength, credit score and so on. However, they have very little knowledge about the customers' behavioral tendencies; for example: Do they tend to take risks? Are they impulsive? What is their personal integrity? And so on, all factors that play a crucial role in determining a client's financial risk and financial behavior, perhaps even more than their financial strength.

At Voicesense, we developed a prosodic based behavioral speech analytics system.

Technology description

The heart of Voicesense technology is the acoustic vocal pattern analysis of a person's natural speech. It measures non-content speech aspects such as intonation, pace and emphasis (aka 'speech prosody'). There is no comprehension of what is being spoken, hence the analysis is completely content free.

Prosodic features are universal by nature—they reflect the physical aspects of our speech pronunciation. To a certain extent, prosodic parameters are language, culture, and gender independent. Voicesense analysis was indeed tested successfully for both genders in different languages, including Western and Asian languages.

Vocal analysis

The vocal analysis first calculates over 200 raw voice parameters per second from the samples of each audio recording.

These raw parameters consist of a wide range of acoustic feature segmentations, including lengths, ranges, slopes, frequencies, values and shapes of pitch extracted parameters, amplitude extracted parameters and silences extracted parameters within the speech recording. Thousands of datapoints are calculated and averaged for each recording to form the over 200 parameter dataset that reflects the individual's personal speech patterns in the given recording.

The raw parameters are then calibrated and normalized to overcome possible biasing effects within the specific recording as a result of amplitude differences, pitch differences, speech type differences (conversation or monologue), gender differences and age differences. The calibration process is performed against a large natural-speech vocal reference datasets – tens of thousands of recordings that were collected by Voicesense over time, covering different speech types, genders, ages and languages.

These calibrated and normalized parameters create the input data for generating the company's predictive models. Machine learning techniques are used to select and weight the vocal parameters that best correlate with the searched behavioral phenomena. The process uses the common K-fold cross-validation method that randomly splits the dataset into training and test sub-samples and repeats the process for multiple iterations in order to reach a stable and reliable predictive model equation. The model provides unified speech-based scores that are associated with the predicted phenomena, e.g. depression probability, financial risk, consumer purchasing probability, employee attrition probability and so on.

Once a new predictive model is generated, it is incorporated into the operational real-time production system. The system enables instant predictive score calculation of streaming speech audio of customers, employees or patients in various real-life settings.

Operational system

The Voicesense operational system receives audio input from different optional sources– enterprise call center, self-recording (mobile or web), mobile conversation recording, and any recorded audio through an API upload.

The system performs the vocal analysis on the audio input, whether in real-time or offline, generating the predictive scores.

The predictive scores, coupled with recommendation guidance and feedback are then provided to users through various optional output means—real-time displays, dashboard, offline reports, API protocols, all according to the specific system use case and configuration.

The operational system supports different optional architectures—as a software cloud service (via API that uploads audio and returns predictions); as an on-premise real-time system—iintegrates with call center PBX to receive audio streaming and provides displays, database and reports; as a hybrid cloud-on premise solution—collecting audio locally on prem and computing predictive scores over the cloud; as an integrated mobile-cloud solution—collecting audio in the mobile-phone and computing predictive scores over the cloud.

Intellectual property

Voicesense has two sets of granted patents. The first set was granted for our analysis method of measuring emotion in speech (sentiment).

The second set was granted for behavioral profiling through speech analysis. It consists of two parts: 1) creating a knowledge base of relations between speech patterns and behavioral tendencies; 2) conducting behavioral speech analysis according to the knowledge base of relations between speech patterns and behaviors. In the USA and Europe the two parts of the patent were separated into a main patent and a continuation patent.

Sentiment patent set:
The measurement of emotion in speech (second patent set) was granted in the following countries:

- USA (US7606701B2)
- Europe (EP1423846B1):
- UK
- Germany
- France
- Spain
- Italy
- Ireland
- Sweden
- Finland
- Canada (CA 2456625)
- India (223756)
- Israel (14481801)4818A

Following are the patent links from the US and the European patent offices:

https://patents.google.com/patent/US7606701B2/en?oq=US7606701B2
https://patents.google.com/patent/EP1423846B1/en?oq=EP1423846B1

Behavioral profiling patent set:
The measurement of emotion in speech (second patent set) was granted in the following countries:

- USA (knowledge base patent) - US8195460B2
- USA (analysis patent) - US8682666B2
- Europe (knowledge base patent) - EP2304718B1
- UK
- Germany
- France
- Spain
- Italy
- Turkey
- Europe (analysis patent) – in final stages of approval
- China (CN2009801320343)
- India (IN331877)
- South Korea (KR1020117001201)
- Israel (IL 209996)

Following are the patent links from the US and the European patent offices:

https://patents.google.com/patent/US8195460B2/en?oq=US8195460B2
https://patents.google.com/patent/US8682666B2/en?oq=US8682666B2
https://patents.google.com/patent/EP2304718B1/en?oq=EP2304718B1

Regulation

Voicesense is certified for the highest data security ISO standards:

- ISO 27001 (ISMS) - Information security management system
- ISO 27799 - Information security in health care

Voicesense was reviewed for GDPR compliance by Kaleidoscope consultancy (UK) and was found to be generally compliant. Voicesense was reviewed for PCI compliance and has a white paper approving product compliance to PCI. Voicesense software was developed in accordance with medical HIPAA standard. The company completed preparation for European CE health certification. Formal registration is expected this year. The company has a letter of opinion approving market readiness for FDA (no need for regulated product).

Voicesense was reviewed for PCI compliance and has a white paper approving product compliance to PCI. Voicesense software was developed in accordance with medical HIPAA standard. The company completed preparation for European CE health certification. Formal registration is expected this year. The company has a letter of opinion approving market readiness for FDA (no need for regulated product).

Technological validation

Voicesense successfully completed two clinical trials validating its ability to link speech patterns to mental health states, including Depression, Schizophrenia, Anxiety, ADHD, all with highly statistical significance.

The clinical trials were conducted in the Neuropsychiatric center in Hamburg, Germany, lead by Dr. Peter Tonn, and in the Beer Yaakov mental health center in Israel, lead by Prof. Yechiel Levkovitz. Both studies are in the process of being published.

Within the enterprise and financial industries, Voicesense conducted several large-scale validation studies, each one consisting of thousands of subjects. These studies validated with high statistical significance the technology's predictive capabilities for the following use cases:

- Loan default prediction - US, Israel, and Guatemalan case studies
- Collection default prediction – US and Brazil case studies
- Insurance claim prediction – Australian case studies
- Employee attrition prediction – UK case study
- Candidate fit screening – US case study
- Buying style prediction – US and Israel case study
- Customer dissatisfaction monitoring – US and Israel case study
- Personality attributes classification – US and Israel case study