



pindrop<sup>®</sup>

FEATURE  
REPRESENTATION OF  
AUDIO SIGNALS AND  
APPLICATIONS

---

Nick Gaubitch  
*Head of Research, EMEA*

## Overview



Part I: Pindrop overview and the world of telephony fraud

Part II: Introduction to feature representation of audio/speech signals

Part III: Advanced audio features: Mel-frequency Cepstral Coefficients (MFCCs)

Part IV: Putting it all together: a basic speaker identification system

## Learning Objectives



To link material from course(s) to practical examples

To understand audio representation in the feature space

To gain some understanding of the use of classification/clustering in audio processing

To appreciate real-world problems and solutions using different audio signal processing tools

2018 Pindrop® 3

## Overview



Part I: Pindrop overview and the world of telephony fraud

Part II: Introduction to feature representation of audio/speech signals

Part III: Advanced audio features – Mel-frequency Cepstral Coefficients (MFCCs)

Part IV: Putting it all together – a basic speaker identification system

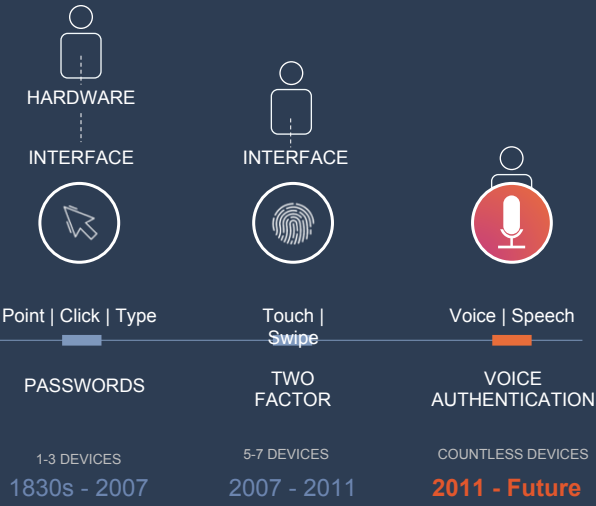
2018 Pindrop® 4

The infographic illustrates the evolution of user interfaces through three stages:

- Stage 1 (Left):** Hardware-based interface. Interaction methods: Point | Click | Type. Security: Passwords. Device usage: 1-3 devices (1830s - 2007).
- Stage 2 (Middle):** Touch-based interface. Interaction methods: Touch | Swipe. Security: Two Factor. Device usage: 5-7 devices (2007 - 2011).
- Stage 3 (Right):** Voice-based interface. Interaction methods: Voice | Speech. Security: Voice Authentication. Device usage: Countless devices (2011 - Future).

The Pindrop logo is located in the top right corner of the infographic.

# THE CONVERSATION HAS SHIFTED TO VOICE.



2018 Pindrop® 5

The infographic features two statistics on a textured background:

- Statistic 1:** 61% OF YOUR CURRENT FRAUD IS ORIGINATING IN YOUR CALL CENTER. Source: AITE ANALYST GROUP.
- Statistic 2:** 20% OF YOUR FUTURE REVENUE OPPORTUNITY DEPENDS ON FRICTIONLESS IDENTITY CORROBORATION. Source: GARTNER.

The Pindrop logo is located in the top right corner of the infographic.

61% OF YOUR CURRENT FRAUD IS ORIGINATING IN YOUR CALL CENTER

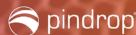
AITE ANALYST GROUP

20% OF YOUR FUTURE REVENUE OPPORTUNITY DEPENDS ON FRICTIONLESS IDENTITY CORROBORATION

GARTNER

650M  
CALLS  
ANALYZED  
EVERY  
YEAR

17 Patents Granted and Pending



PINDROP'S MISSION

Provide real time identity, security, and trust on every voice interaction

Employees

300+ Employees  
50 PhD specializing in machine learning and audio processing

Customers

70% of our customers are Fortune 500, from every industry

Innovation

Phoneprinting™ Technology  
Deep Voice™ Biometrics Engine  
Toneprinting™ Technology

2018 Pindrop® 7

THE NEW  
STANDARD FOR  
VOICE  
AUTHENTICATION  
AND SECURITY.

Securing the most sensitive organizations in the era of voice.




5 of the top

8 of the top 10

3 of the top 5

Insurance Retail Government Banks Hospitality Telc Stock Brokers

2018 Pindrop® 8



# THE BALANCING ACT

Why enterprises struggle to stop treating customers like criminals

**SATISFACTION**  
10-30% of Customers  
Unable to Authenticate

**CUSTOMER EXPERIENCE**

AUTHENTICATION

**FRAUD**  
1 in 937 Calls is  
Fraud Related on Average


**SECURITY ENFORCEMENT**

FRAUD

---

**OPERATIONS**  
\$0.30 - \$1.00 Added Cost Per  
Call to Authenticate

2018 Pindrop® 9



# WHAT'S DRIVING VALUE?

Financial | Insurance | Telco | Government | Retail | Hospitality ...

**PASSIVE AUTHENTICATION**  
Up to \$1M savings for every 1M calls pre-authenticated

**+30%**

Authentication

**60 sec.**

AHT Reduction

**CONTAINMENT**

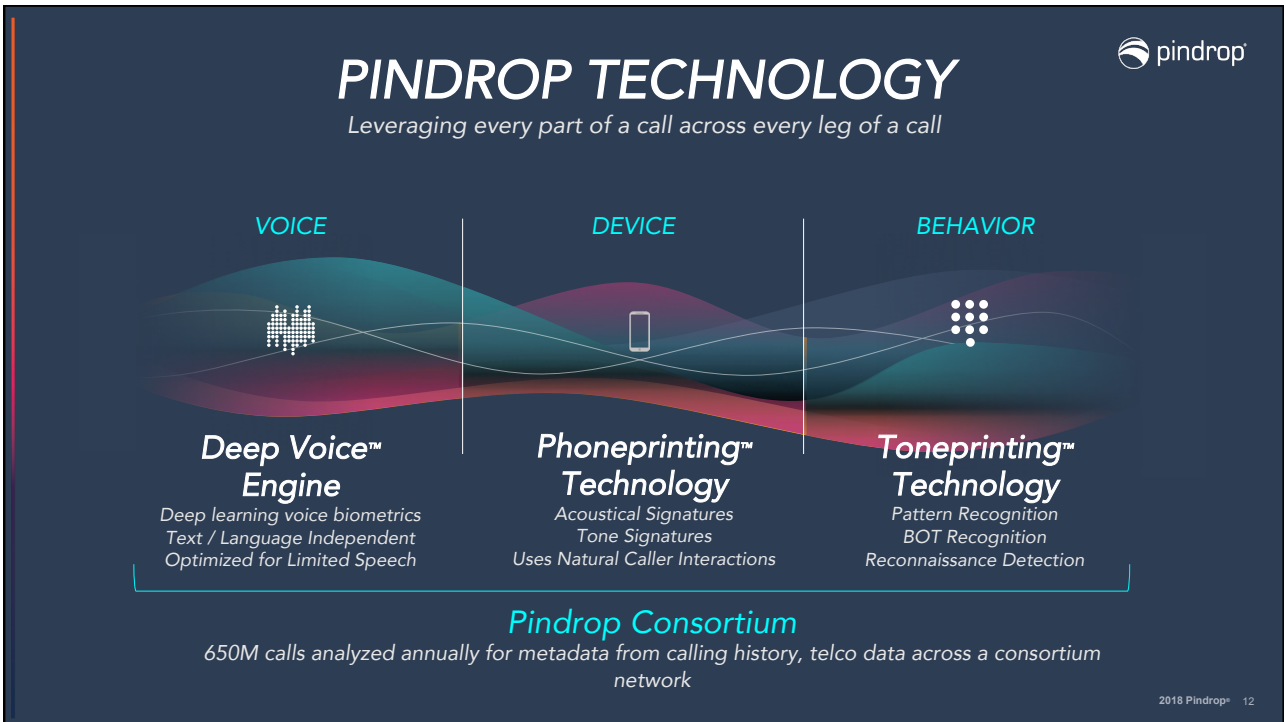
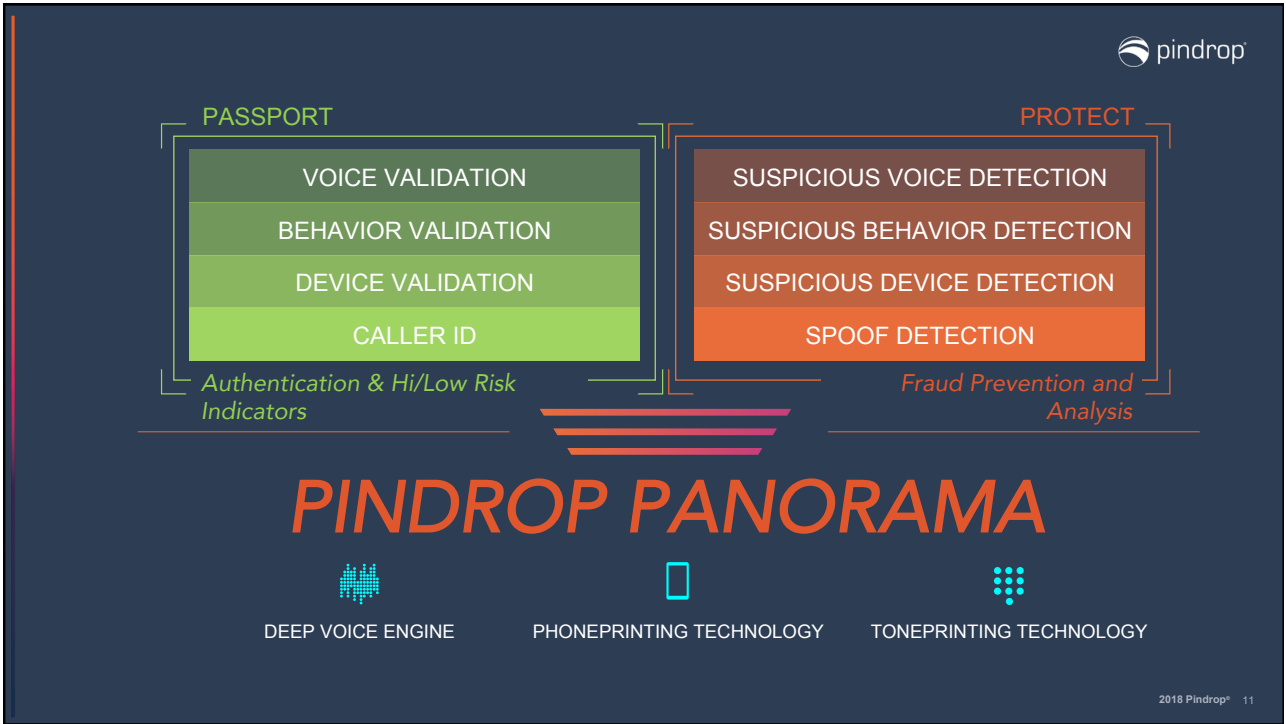
Up to \$4.5M for every 1m calls that are contained in the IVR


**FRAUD PREVENTION**  
Up to \$5.8M in fraud loss for every 10M calls

**80%**




Fraud Detection

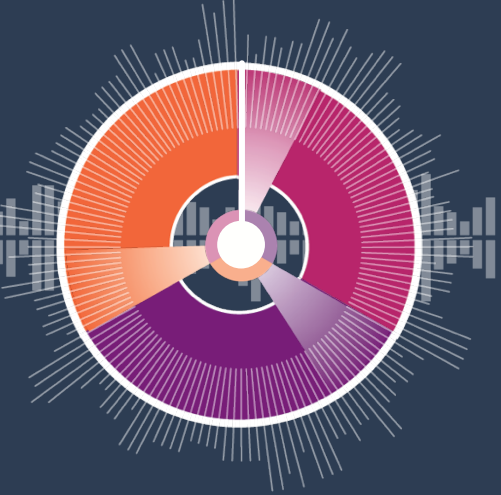
2018 Pindrop® 10








**Phoneprinting™ Technology** 

1480 FACTORS analyzed to create a distinctive telephony signature

SPECTRUM	LOSS	NOISE
 <ul style="list-style-type: none"> <li>Quantization</li> <li>Frequency filters</li> <li>Codec artifacts</li> </ul>	 <ul style="list-style-type: none"> <li>Packet loss</li> <li>Robotization</li> <li>Dropped frames</li> </ul>	 <ul style="list-style-type: none"> <li>Clarity</li> <li>Correlation</li> <li>Signal-to-Noise Ratio</li> </ul>



**PHONEPRINTING RESULTS**

TRUE CHARACTERISTICS	HISTORY	PROFILE
 DEVICE TYPE  GEO-LOCATION  CARRIER	 KNOWN FRAUD MATCH	 UNIQUE DEVICE
Detect Fraud		Authenticate

2018 Pindrop® 13



# DEEP VOICE ENGINE

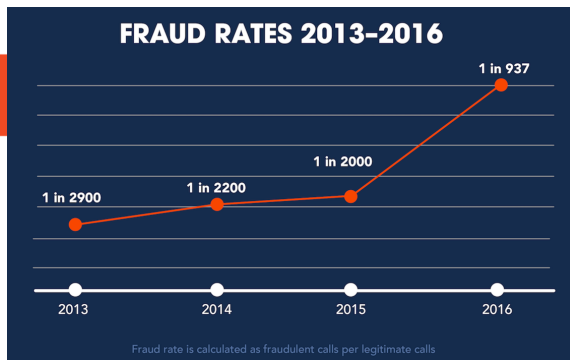
27% higher accuracy compared to competitive solutions

- Passively analyzes a caller's voice for fraud detection and authentication.
- Powered by patented deep neural networks
- Requires nominal speech
- Resilient to voice synthesis, voice morphing, and replay attacks
- Robust to noise, channel, language, & voice aging
- Developed by Pindrop researchers & tested on 20,000+ speakers

HISTORY	PROFILE
 KNOWN FRAUD MATCH	 UNIQUE VOICE
Fraud Detection	Authentication

2018 Pindrop® 14

## Global Fraud Call Trend



THE INCREASE IN  
GLOBAL FRAUD  
CALL RATE 2016-2017

113%

2018 Pindrop® 15

## Fraud Vectors

### DATA DEALING

- ✓ Collect and repackage profiles for sale
- ✓ Purchase in bulk months ahead of breach announcements
- ✓ Not the same fraudster attacking call centers

### SOCIAL ENGINEERING

- ✓ Vishing: The practice of impersonating a legitimate caller in order to elicit information or influence action over the telephone
- ✓ Psychology: Fear, Intimidation, evoke flight or fight reflex
- ✓ Distraction: Diverts agents attention with annoyance, complex requests, or loud background noises
- ✓ Transfer: Simple technique, lets agent vouch for caller

### RECONNAISSANCE

- ✓ Assume fraudsters have a perfect working knowledge of your call center
- ✓ Fraudsters understand weaknesses in anti-fraud procedures
- ✓ Fraudsters call an average of 5 times into the call center before making a transaction or any changes to the account

### VOICE MORPHING

- ✓ Masks fraudster's identity
- ✓ Ages up or down to match target account
- ✓ Changes gender to add legitimacy to the scheme
- ✓ Used to bypass voice biometric solutions

2018 Pindrop® 16



## Attack Characteristics



### WHAT FRAUDSTERS DO

- Information phishing / confirmation
- Address / telephone number changes
- Credit line creation
- Unblocking online access / telephony security numbers
- Replacement debit / credit cards
- Open additional accounts
- Order products
- Change delivery address

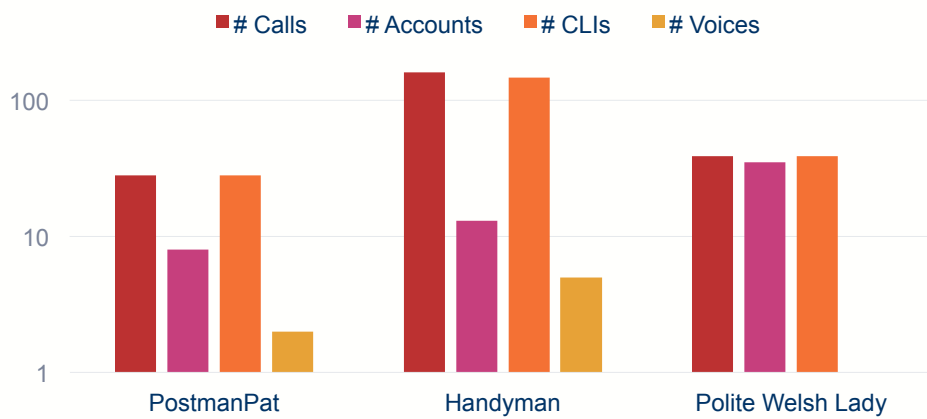


### HOW FRAUDSTERS DO IT

- Voice disguise (manual / software)
- Hiding phone numbers  
70-80% of fraudulent calls have restricted CLI
- Spoofing phone numbers  
[SpoofCard](#)
- Mail intercept
- SIM card take-overs
- 3-5 calls before actual attempt

2018 Pindrop® 17

## Prolific Fraudsters



2018 Pindrop® 18

## Video: Social Engineering



2018 Pindrop® 19

## Overview



Part I: Pindrop overview and the world of telephony fraud

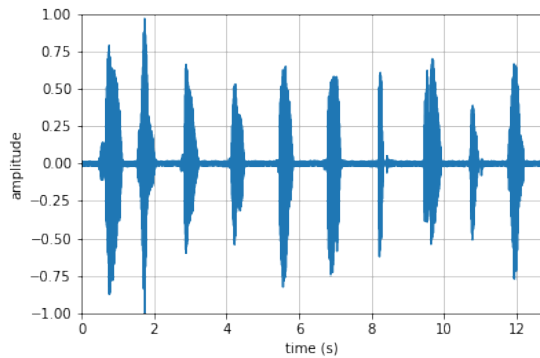
Part II: Introduction to feature representation of audio/speech signals

Part III: Advanced audio features: Mel-frequency Cepstral Coefficients (MFCCs)

Part IV: Putting it all together: a basic speaker identification system

2018 Pindrop® 20

## Sampled signal in the time-domain



- Sampling rate = 8kHz
- Good for a human listener
- *Not informative* for a machine
- *Not compact*
  - 8000 numbers (samples) / second

2018 Pindrop® 21

## Feature representation of audio: motivation



- Why feature representation?
  - more *informative* than sample values
  - more *compact* than sample values
- Many [machine learning] applications use feature representations of audio
  - Speech recognition
  - Speaker identification
  - Speech quality/intelligibility estimation
  - Identifying parameters of a recording
    - what codec was used?
    - what was the acoustic scene?
  - Music search/indexing
  - Music transcription

2018 Pindrop® 22

## Feature representation of audio: basics



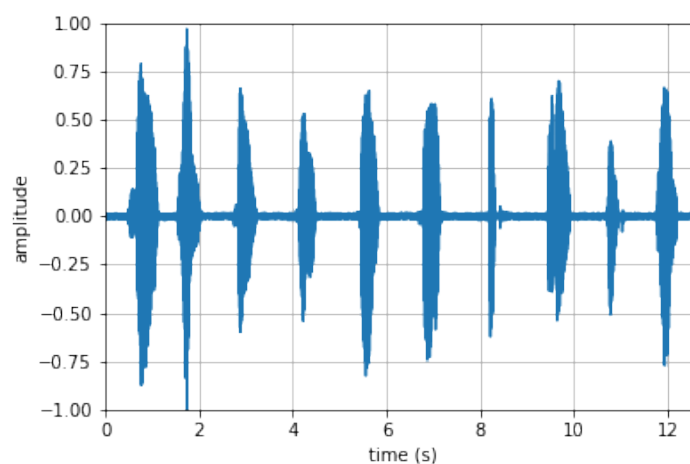
- Features vary from very simple (low-level signal features)...
  - e.g. short-term energy
- ...to advanced
  - e.g. Mel-Frequency Cepstral Coefficients – MFCCs
- Representing an audio signal in terms of  $K$  features  $F \downarrow 1 \dots F \downarrow K$ 
  - Collection of features – *feature vector* -  $F$
- Generic approach to obtaining features
  - Divide signal into short and possibly overlapping segments (typically 10-30ms)
  - For each segment calculate some entity of the audio
  - (Optional) calculate statistics of the short-term features for a longer super-frame
    - Mean, variance, skewness, kurtosis

2018 Pindrop® 23

## Feature representation of audio by example



- What portions are Voice Activity and what are Noise/Silence?



2018 Pindrop® 24

## Short-term processing



- a speech signal  $s(n)$  is divided into short frames for all feature calculations

$$s_{\ell}(n) = s(n + \ell R)w(n), \quad n = 0, \dots, N - 1$$

- $R$  – hop length
  - $N$  – frame length
  - $w(n)$  – windowing function
  - $\ell$  – the frame index
- 
- $N$  is often in the range 10-30ms and
  - Hop length typically 25 - 75% of the frame length
    - in the VAD example we use 32ms with 8ms hops  $\rightarrow$  256 samples and 64 samples at 8kHz
    - rectangular window

2018 Pindrop® 25

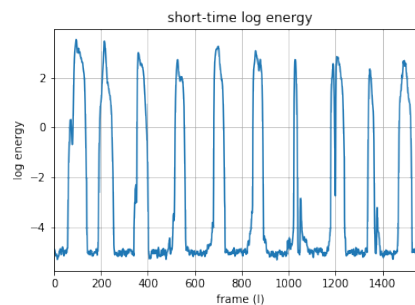
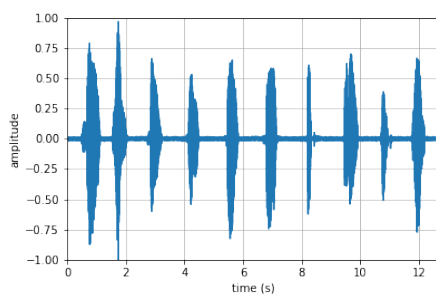
## Log-energy



- The log-energy of a frame is calculated as

$$E_{s,\ell} = \log \left( \frac{1}{N} \sum_{n=0}^N s_{\ell}^2(n) \right)$$

- energy of speech is greater than energy of noise



2018 Pindrop® 26

## Zero-crossing rate

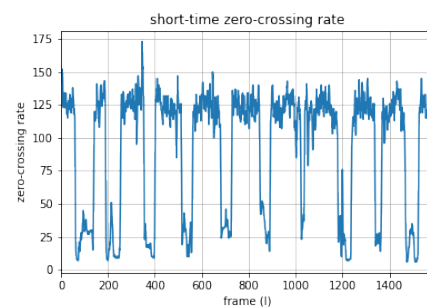
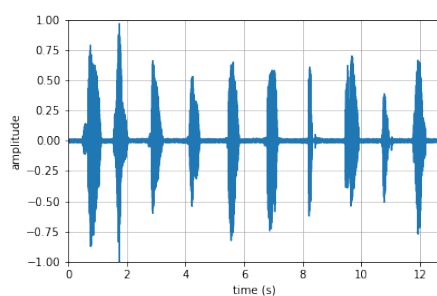


- Number of times the signal crosses the zero line

$$ZCR_{\ell} = \sum_{n=1}^N I\{s_{\ell}(n)s_{\ell}(n-1) < 0\}$$

$$I\{a < 0\} = 1 \text{ if } a < 0 \text{ and } 0 \text{ otherwise}$$

- zero-crossing rate is greater for noise than for speech



2018 Pindrop® 27

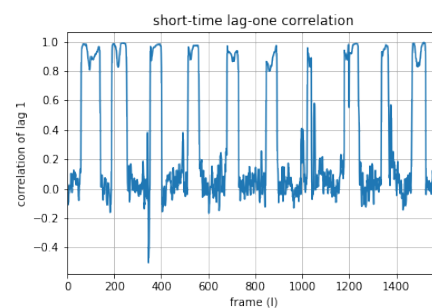
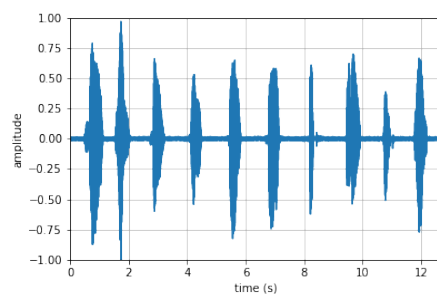
## Lag-one autocorrelation



- The signal correlated with itself delayed by one sample

$$\gamma_{\ell}(1) = \frac{\sum_{n=1}^N s_{\ell}(n)s_{\ell}(n-1)}{\sqrt{\sum_{n=1}^N s_{\ell}^2(n) \sum_{n=0}^{N-1} s_{\ell}^2(n)}}$$

- voiced sounds highly correlated
- noise not correlated



2018 Pindrop® 28

## What do we do with the features?



- **Threshold for each feature**
  - can be difficult to find universal threshold values
  - not scalable
- **Clustering**
  - No need for training data
  - ad-hoc decision rules
  - K-means; Gaussian Mixture Model (GMM)
- **Classification**
  - supervised learning by example – requires annotated training data
  - Gaussian Mixture Models (GMMs); Support Vector Machines (SVMs); Neural networks

Further reading: C. M. Bishop, "Pattern Recognition and Machine Learning," Springer, 2007

2018 Pindrop® 29

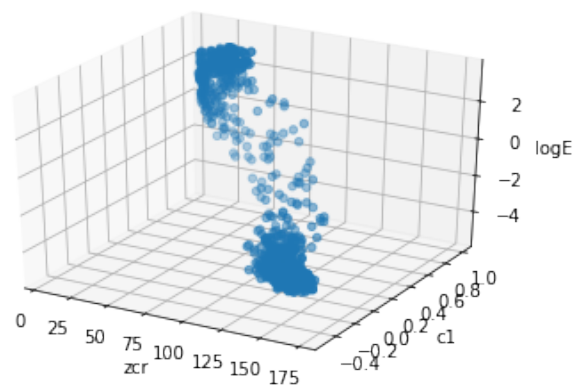
## Feature vectors



- Collecting all features for a frame in one vector

$$\mathbf{F} = [E_{s,l} \ ZCR_l \ \gamma_l(1)]$$

- in this case a 3-d feature vector
- two visibly well-defined clusters

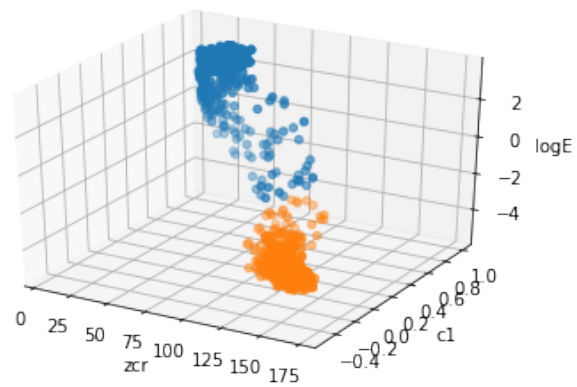


2018 Pindrop® 30

## Feature clustering

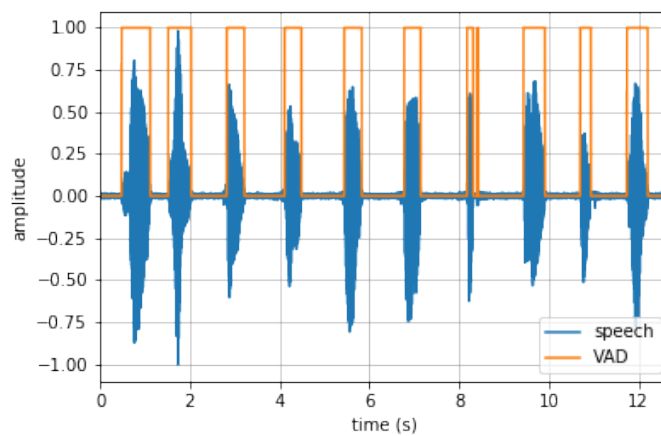


- Use a two-component GMM
- Basic VAD decision rule
  - Cluster with highest mean log energy is voice activity (1)
  - The other cluster is noise (0)



2018 Pindrop® 31

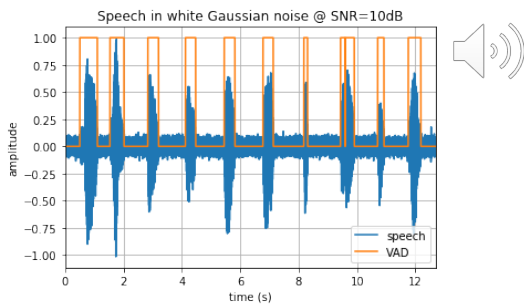
## Voice activity detection



2018 Pindrop® 32



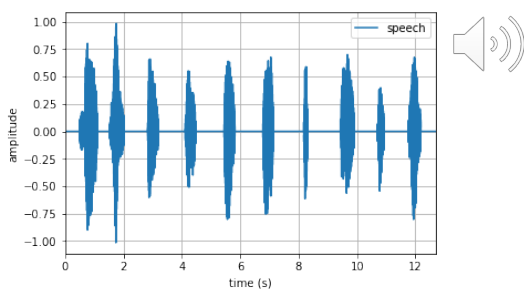
## Related application: Noise-gating



- Noise gating suppresses the portions that are not active voice

- Simple and not too pretty noise reduction but often used

- Some codecs don't transmit signal in silence/noise



2018 Pindrop® 33

## Summary



- *Pindrop* and the world of telephony fraud
- *Audio features* are a more compact and more informative representations of a speech/audio signal compared to sample values
  - designed depending on the application
- Threshold for each feature → not scalable so better use machine learning/ clustering
- *Voice activity detection* is a key component in most speech related systems

2018 Pindrop® 34

## Overview



Part I: Pindrop overview and the world of telephony fraud

Part II: Introduction to feature representation of audio/speech signals

Part III: Advanced audio features: Mel-frequency Cepstral Coefficients (MFCCs)

Part IV: Putting it all together: a basic speaker identification system

2018 Pindrop® 35

## Recap



- **Pindrop** and the world of telephony fraud
- **Audio features** are more *compact* and more *informative* representations of speech/audio signals compared to sample values
  - designed depending on the application
- **Threshold** for each feature → not scalable so better use machine learning/ clustering
- **Voice activity detection is a key component in most speech related systems**

2018 Pindrop® 36

## Simple vs advanced features



- Low-level features
  - single calculation based on some observation about the signal
  - for example, short-term energy
- Advanced features
  - series of calculations and/or transforms
  - algorithms to estimate some parameters
    - SNR, reverberation time, packet loss
  - related to more rigorous studies of the signal (speech production or perception)

2018 Pindrop® 37

## Example advanced features



- Perceptual spectral features
- Linear prediction
- Perceptual linear prediction (PLP)
- Modulation domain features
- Mel-frequency cepstral coefficients (MFCC)
- Linear-frequency cepstral coefficients (LFCC)

2018 Pindrop® 38

## MFCC



- Mel-Frequency Cepstral Coefficients
  - probably the most widely used feature representation for speech and audio
  - originally used for speech recognition
  - inspired by human hearing and speech production
  - ETSI standard for use in mobile phones
- Most [machine learning] examples from last week use MFCCs
  - Speech recognition
  - Speaker identification
  - Speech quality/intelligibility estimation
  - Identifying parameters of a recording
    - what was the acoustic scene?
  - Music search/indexing

2018 Pindrop® 39

## Short-term Fourier Transform

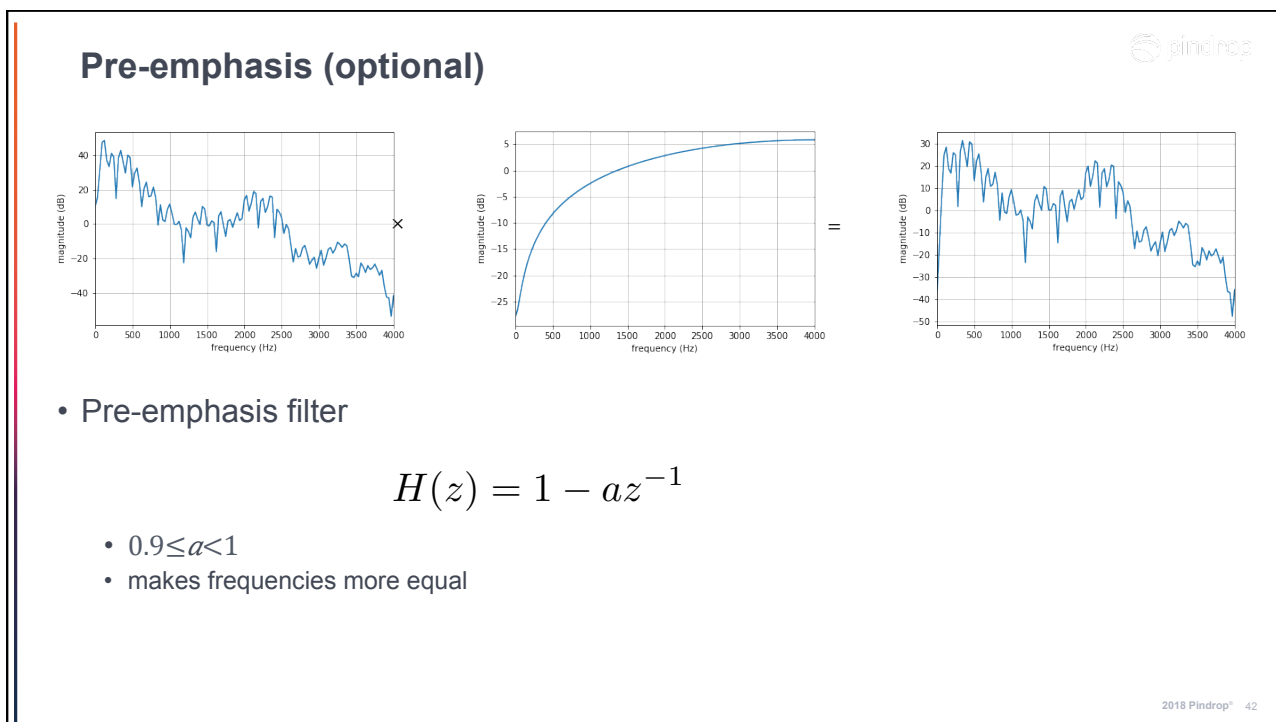
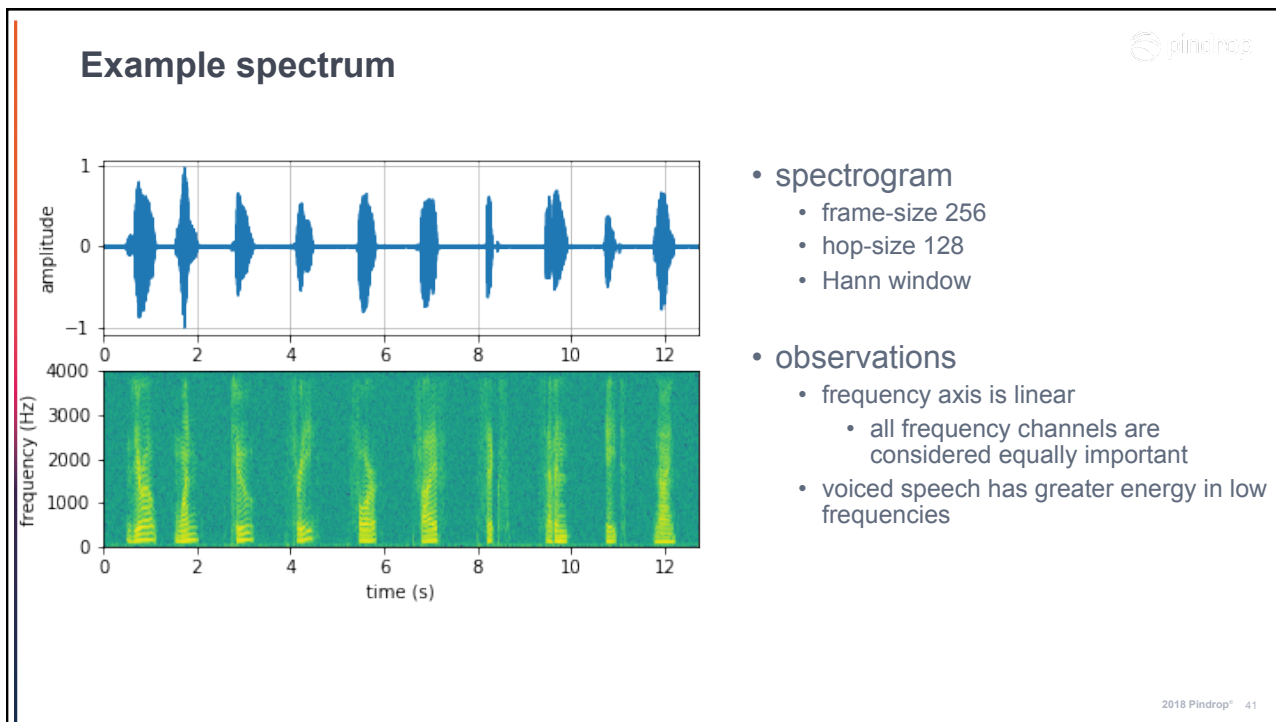


- a speech signal  $s(n)$  is divided into short frames for all feature calculations

$$s_\ell(n) = s(n + \ell R)w(n), \quad n = 0, \dots, N - 1$$

- $R$  – hop length
- $N$  – frame length
- $w(n)$  – windowing function
- $\ell$  – the frame index
- short-term Fourier transform (STFT)
 
$$S_\ell(k) = \mathcal{F}\{s_\ell(n)\} = \sum_{n=0}^{N-1} s_\ell(n)e^{-\frac{2j\pi kn}{N}} \iff s_\ell(n) = \mathcal{F}^{-1}\{S_\ell(k)\} = \frac{1}{N} \sum_{k=0}^{N-1} S_\ell(k)e^{\frac{2j\pi kn}{N}}$$
  - $k$  – the frequency index

2018 Pindrop® 40



## MFCC calculation



2018 Pindrop® 43

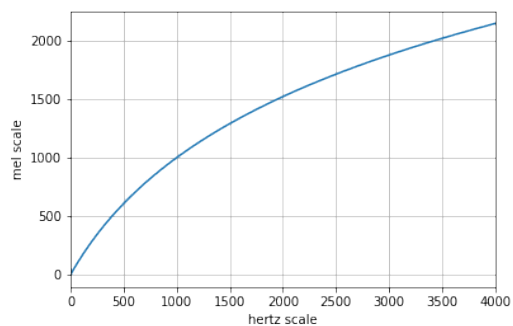
## Mel scale



- Mel scale – a perceptual scale based on perceived pitches to be equal
  - Human hearing is less sensitive to changes in high frequencies
- Relationship between Mel and Hertz

$$m(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)$$

$$f(m) = 700 \left( 10^{m/2595} - 1 \right)$$



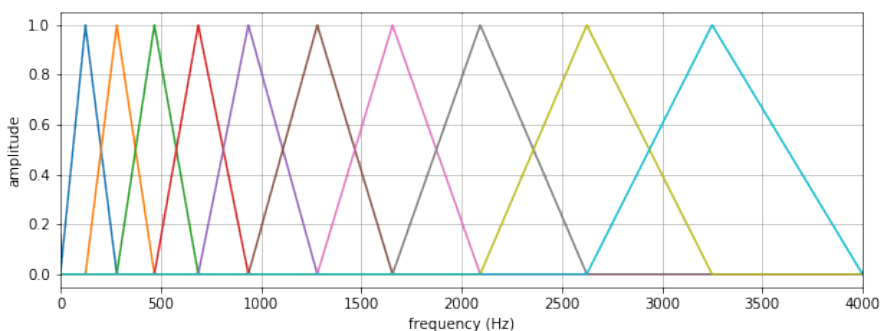
2018 Pindrop® 44

## Mel filterbank



- Implemented as a filterbank using triangular filters

$$H_m(i) = \begin{cases} 0 & i < f(m-1) \\ \frac{i-f(m-1)}{f(m)-f(m-1)} & f(m-1) \geq i \leq f(m) \\ \frac{f(m+1)-i}{f(m+1)-f(m)} & f(m) \geq i \leq f(m+1) \\ 0 & i > f(m+1) \end{cases}$$

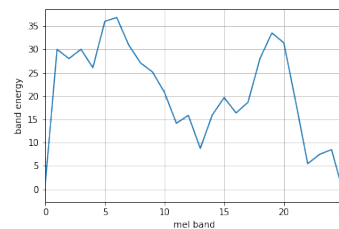
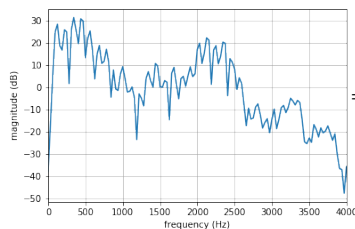
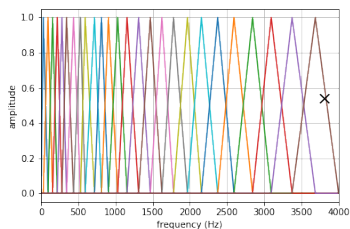


2018 Pindrop® 45

## Mel Spectrum



- Applying the mel-filterbank to the spectrum
  - typically 20-40 mel-bands (example below uses 26)
  - other shapes than triangular are possible



2018 Pindrop® 46

## MFCC calculation



2018 Pindrop® 47

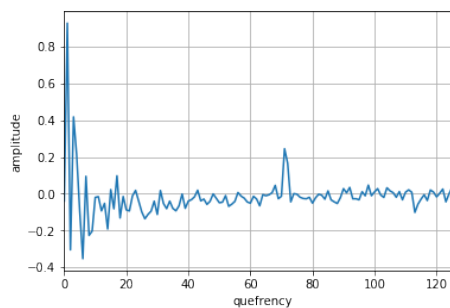
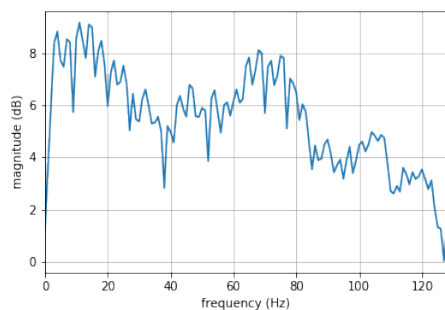
## Cepstrum



- the cepstrum is defined as

$$C_\ell = |\mathcal{F}^{-1}\{\log(|\mathcal{F}\{s_\ell(n)\}|^2)\}|^2$$

- $C_\ell$  has units *quefrequency* – a measure of time but not as time-domain  $n$



2018 Pindrop® 48

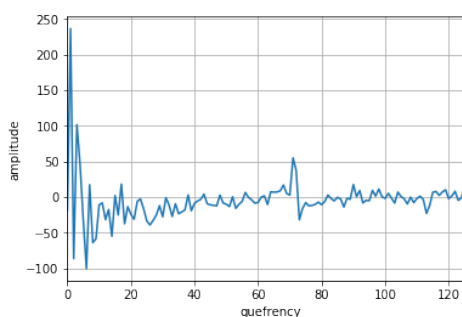
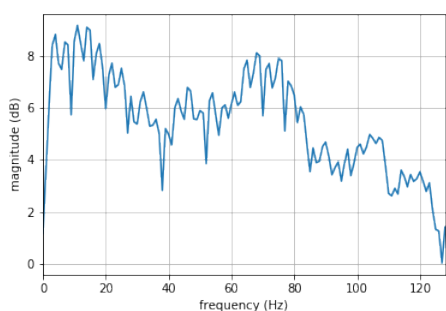


## Cepstrum with DCT



- the inverse Fourier transform is replaced by the discrete cosine transform (DCT)

$$S_\ell(k) = 0.5s_\ell(0) + \sum_{n=1}^{N-1} s_\ell(n) \cos \left[ \frac{\pi n}{N} (k + 0.5) \right]$$



2018 Pindrop® 49

## MFCC calculation



- feature vector  $c_{ll}$  obtained by selecting 10-20 lowest mel-cepstral coefficients as that is where most of the information is
- compact
- contains relevant information for speech
  - but seems to work well for many other audio applications

2018 Pindrop® 50

## Delta and Delta-Delta features



- In addition to the per-frame MFCCs we can calculate two more entities, delta and delta-delta coefficients
  - describe the dynamics of speech
- differential MFCC coefficients

$$\Delta_{\ell} = \frac{\sum_{i=1}^I i (\mathbf{c}_{\ell+i} - \mathbf{c}_{\ell-i})}{2 \sum_{i=1}^I i^2}$$

- acceleration MFCC coefficients

$$\Delta\Delta_{\ell} = \frac{\sum_{i=1}^I i (\Delta_{\ell+i} - \Delta_{\ell-i})}{2 \sum_{i=1}^I i^2}$$

2018 Pindrop® 51

## MFCC feature vector



- A typical output of feature vectors

$$\mathbf{F} = \begin{bmatrix} \mathbf{c}_0 & \Delta_0 & \Delta\Delta_0 \\ \vdots & \vdots & \vdots \\ \mathbf{c}_{\mathcal{L}-1} & \Delta_{\mathcal{L}-1} & \Delta\Delta_{\mathcal{L}-1} \end{bmatrix}$$

- for 12 MFCCs the feature vector will have 36 entries (if all delta and delta-delta features are utilized)
  - compare with an original frame size of 256
  - compact
  - informative

2018 Pindrop® 52

## MFCC in the real [acoustic] world



- MFCCs work very well when you have clean speech
- More often, speech signals are captured in noisy and reverberant environments
  - hands-free telephony
  - mobile telephony / laptops
- MFCCs are not robust to noise and reverberation
  - for example, try Siri at a couple of meters distance from the phone
  - Amazon Alexa uses 8 microphones to improve audio before recognition

2018 Pindrop® 53

## Summary



- advanced audio features tend to draw inspiration from studies in audio perception and production
- MFCCs are a good example of feature design that combines signal processing with knowledge from other fields of science
- MFCCs are fundamental in most speech related applications
- require extra processing to make robust to the acoustic environment

2018 Pindrop® 54

## Overview



Part I: Pindrop overview and the world of telephony fraud

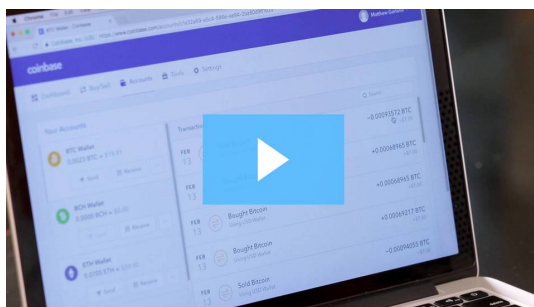
Part II: Introduction to feature representation of audio/speech signals

Part III: Advanced audio features: Mel-frequency Cepstral Coefficients (MFCCs)

Part IV: Putting it all together: a basic speaker identification system

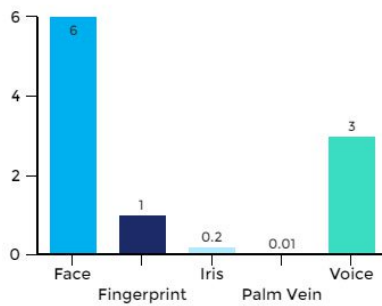
2018 Pindrop® 55

## Voice recognition video

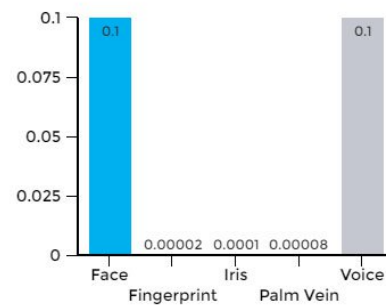


2018 Pindrop® 56

## Biometrics compared



False rejection rate



False acceptance rate

Source: <https://www.bayometric.com/biometrics-face-finger-iris-palm-voice/>

2018 Pindrop® 57

## Why voice recognition?



- It is a very hot topic in the audio world
- Many large EU banks are rolling out voice biometrics (verification) to replace passwords with the voice
- Will become even more important as voice control of devices increases
- Issues:
  - voice verification is not 100% accurate (think of MFCCs in noise)
  - you need a model for each customer which requires some form of enrolment
  - several attacks are possible
    - voice synthesis
    - recording of your voice [replay attack]

2018 Pindrop® 58

## Voice recognition



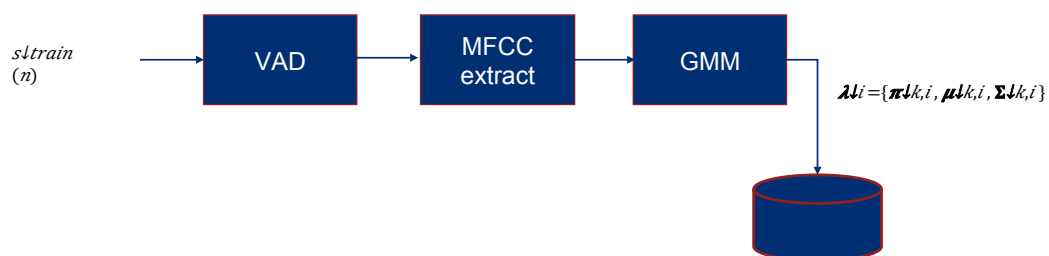
- Two main applications: verification and identification
- Voice identification
  - many models for different speakers are stored
  - new incoming speech from an unknown speaker is compared to all existing models to determine the identity of the speaker
  - one-to-many match [1:N]
- Voice verification
  - a model of the speaker is first stored
  - new incoming speech from a known speaker is compared to their model
  - voiceprint
  - one-to-one match [1:1]
- *text-dependent* or *text-independent*
  - somewhat different approaches in terms of machine learning

2018 Pindrop® 59

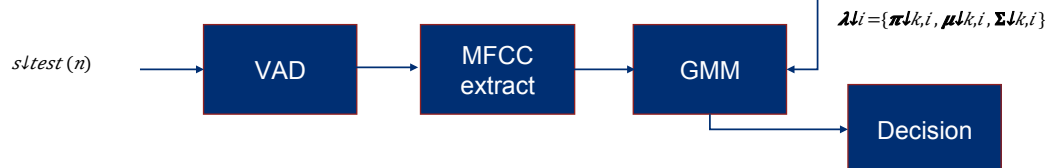
## Overview of a basic voice-id system



### Train speaker models



### Test speaker models



2018 Pindrop® 60

## Gaussian mixtures



- A mixture of Gaussians (GMM) often used to model speaker data (features)

$$p(\mathbf{f}_\ell) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{f}_\ell \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$\mathcal{N}(\mathbf{f}_\ell \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{f}_\ell - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{f}_\ell - \boldsymbol{\mu}_k)\right)$$

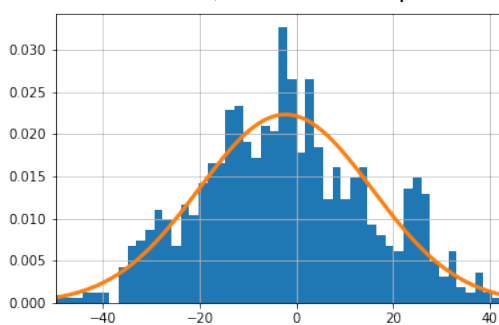
- $0 \leq \pi_k \leq 1$  - weight coefficients with  $\sum_{k=1}^K \pi_k = 1$
- $\boldsymbol{\mu}_k$  and  $\boldsymbol{\Sigma}_k$  are means and variances (covariance matrix), respectively
- for voice identification we would use  $K=16$

2018 Pindrop® 61

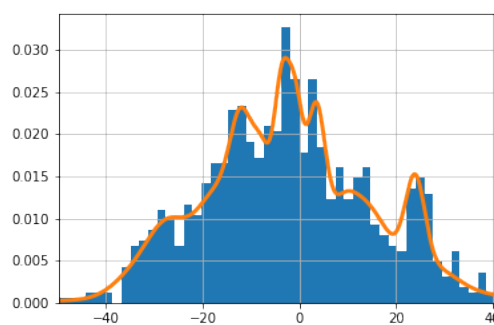
## Gaussian mixtures: example



GMM with K=1, i.e. a Gaussian pdf



GMM with K=16



2018 Pindrop® 62

## Speaker Models



- GMMs can model arbitrarily shaped pdfs which is often the case of MFCCs in speech
- the key in model training is to estimate parameters  $\pi_{k,i}, \mu_{k,i}, \Sigma_{k,i}$
- $\lambda_i = \{\pi_{k,i}, \mu_{k,i}, \Sigma_{k,i}\}$  define the model for speaker  $i=1, \dots, M$ 
  - can be estimated efficiently using the Expectation-Maximization (EM) algorithm
  - iterative likelihood maximization with respect to the parameters

2018 Pindrop® 63

## Speaker identification



- Assume a collection of  $M$  speaker models  $\mathcal{M} = \{1, 2, \dots, M\}$  represented by their GMMs  $\lambda_1, \lambda_2, \dots, \lambda_M$
- For a new utterance, extract feature vectors  $\mathbf{F}$
- then select the model with maximum likelihood

$$\hat{\mathcal{M}} = \arg \max_{1 \leq i \leq M} \sum_{\ell=1}^{\mathcal{L}} \log p(\mathbf{f}_\ell | \lambda_i)$$

- with  $\log p(\mathbf{f}_\ell | \lambda_i) = \sum_{k=1}^K \pi_{k,i} \mathcal{N}(\mathbf{f}_\ell | \boldsymbol{\mu}_{k,i}, \boldsymbol{\Sigma}_{k,i})$

2018 Pindrop® 64



## Speaker verification



- For a new utterance, extract feature vectors  $F$
- Calculate the log-likelihood ratio for voice belonging to the claimed speaker vs any other speaker
- Universal background model (UBM) to model "all speakers in the world"
- If text dependent you can apply speech recognition-like methods in addition

2018 Pindrop® 65

## Summary



- Speaker recognition will be important building block in many systems
  - voice control
- Basic GMM and MFCC based speaker-id system
  - current systems in use are more advanced but the basic principles are similar
- Voice (voiceprint) is not as robust personal identifier as fingerprint
  - what if someone steals your voice?
- Feature tailoring is an interesting area of research and still remains an important building block for machine learning systems
  - competition from Deep Learning Neural Networks

2018 Pindrop® 66