PinDr0p:
Telephony Provenance

Motivation

Robust Identification is important

New IP infrastructure allows mass phone spam

Phone scams spoofing Caller ID is more common

Current Methods

Caller ID

- Requires an honest caller/infrastructure
- Easy to spoof, hard to detect beforehand

PinDr0p's Solution

Develop a detection scheme based on channel and codec artifacts

- Eliminating channel artifacts is a hard problem
  - Nyquist/Shannon Theorem
  - Different channels have different problems

Channel Artifacts

Nyquist Shannon Theorem puts upper bound of ½ the sample rate on transmittable frequencies

- Low pass filter artifacts or aliasing artifacts

Real time linear filters never perfect

- Nonuniform attenuation/amplification of frequencies unavoidable

Channel Artifacts

Quantization error leads to bit flips

- Easy but expensive to reduce probability
- Can lead to lost packets or incorrect output
- Megabytes of data, both situations inevitable

Packet Loss

- Can be obscured or zeroed output

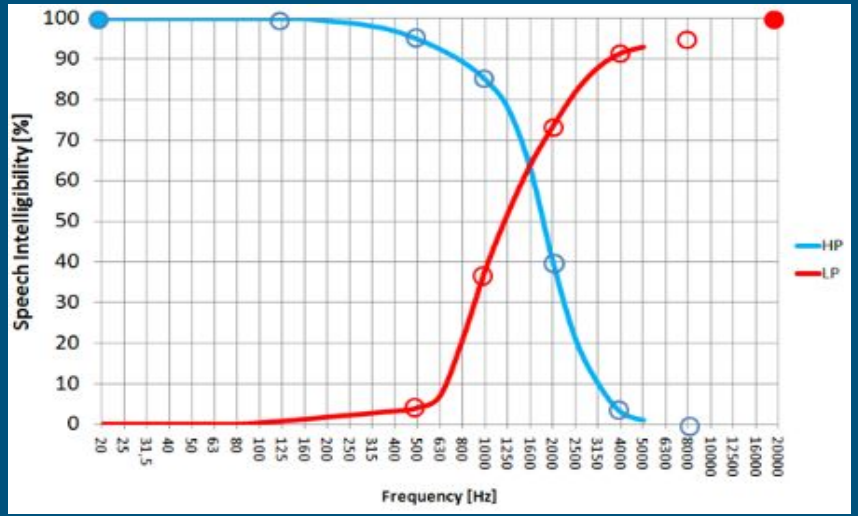
Voice Communication

Human Voice ranges from ~300 Hz to ~8 kHz

85% Intelligibility

only requires the band

~500 Hz to 4 kHz



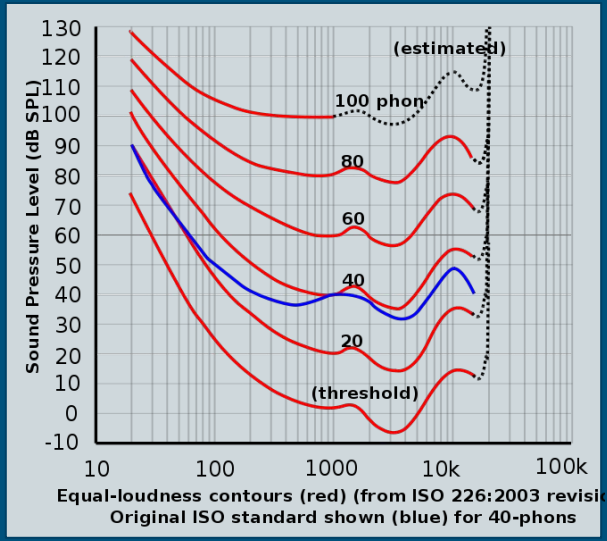
http://www.dpamicrophones.com/en/Mic-University/Tech-Guide/Facts-about-speech-intelligibility.aspx

Voice Communication

Human hearing dynamic range is amplitude and frequency dependent, given by the Fletcher Munson Curves

Bark is a measure of perceptual frequency

Sone is a measure of perceptual loudness



https://en.wikibooks.org/wiki/Engineering_Acoustics/The_Human_Ear_and_Sound_Perception

Voice Communication

Pitch Synchronous Residual

- Computed difference between estimated “pitch” and voice signal
- This allows us to reduce the information needed for the actual signal

Pitch Synchronous Analysis

- Fourier analysis of pitch periods
- Sliding window of periods of constant pitch

Call Quality Metrics

Spectral Clarity: How “crisp” audio is, how distinguishable different sounds are

Mean Opinion Score

- Subjective measure of quality
- Expert listeners rate quality on scale of 1 to 5, 1 being the worst

Perceptual Evaluation of Speech Quality

- Objective measure of quality
- Software simulation of ear’s perceptual ability

PESQ

Compares the original signal with the degraded signal

Outputs a prediction of the perceived quality by subjects

Computes a series of delays between the time intervals of the inputs to compare corresponding time intervals

Transforms the inputs to approximate psychophysical representations like the bark and sone

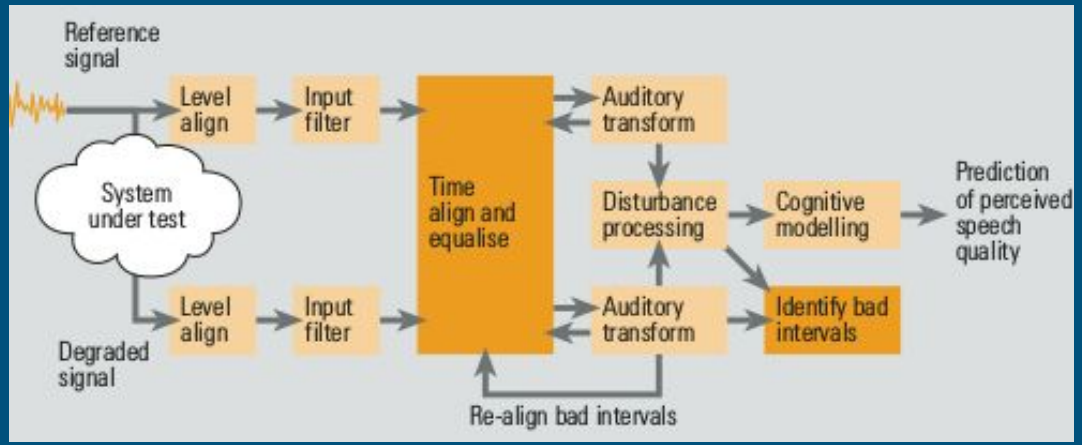
PESQ

Computes difference between psychophysical representations

Cognitive modelling is performed by a weighted sum of distortions determine how disruptive the differences are

Uses a regression process of MOS scores and distortions to determine weights for estimating subjective quality

Shown to be within 0.5 of the actual MOS



Linear Block Code (LBC)

Block

$$\vec{v} = [v_1, ..., v_n] \in GF(2)^n \rightarrow v_i \in \{0, 1\}$$
$$\vec{u} + \vec{v} \cong [u_1 + v_1, ..., u_n + v_n](mod\ \ 2)$$

A linear space of blocks

$$a, b \in GF(2), \vec{u}, \vec{v} \in LBS, a\vec{u} + b\vec{v} \in LBS$$

Block Code: Each block maps to a symbol

Linear Predictive Coding

Represents the spectral envelope in compressed form

Assumes speech is produced with a buzzer and additional hissing and popping sounds

LPC estimates formants and computes the speech signal without them

- Formants are frequency bands in the produced sound from the buzzer

PSTN Network

PSTN uses G.711 signals

- Signal Transmitted: 8 kHz, 8 bits/sample
- Low quantization noise
- Most noise is highly correlated to the signal
- Spectral clarity is high

Cellular Network

Cellular Networks encode voice with GSM-FR

- Uses a lossy compression algorithm
- Signal Transmitted: 13 kbps
- High quantization noise
- Low spectral clarity

VoIP iLBC

Lossy Compression

13.3 kb/s at frames of 30 ms, 15.2 kb/s at 20 ms

Uses a speech model for compression

Uses a block independent linear predictive coding

Uses Packet Loss Concealment based on the speech model

VoIP iLBC

For a block of speech, create a set of LBC filters

The speech signal is inverse filtered, leaving the sound energy residual (SER)

- Sound energy in the form of formants from the voice box
- Filtered by geometry of mouth to create speech
- Inverse filter removes formants to make the sound energy residual

iLBC Packet Loss Concealment

When a packet is dropped, iLBC attempts to recover

Uses the previous packets SER plus some random noise

Uses the same LBC filters as well

Opus - A more recent codec

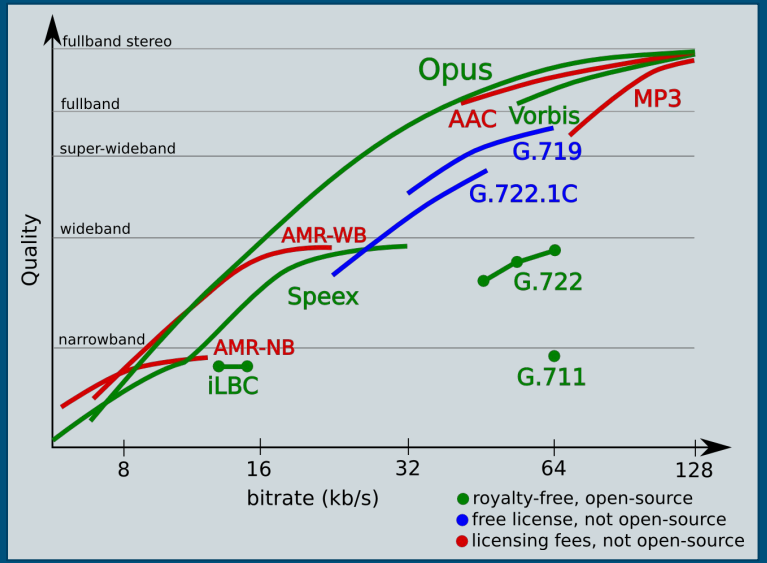
Open codec accepted as standard in 2012

Built off of Silk and CEPH codecs

Can be encapsulated in OGG

Implemented in WebRTC

Scales from 6 kbps to 510 kbps



Codec Feature Comparison

Feature\Codec	G.711	GSM-FR	iLBC	Opus
Bitrate (kbps)	64	13	13.3, 15.2	8, 12, 16, 24, 32
Quantization Noise	Low	High	Unknown	Unknown
Spectral Clarity	High	Low	Unknown	Unknown
Sample Rate (kHz)	8	8	8	8-44
Frame Size (ms)	Sample Rate	22.5	20, 30	2.5-60
MOS	4.1	3.5	3.9	4.5

VoIP Codec Usage

Skype - Silk, Opus, G.729, G.722, G.711, H.264

Google Hangouts (WebRTC) - Opus, G.711

Snapchat (WebRTC) - Opus, G.711

Vonage - G.711, G.726, G.729

Whatsapp - Opus (probably)

Google Talk (dead) - iLBC, Speex

MSN Messenger (dead as of Oct 31) - iLBC, Speex, Siren

C4.5 Decision Tree

Binary tree, splits sets based on attribute values with most information

Uses the Expected Kullback-Leibler Divergence to determine the attribute with maximum information

Expected Kullback-Leibler Divergence is the cost of using a different code

Splits on the attribute with the most information

k-fold Validation

Divide training set into k sets

Train k classifiers, with a unique set of k-1 sets

Validate each classifier on it’s excluded set

Combine scores to estimate success of the fully trained classifier

- Scores can be averaged, or use the median to estimate fully trained success

PinDr0p Training Data

Multilabel Classifier is trained against a repository of speech samples

- 20 samples of 15 s of speech from the Open Speech Repository
- Used for standardized PSTN tests
- Samples are transformed with real world degradations
- Only 1 to 3 different traversed networks are assumed
- Samples traversing multiple networks are converted to G.711 between networks

PinDr0p Testing

Training data used with a C4.5 decision tree

10 fold cross validation is used to evaluate success

- 3 reduction techniques used to estimate accuracy, all estimate >90%
- None of these estimates seem to be commonly used, are they valid?

PinDr0p Real World Evaluation

Testers each make 10 calls 20 s long; each in one of 16 places around the world

- 10 call sets built from the data
- Neural Network trained on some of the call sets
- Accuracy of classification of location is high even for training on one set
- No guarantees of robustness as more locations are added
  - Location classification is way too sparse

Pindrop Now

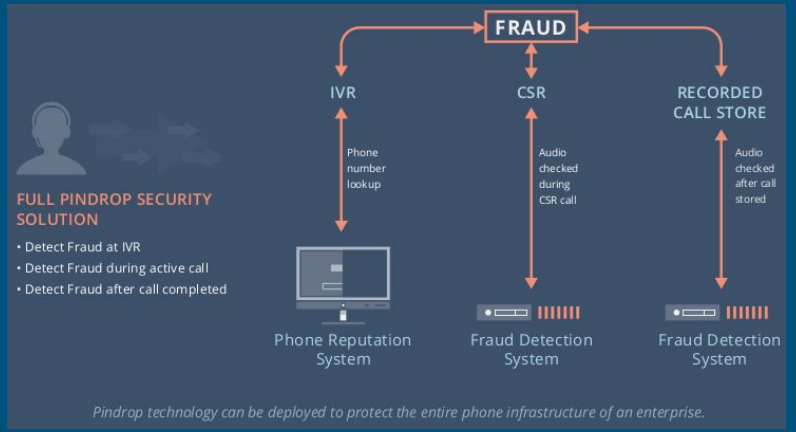
Significant improvements in provenance

Determines geography and calling device type

80% of fraud calls detected on first call

147 features analyzed

$4 billion in committed capital in 2015



Pindrop Features

Packet loss

Dropped frames -

Quantization - Number of intensity levels

Clarity - Distinguishability of noises

Noise correlation -

Signal to Noise Ratio

Sources

Pitch Synchronous Analysis of Voiced Sounds, M. V. Matthews, Joan E. Miller, E. E. David Jr.
Audio codec identification from coded and transcoded audios, Samet Hicsonmez, Husrev T. Sencar, Ismail Avcibas
On Estimating Accuracy with Repeated Cross-Validation, Gitte Vanwinckelen, Hendrik Blockeel
http://www.dpamicrophones.com/en/Mic-University/Tech-Guide/Facts-about-speech-intelligibility.aspx
https://en.wikibooks.org/wiki/Engineering_Acoustics/The_Human_Ear_and_Sound_Perception#cite_note-Hirsh-4
http://hyperphysics.phy-astr.gsu.edu/hbase/sound/phon.html
http://www.itu.int/rec/T-REC-P.862/en
http://tools.ietf.org/html/rfc3951
https://en.wikipedia.org/wiki/PESQ
https://en.wikipedia.org/wiki/C4.5_algorithm
https://en.wikipedia.org/wiki/Information_gain_in_decision_trees
https://developers.google.com/talk/open_communications
https://en.wikipedia.org/wiki/Google_Talk
https://en.wikipedia.org/wiki/Opus_(audio_format)
https://tools.ietf.org/html/rfc6716
https://bloggeek.me/whatsapp-voip-implementation/
https://webrtchacks.com/hangout-analysis-philipp-hancke/
http://www.opticom.de/download/SpecSheet_PESQ_05-11-14.pdf
http://www.pindropsecurity.com/pindrop-security-closes-35-million-investment-to-extend-leadership-in-call-center-anti-fraud-and-authentication/
http://www.en.voipforo.com/codec/codecs.php
http://www.SampleSwap.org
https://en.wikipedia.org/wiki/Linear_predictive_coding