Audio processing

Mathieu Lagrange 

February 6, 2019

Course website: https://github.com/mathieulagrange/datasim

Outline

Outline

Outline

Outline

Data processing

Data: the challenge of the next century

Data processing

Data: the challenge of the next century
"We are drowning in information but starved for knowledge",

J. Naishbitt

Types of data

- dictionary based: text, symbolic music, ...
- spatial sampling based: image, ...
- time sampling based: audio, video

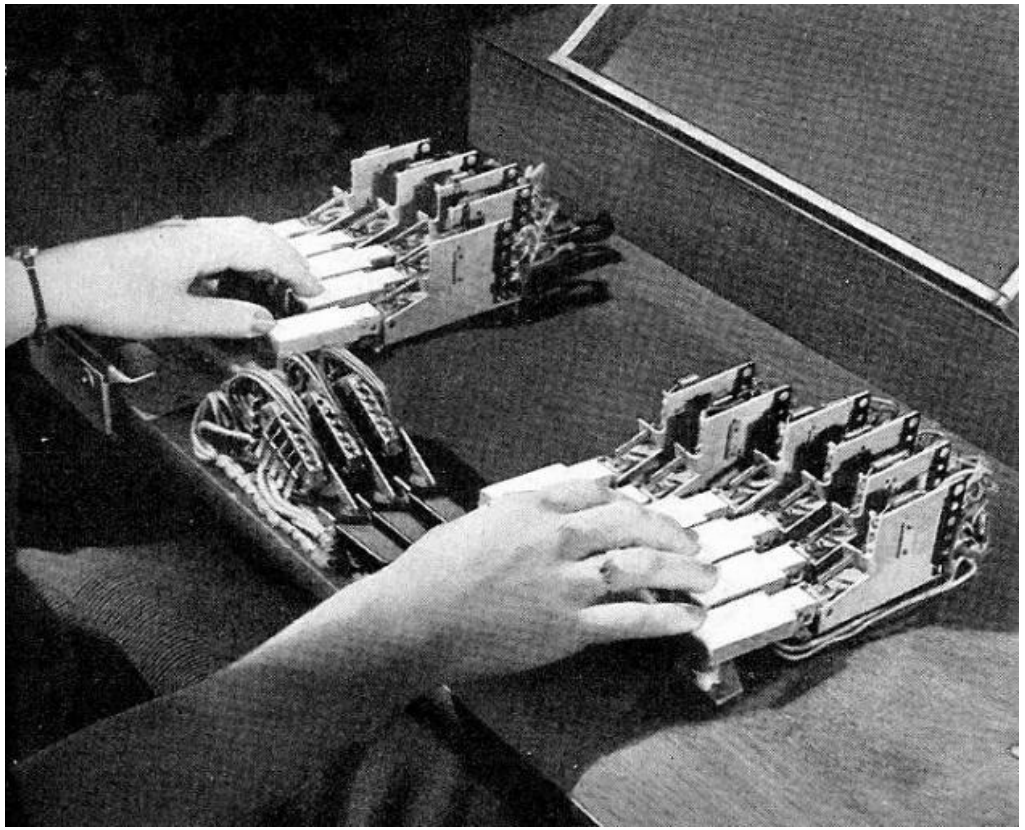
Sampled Data

- recording: cheap
- storing: cloud
- transmitting: fiber
- **accessing**

Speech processing

- the early application topic
- dream of conversational AI (1950)
- need for speech synthesis: text to speech (TTS)
- need for speech recognition: automatic speech recognition (ASR)

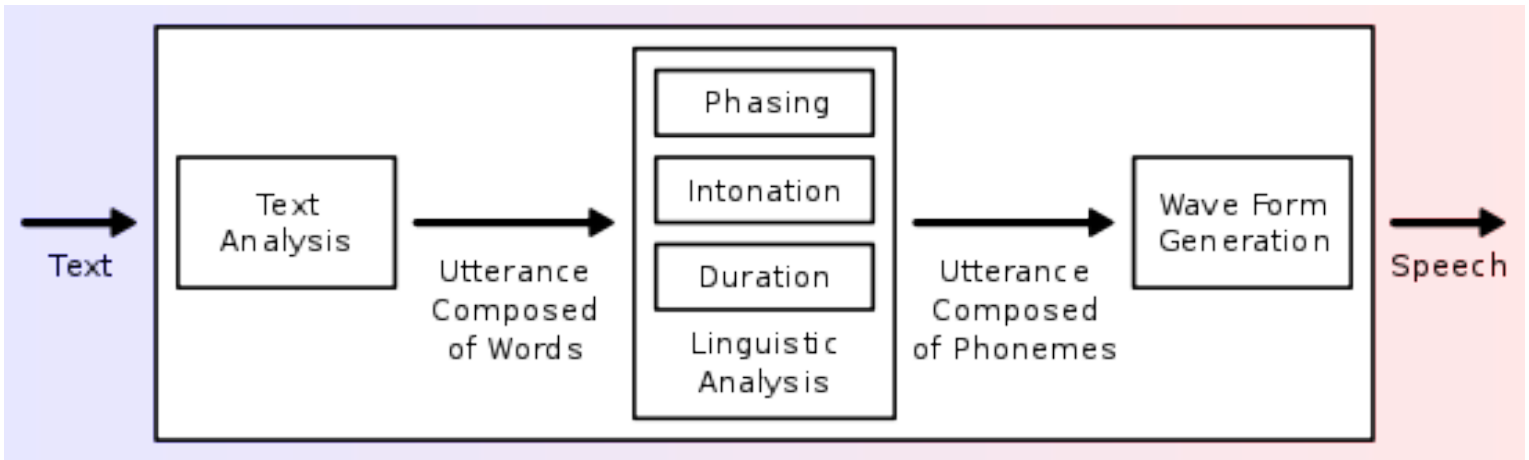
Speech synthesis



The VODER

https://www.youtube.com/watch?v=0rAyrmm7vv0

Speech synthesis



Speech synthesis

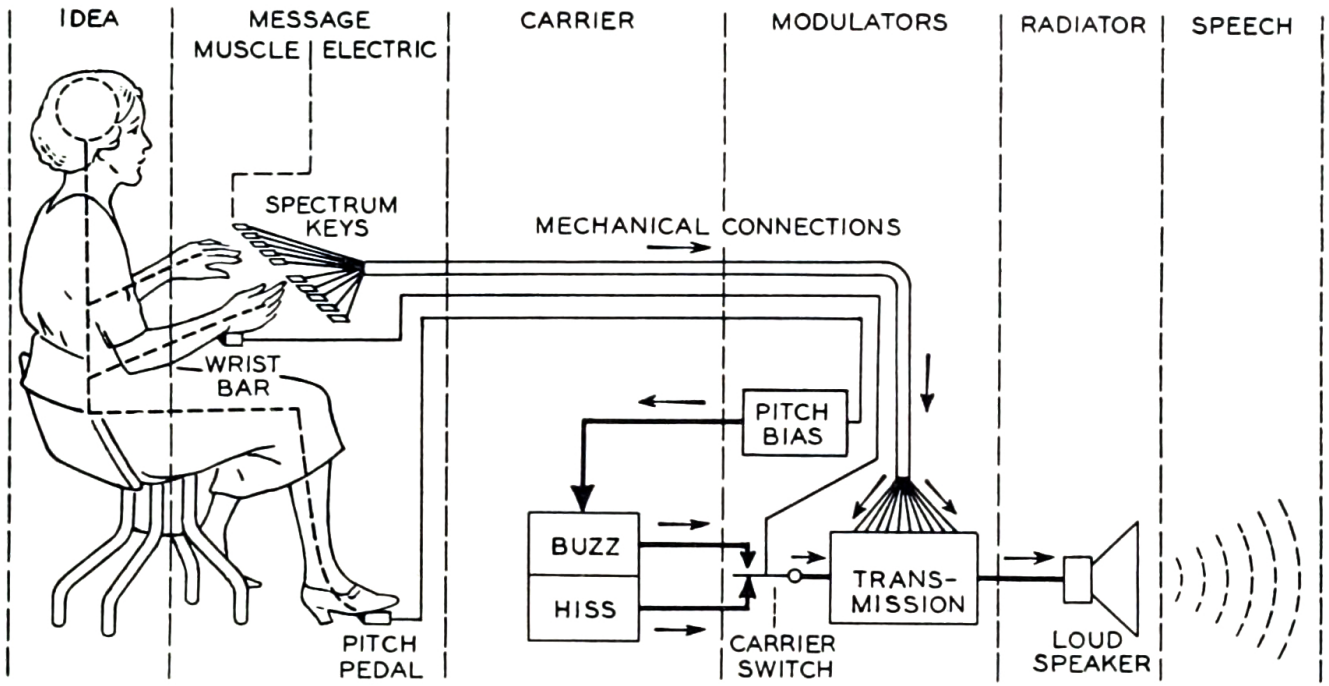
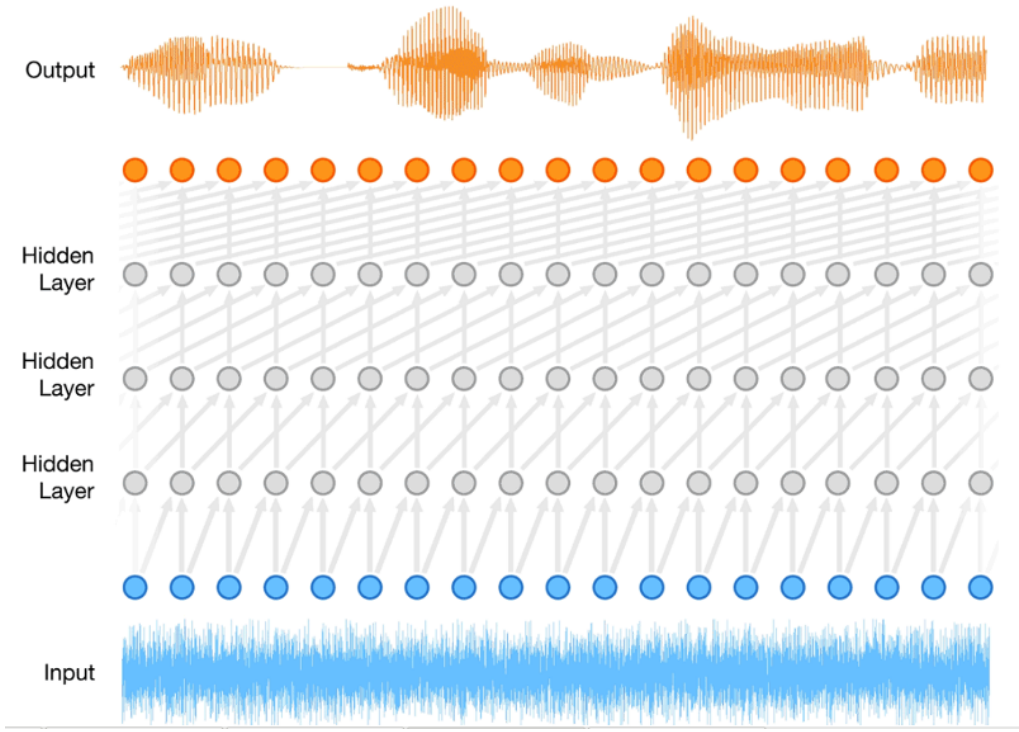


Fig. 8—Schematic circuit of the voder.

Wavenets



Concatenative Parametric Wavenet Unconditioned Wavenet

Oord, Aaron van den, et al. Wavenet: A generative model for raw audio 2016

Speech recognition

- Task: from raw audio predict phonemes, words, phrases
- Metric: word error rate (human 5 %)

Speech recognition history

- 50-60s: Bell Laboratories designed the "Audrey" system which could recognize a single voice speaking digits aloud.
- 70s: Carnegie Mellon's "Harpy" speech system came from this program and was capable of understanding over 1,000 words which is about the same as a three year old's vocabulary.
- 80s: vocabulary go from a few hundred words to several thousand words. Introduction of the Hidden Markov Models for modeling phoneme transitions.

Speech recognition history

- 90s: BellSouth introduced the voice portal (VAL) which was a dial-in interactive voice recognition system. This system gave birth to the myriad of phone tree systems that are still in existence today.
- 00s: Move to the cloud: Google's English Voice Search System included 230 billion words from user searches.
- 10s: Apple Siri. speech accuracy title is subject of competition from majors companies (ibm, microsoft, google): 4.9 % WER

Environmental audio processing

Environmental audio processing is about

- detection of failures: production chain, structure monitoring
- security: aggression, detection, monitoring
- environmental monitoring: ecoacoustics, urban acoustic quality monitoring

Potentials

Using airborne acoustics has some advantages

- contactless
- non intrusive (despite privacy issues)
- cheap sensors
- relatively low bandwidth

Security monitoring

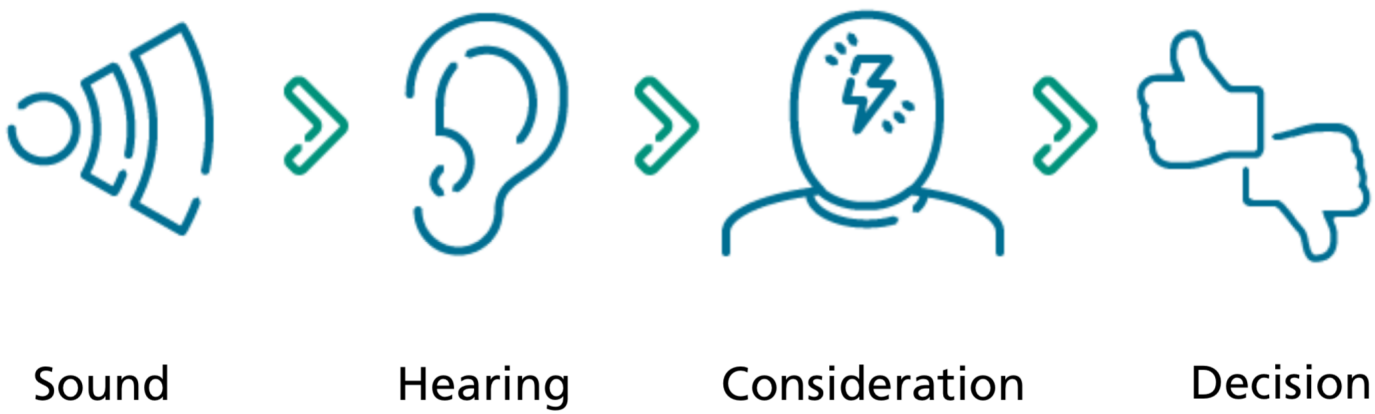
- aggression: verbal, gunshots
- health care, elderly care
- intrusion: breaking glass

https://www.soundintel.com/

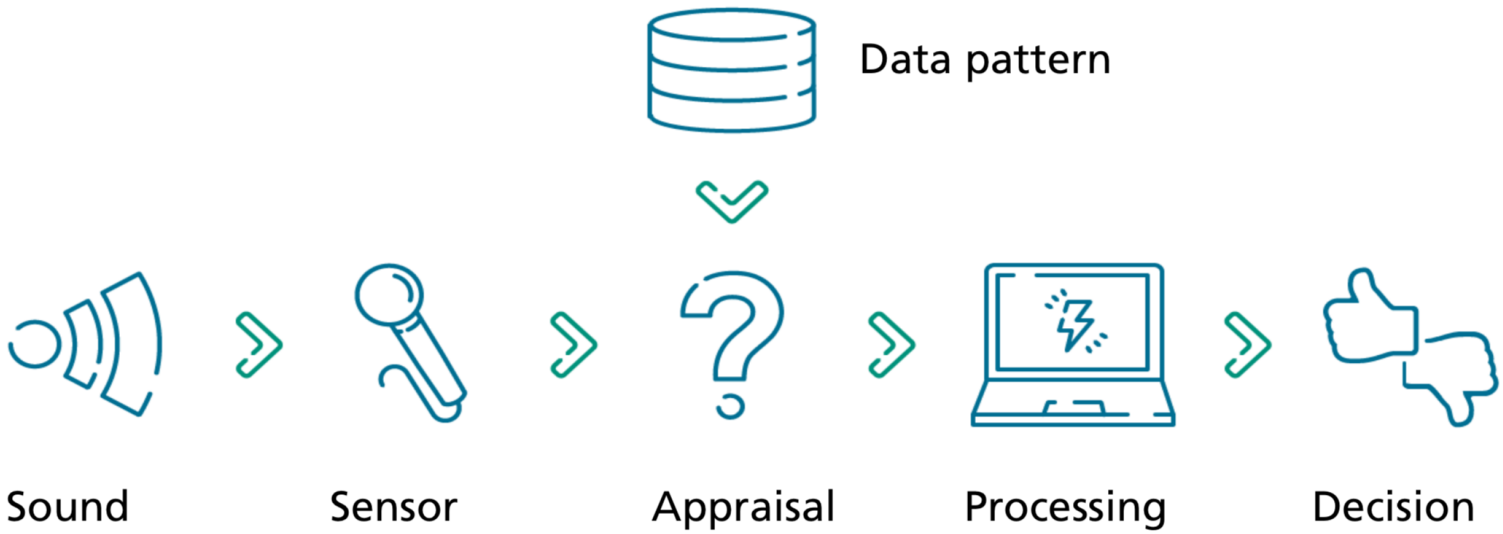
Acoustic condition monitoring (acm)

- many moving parts make noises (motors, fans, ...)
- many companies rely on human ears
- some are accompanied by automated procedures
- Applications: Quality control, predictive maintenance, end-of-line testing, process / workflow monitoring

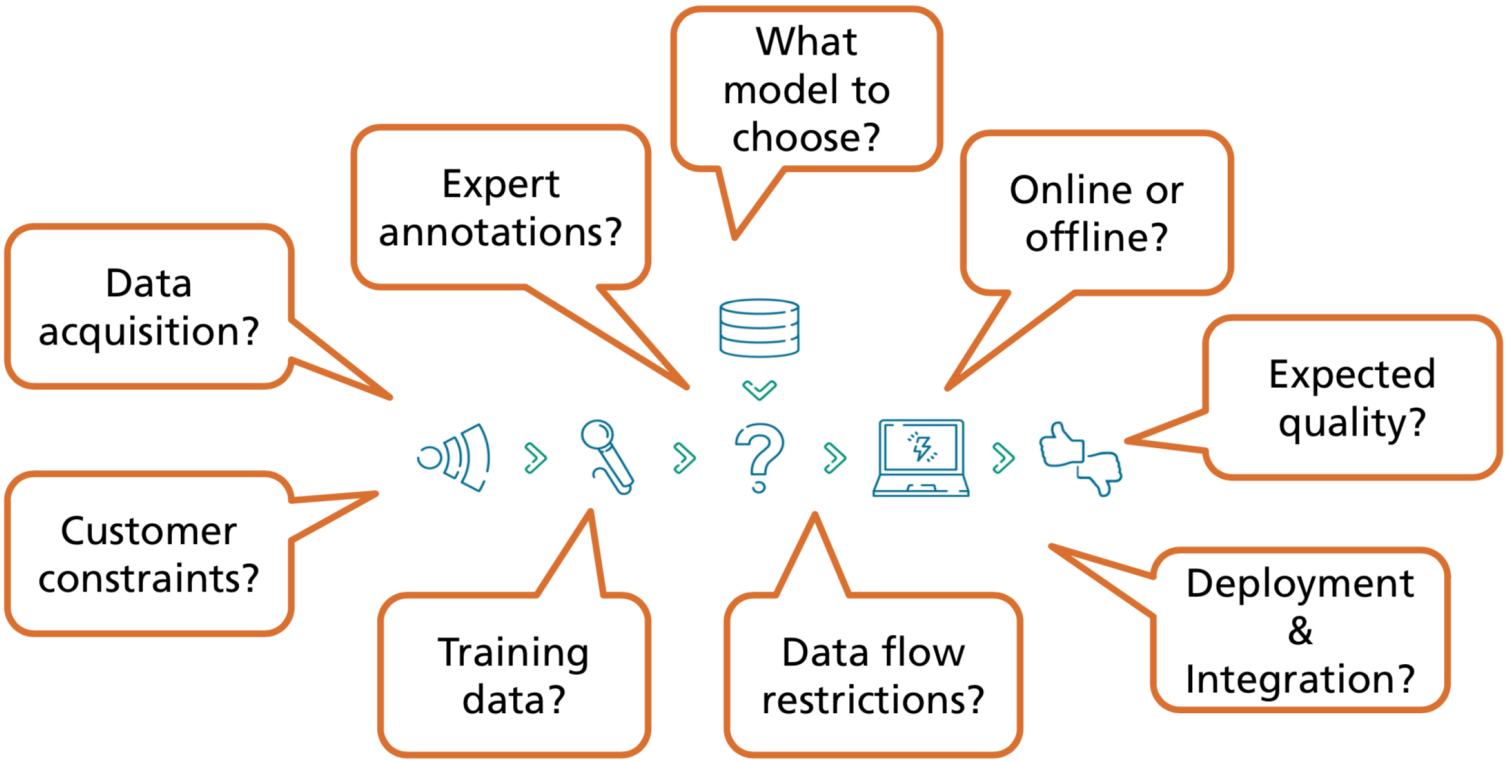
Human based Acm



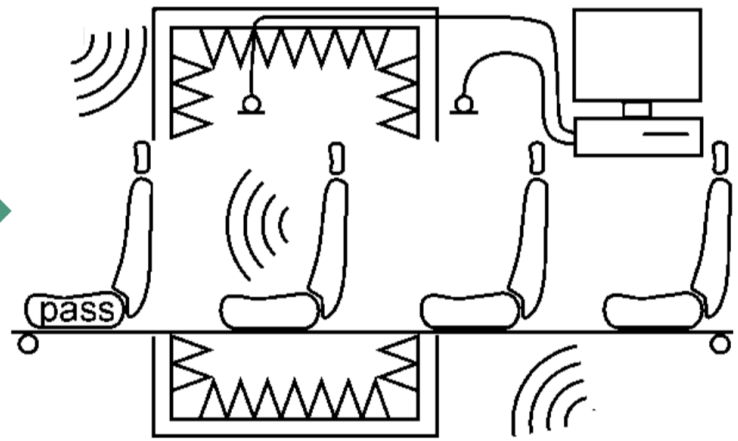
Computer based Acm



Computer based Acm



Acm use case 1

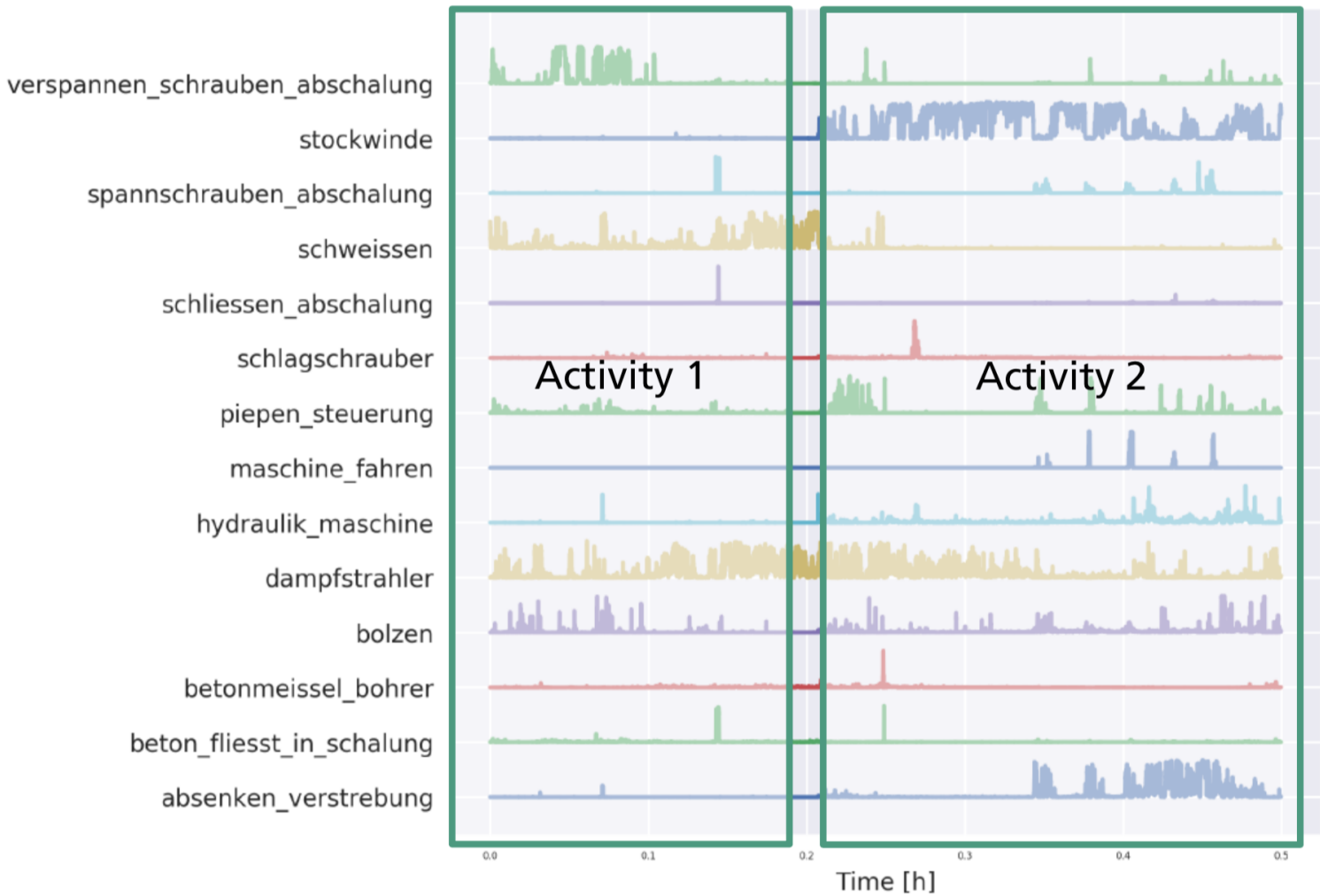


provided by Hanna Lukashevich @ Fraunhofer

Acm use case 2

provided by Hanna Lukashevich @ Fraunhofer

Acm use case 2



Ecoacoustics

Ecoacoustics

Audio is used

- to study and monitor animal diversity, abundance, behavior, dynamics and distribution,
- and their relationship with ecosystems and the environment.

Ecoacoustics

Ecoacoustics

- investigates natural and anthropogenic sounds and their relationship with the environment
- recognizes that sounds can be both the subject and tools of ecological research.
- As the subject, sounds are investigated in order to understand their evolution, functions and properties under environmental pressures

https://sites.google.com/site/ecoacousticssociety/about

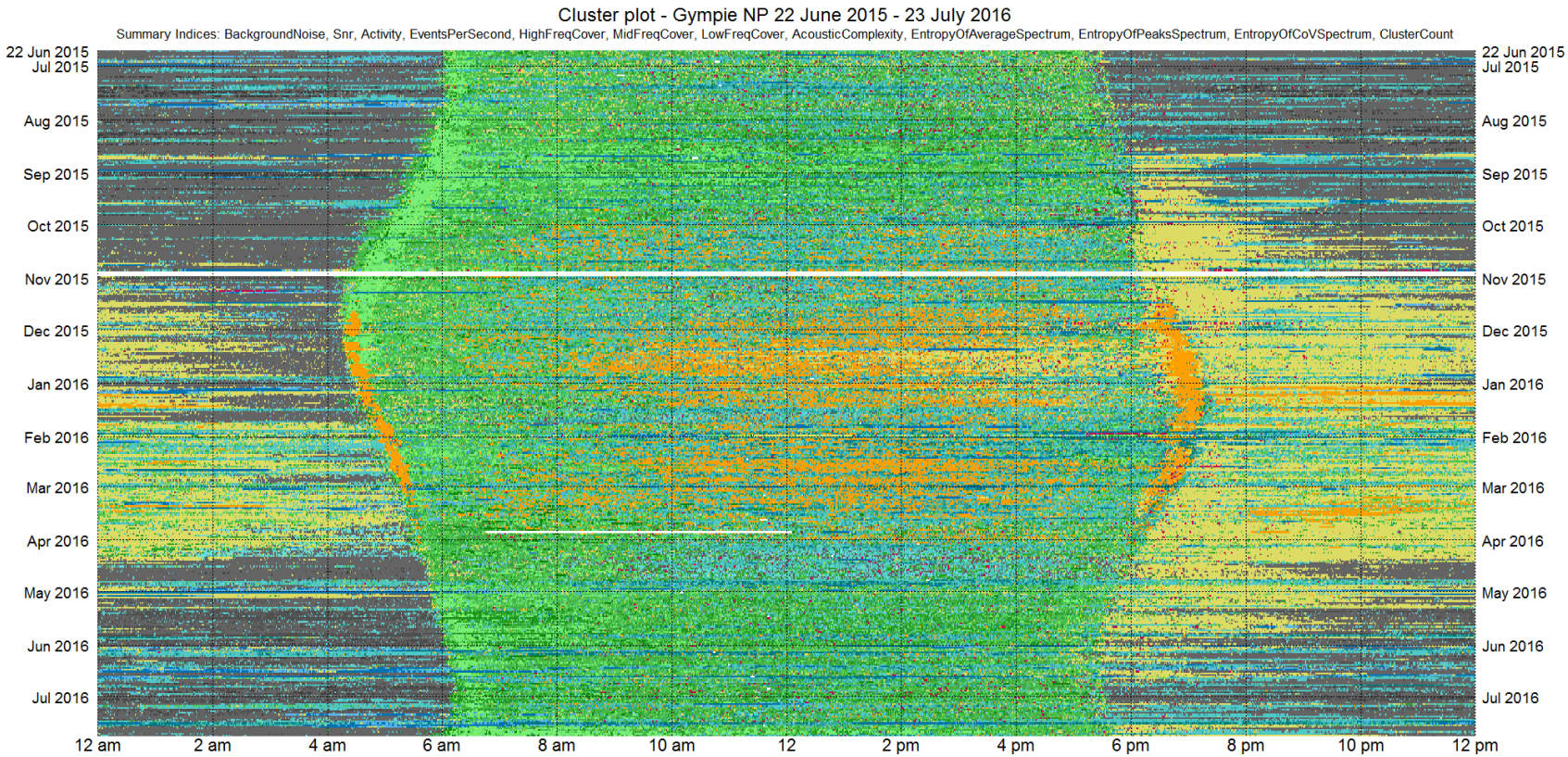
Diel plot

The next graphic shows

- 1.1 million vectors of 12 summary indices
- that represents 13 months of continuous audio, recorded in Gympie National Park.
- colour coding was simplified to seven basic sound categories or acoustic states:
  1. silence clusters are represented in grey;
  2. bird cluster in green;
  3. rain clusters in dark blue;
  4. wind clusters in pale blue;
  5. insect cluster in yellow;
  6. cicada chorus clusters in orange; and
  7. plane sounds in red.

Phillips & al Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation 2018

Diel plot

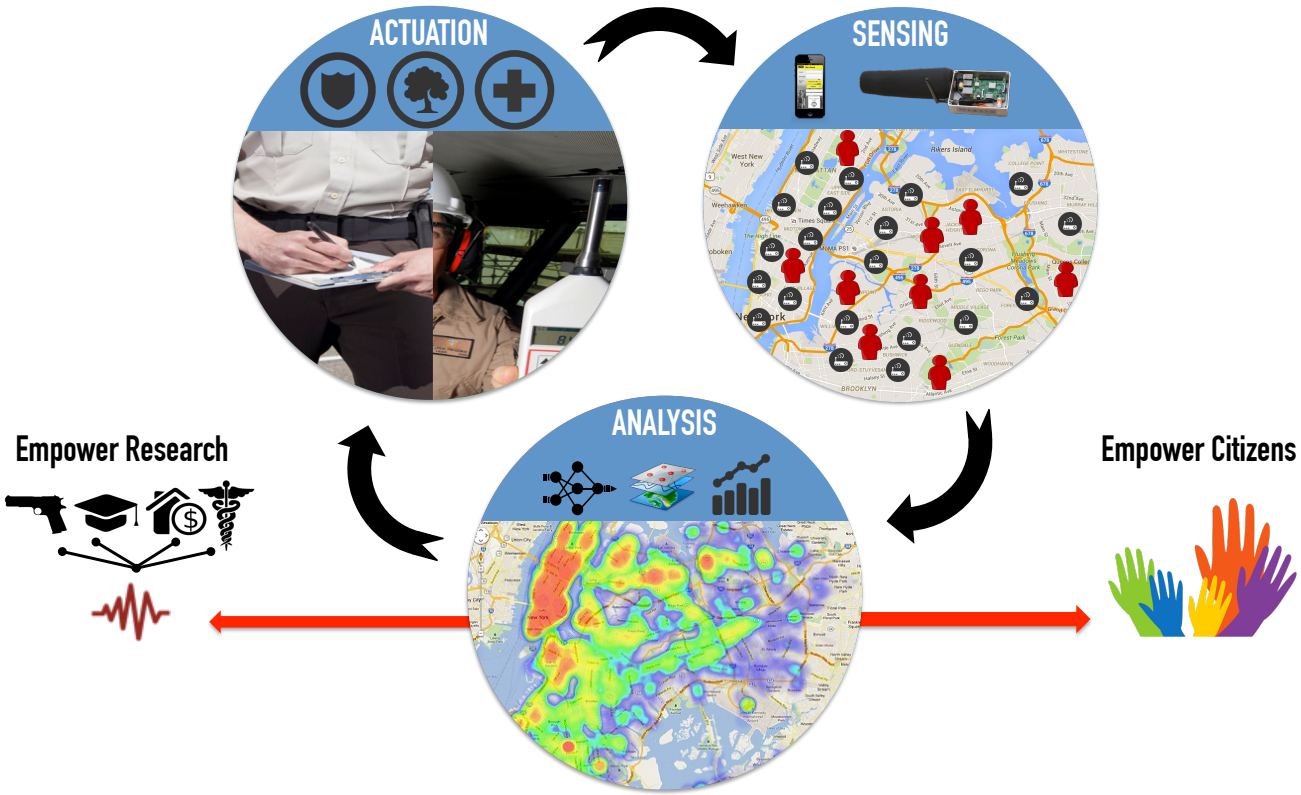

Cluster plot - Gympie NP 22 June 2015 - 23 July 2016
Summary Indices: BackgroundNoise, Snr, Activity, EventsPerSecond, HighFreqCover, MidFreqCover, LowFreqCover, AcousticComplexity, EntropyOfAverageSpectrum, EntropyOfPeaksSpectrum, EntropyOfCoVSpectrum, ClusterCount
22 Jun 2015
Jul 2015
Aug 2015
Sep 2015
Oct 2015
Nov 2015
Dec 2015
Jan 2016
Feb 2016
Mar 2016
Apr 2016
May 2016
Jun 2016
Jul 2016
12 am
2 am
4 am
6 am
8 am
10 am
12
2 pm
4 pm
6 pm
8 pm
10 pm
12 pm

Urban acoustic quality monitoring

New set of tools

- based on distributed network of both sensors and people for large-scale noise monitoring.
- using big data solutions to analyze, retrieve and visualize information from sensors and citizens,
- aims at creating a comprehensive acoustic model of the city that can be used to identify significant patterns of noise pollution.

SONYC: Sounds of New York City



Bello & al SONYC: A System for the Monitoring, Analysis and Mitigation of Urban Noise Pollution 2018

Music Information Retrieval

MIR is the interdisciplinary science of retrieving information from music. It has background in

- musicology, psychoacoustics, psychology
- signal processing, informatics, machine learning.

The analysis can be performed at 2 levels:

- song level analysis,
- collection level analysis.

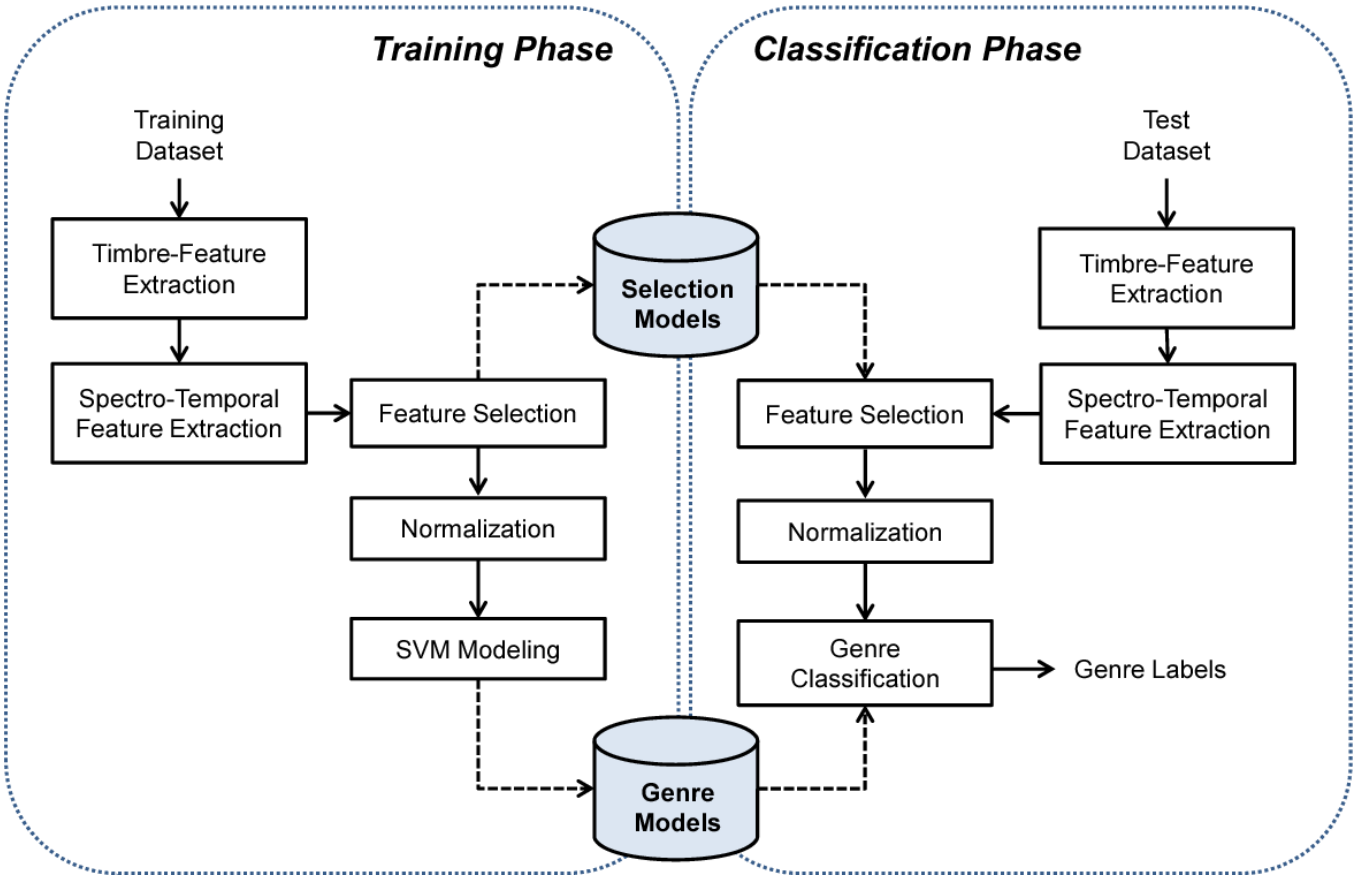
Song level analysis

- transcription
- rhythm, melody, chords
- demixing
- generation

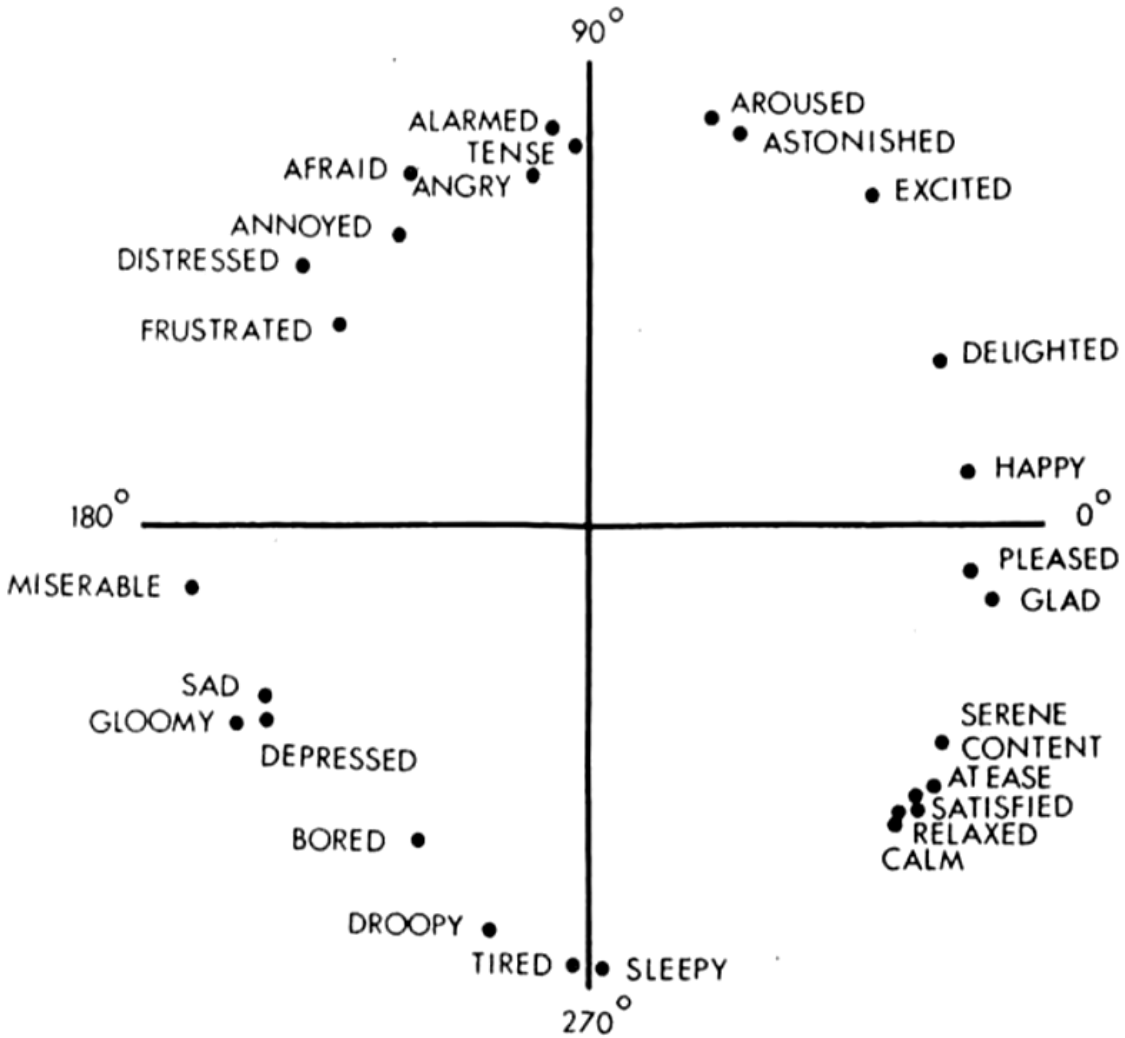
Collection level analysis

- genre classification
- mood regression
- fingerprinting
- cover song detection

Genre classification



Mood regression



Fingerprinting

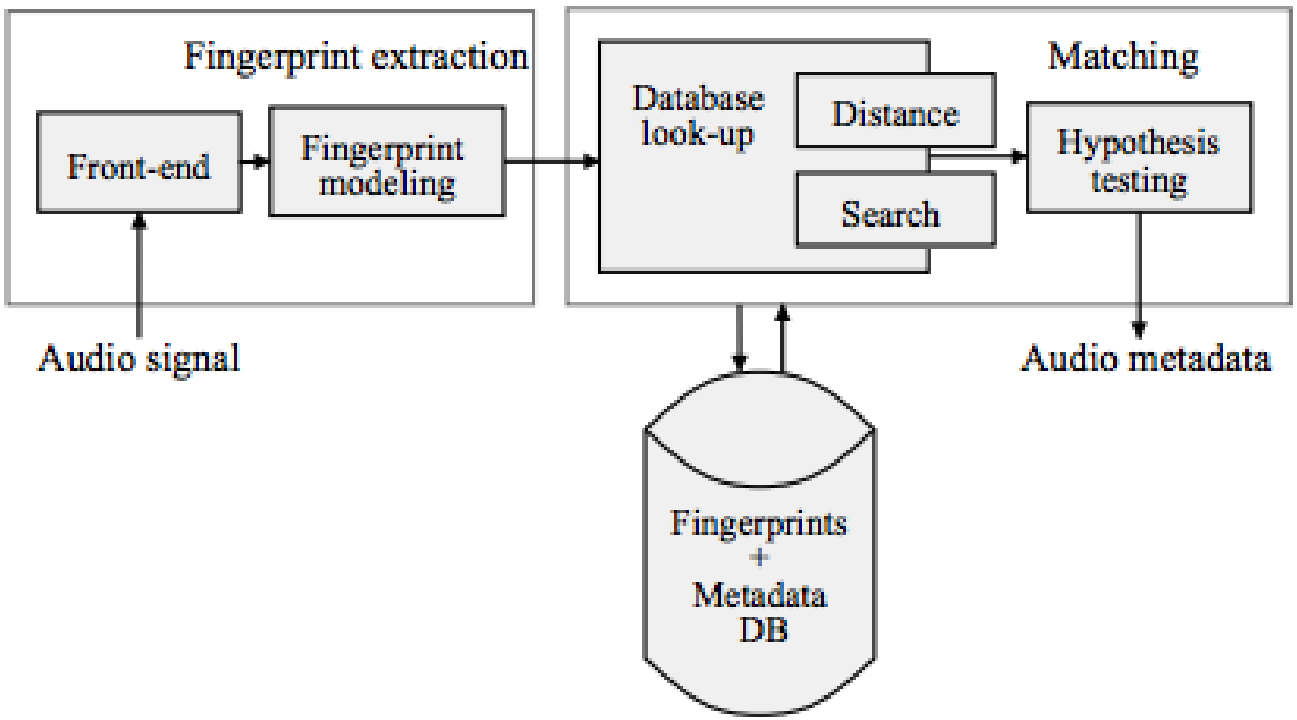
An acoustic fingerprint

- is a condensed digital summary, a fingerprint, deterministically generated from an audio signal
- that can be used to identify an audio sample or quickly locate similar items in an audio database.
-

Application scenarios:

- meta data retrieval: Shazam
- monitor the use of specific musical works and performances on radio broadcast, records, CDs, streaming media and peer-to-peer networks.
- This identification has been used in copyright compliance, licensing, and other monetization schemes.

Processing pipeline

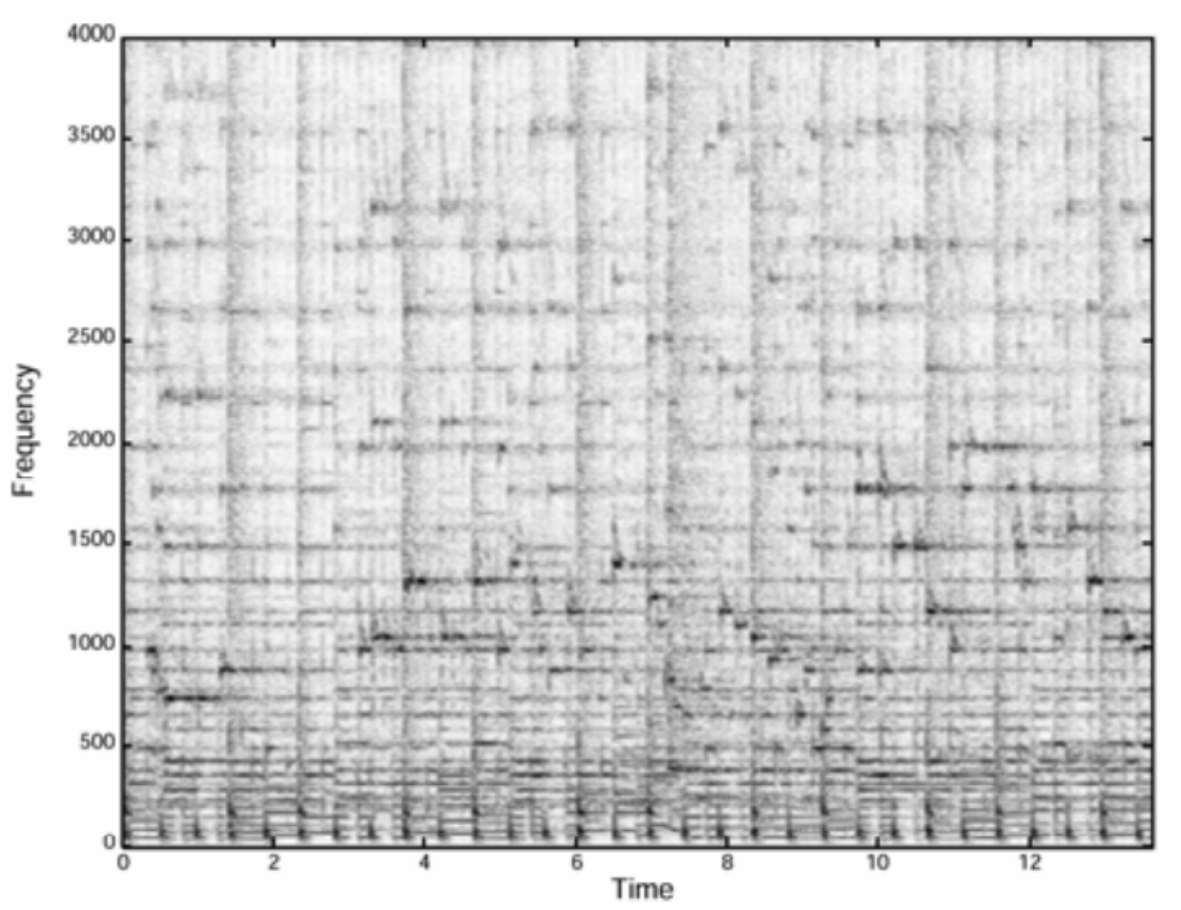


Cano & al A review of algorithms for audio fingerprinting 2002

Challenges

- Immunity to channel added noise
- Recognize fragments from anywhere in the track (the shorter, the better)
- Large corpus of reference items

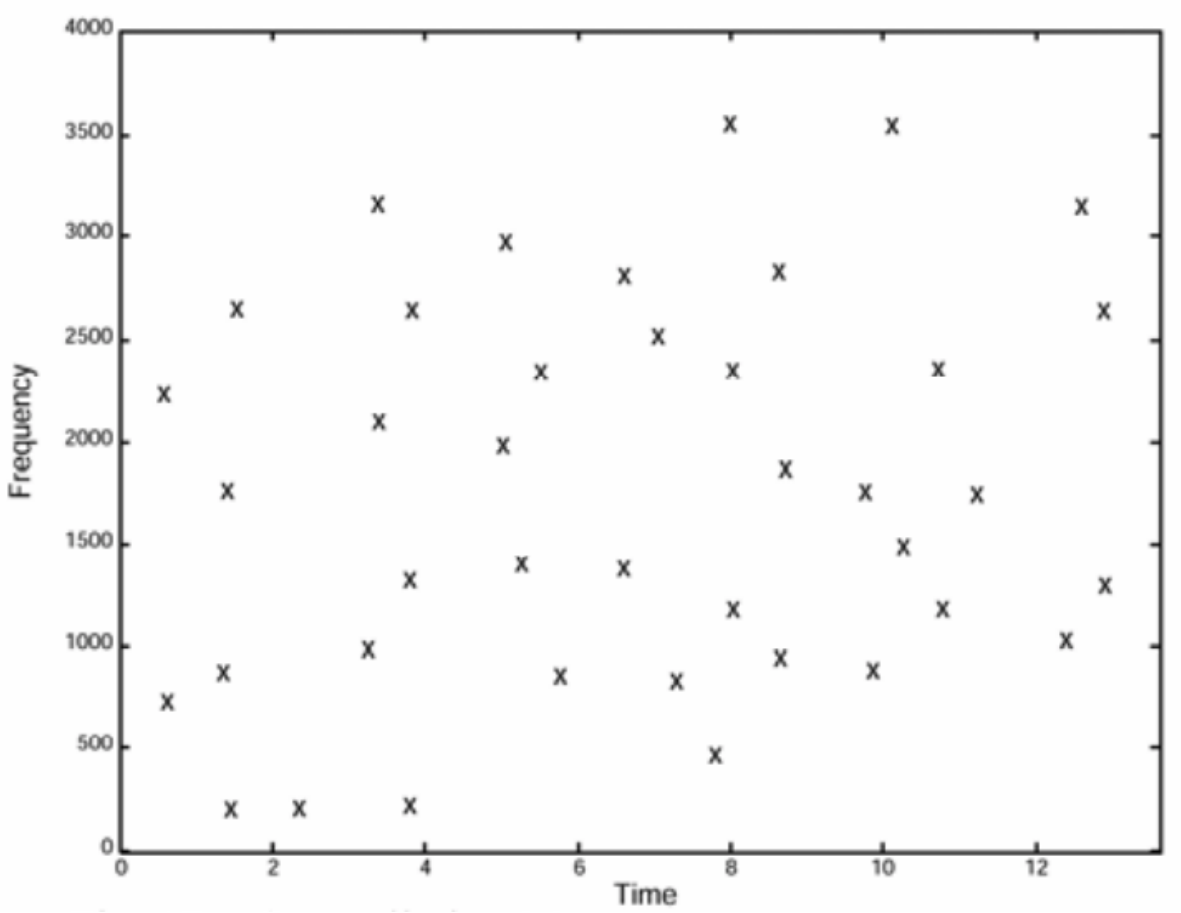
The Shazam algorithm



Spectrogram

Wang & al An industrial-strength audio search algorithm 2003

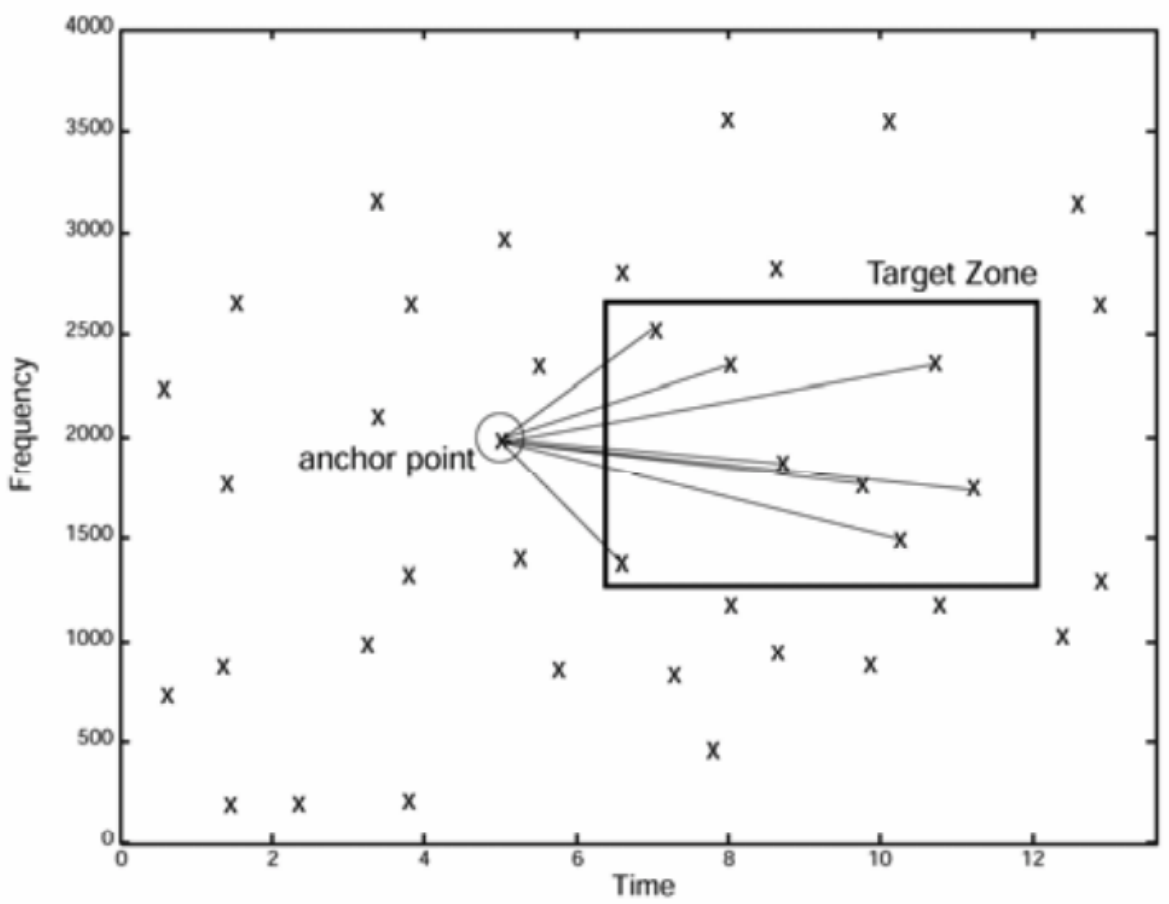
The Shazam algorithm



Peak selection

Wang & al An industrial-strength audio search algorithm 2003

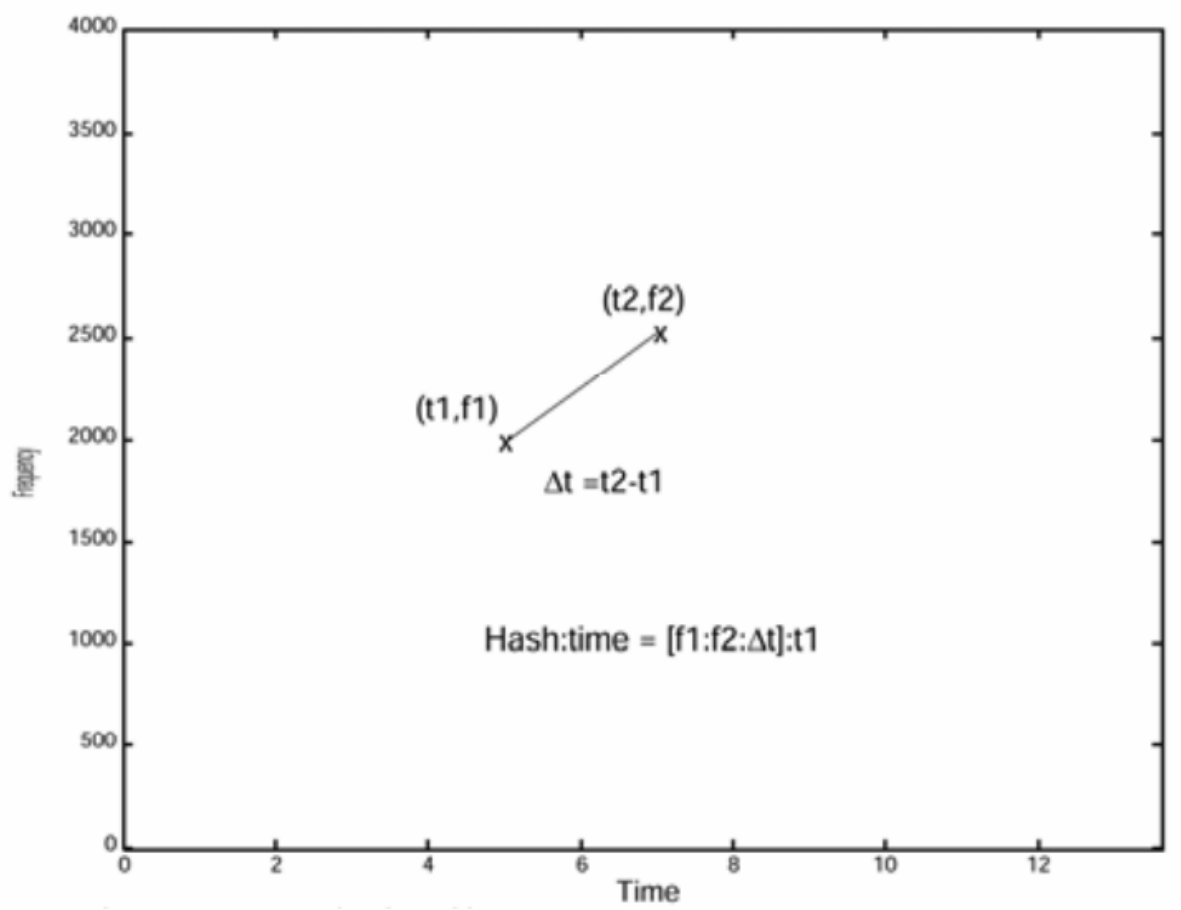
The Shazam algorithm



Pairing lookout

Wang & al An industrial-strength audio search algorithm 2003

The Shazam algorithm



Hash composition

Wang & al An industrial-strength audio search algorithm 2003

euscript,amsmath,amssymb,amsfonts,amsthm,epsfig,subfigure,color
[latin1]inputenc [T1]fontenc pgf tikz smartdiagram
letltxmacro beamerthemedefault, multimedia, wasysym,
amssymb, kpfonts
sech *sech* csch *csch* arcsec *arcsec* arcCot *arcCot* arcCsc *arcCsc* arcCosh *arcC*
[]beamerouterthemesmoothbars
[shadow=true]beamerinnerthemerounded
ping