## Sound Recognition

#### ~ CSE 352 Team 3 ~

#### $\bullet \bullet \bullet$

Jason Park Kevin Lui Evan Glover Aman Rawat Prof. Anita Wasilewska

## What is Sound?

Sound is a vibration that propagates as a typically audible mechanical wave of pressure and displacement through a medium.

Elements of sound perception:

Pitch: Frequency of the sound

<u>Duration</u>

Loudness

Timbre: How a sound changes over time

DOT: DarAmet

Sound traveling in a gas or liquid medium



Sound traveling through solid medium

Sonic Texture: Interactions between multiple sound sources Spatial Location: Where the sound comes from

## What is Sound Recognition?

A subset of pattern recognition

Depending on purpose, different AI discipline can be used

Neural networks for Speech.

Adaptive algorithms for noise cancellation and virtual sur

General pattern recognition for Music Recognition

Classification



## Common Uses

**Event Detection** 

Song Recognition

**Noise Cancellation** 

Voice/Speech Recognition

**Environmental Condition Detection** 

Mapping



Music Composition

etc.

## Audio Signal

An electrical representation of sound

Sends information along a signal flow

From source to speaker or recording device.

Frequency Range: 20 to 20,000 Hz (limits of human hearing)

Can be synthesized or originate from a transducer.

Parameters:

**Bandwidth** 

لا يوسى Difference between upper هيد بريد r frequencies , set of frequerions:

## Acoustic Fingerprint

Condensed digital summary - "fingerprint" - used to identify audio samples or items in an audio database.

Key characteristics:

Estimated tempo

Average Zero crossing rate

Average Spectrum

Spectral Flatness



Tones Bandwidth

## Automatic Content Recognition

Used to identify content element without user input,

Commonly uses acoustic fingerprinting and watermarking

Associates content and associated information in a database, and allows for the return of metabata to a client.



#### Feature Extraction

When the input data to an algorithm is too large, it can be transformed into a reduced set of features.

Reducing the amount of resources required to represent a large set of data, referred to as feature vectors.

The process to reduce variables involved is called dimensionality reduction.

Plays a major role in Digital Signal Processing (DSP).

Few of the models used for DSP are:

LPC - Linear Predictive analysis

MFCC Mel-frequency cepstral coefficients

#### Feature Extraction

Speech is highly variable

- different speakers
- Speaking rates
- Content
- Acoustic conditions (ambient sounds)

Theoretically, it is possible to recognize speech directly from the digitized waveform, but because of the large variability of the speech signal. This is where Feature Extraction plays a role in the speech signal sector.

## Noise Cancellation

Emit a sound wave with the same amplitude, but with inverted phase.

Phase: the position of a point in time on a waveform cycle

Waveform: the shape and form of a signal

The crest of one wave meets the trough of another wave.





#### Adaptive Noise Cancellation



#### Shazam

Shazam is an app for PC, Macs and smartphones that identifies music

Mainly uses fingerprints to recognize the songs

Fingerprinting the song:

Analyzes a "chunk" of the song and get the frequency makeup of the audio

Determine which free	Hash Tag	Time in Seconds	Song	
	30 51 99 121 195	53.52	Song A by artist A	
To allow for easy acc	33 56 92 151 185	12.32	Song B by artist B	d into a hash table
	39 26 89 141 251	15.34	Song C by artist C	

## Shazam

Matching the song:

Capture the audio and perform a fingerprinting of it.

Compare the fingerprint pattern to those stored in the database (hashtable)

Commonly, the pattern will match to multiple songs.

Usually use relative timings

Allows for greater flexibility for the captured sound.



## **Music Composition**

#### Algorithmic composition

Provides notational information (sheet music)

Provides composition (music synthesis)

#### Many types of models:

Grammars - Creates distinct musical grammars. Composed of harmonies and rhythms instead of single notes

Knowledge-based systems - Isolates the aesthetic code of a certain musical genre

Evo-Devo approach - Transforms a very simple composition (of a few notes) into a fully fledged

#### **Algorithmic Music Composition**



The AI system, called FlowMachines, works by first analyzing a database of songs, and then following a particular musical style to create similar compositions. (The Beatles)

Was composed by the AI but the arrangement was done by a French composer.

#### **Emotion Recognition**

Subset of Speech Recognition

Use Neural Networks to determine emotion in a sound clip

Obtain waveform of a certain speech pattern and examine different factors to determine emotion

Pitch

Decibels

Formant



# Waveform samples of different emotions

## Emotion Recognition - How does it work?

#### 1. Feature extraction



#### 2. Feature selection

Select features that best identify a class

3. Classification



https://arxiv.org/ftp/arxiv/papers/1305/1305.1145.pdf https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition http://www-personal.umich.edu/~gowtham/bellala\_EECS452report.pdf http://willdrevo.com/fingerprinting-and-audio-recognition-with-python/ http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.437.2775&rep=rep1&type=pdf http://www.docsity.com/en/news/physics/physics-sound-visual-representation-gifs/

