Mobile Health
Lecture 5
Audio Signal and Health

Cecilia Mascolo
An Affordable Sensor
Automated Sound based Diagnostics
Voice-based Diagnostics

Voice Analysis Tech Could Diagnose Disease

Researchers enlist smartphones and machine learning to find vocal patterns that might signal post-traumatic stress disorder or even heart disease.

by Emily Mullin  Jan 19, 2017
Type of diseases for which audio has been tried on

- Depression and PTSD
- Sleep Apnea
- Wheezing (Asthma)
- Parkinson’s
- Alzheimer’s
- Autism
- Cardiovascular: coronary heart disease, arteriosclerosis
- ...
Wheezeing Detection

- Wheezeing sound is a continuously abnormal breathing sound
Preprocessing of this signal

• Band pass filter applied: (150 Hz–1000 Hz)
  • Cut out heart sound, muscle and blood interference sound.
• Raw breathing sound split into 250-ms breathing sound segments with 200-ms overlapping.
• Power spectrums of these breathing sound segments calculated by using DFT with a Hann window.

Wheezing Features

Important Features

- **ASE** = audio spectral envelope
- **TI** = tonal index
- **CF1** = correlation feature
- **ER** = Energy ratio
- **K** = Kurtosis, Difference to mean ratio
- **EVD** = Eigen Value Decomposition
- **VC** = Vector Comparison
- **LP** = Linear prediction
- **SPE** = Spectral peak entropy
- **SF** = Spectral flatness

Fig. 2. Relevance of the tested features calculated for real lung recordings using the mRMR algorithm

Frequency Spectrum Envelope

• Identifies the max frequency of each frame and creates a curve which represents all those maximum frequencies.

• Sometimes the frequency spectrum is “smoothened” first.

Figure from F. Tesser, E. Zovato, P. Cosi (2022). Statistical spectral envelop transformation applied to emotional speech. 13th Int. Conference on Digital Audio Effects. 2010.
Energy of the Signal

\[ \sum_{n=-\infty}^{\infty} |x(n)|^2 \]
Linear Prediction Coefficients

Technique that calculates coefficients of a linear prediction model which predicts the next sample of audio from a sequence of k previous samples. The coefficient of this linear model are those coefficients. This technique is used to predict pitch period accurately (ie the period at which a signal pattern repeats).
MFCCs (Mel-frequency cepstral coefficients)

Features

• Very used in audio processing

• The intuitively match our auditory way of perceiving sounds
From Spectrograms to Mel Spectrogram

• Spectrogram 1 is the output of standard DFTs.
• However humans perceive frequency of sound “logarithmically”
  • Difference between sounds at lower frequency seems more than difference of sounds at higher frequency

Whale song 😊
Mel Scale

• Mel Scale is perceptually relevant scale for frequency
  • Matching our hearing

\[ m = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \]
Spectrogram to Mel Spectrogram

- Mel spectrograms use Mel Filter banks
  - Effectively bins for the frequency mapping
- Shape retained. Fine structure smoothened.

Figure from
https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC
Mel Frequency Cepstral Coefficients

• Steps
  • Apply a DFT to generate a frequency plot from time domain plot.
  • Log the data and apply Mel Scaling.
  • Discrete Cosine Transform (a transformation which highlights the important parts of the log DFT graph).
  • Result is a number of “coefficients”: first 12-13 are the most relevant generally for audio.

Figure from https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC
Respiratory Pattern Detection from Audio

• Green/yellow line is ground truth data from chest strap.
• In-out breathing audio patterns can be visibly distinguished on time domain audio plot.

E Doheny, Ben P.F. O’Callaghan, Vitória S. Fahed, Jérémy Liegey, Cathy Goulding, Silke Ryan, Madeleine M. Lowery, Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones, Biomedical Signal Processing and Control, Volume 80, Part 1, 2023,
Breathing Pattern Detection

- Signal passed with filters to isolate the right frequencies.
  - 8th order Butterworth low-pass filter with cut-off frequency of 1 kHz.
- Elimination of signal with coughs, yawning etc: windowing and max and median peak amplitude used.
- Time domain features used to detect pauses in respiration.
Snoring-non Snoring

- Identification of voiced snoring, breathing and silence.
- Features
  - number of zero crossings in a given length of time
  - the energy of the signal
  - normalized autocorrelation coefficient at 1 ms delay
  - first predictor coefficient of linear predictive coding (LPC) analysis

Zero Crossing Rate

Measures the number of times a signal crosses over from positive to negative and vice versa (per time period). This is very correlated to frequency generally.
Autocorrelation (useful to find period)

\[ R(m) = \sum_{n=0}^{N-1-m} s(n)s(n+m) \]
Performance: Probability distribution of the features for classification

PDF of
- zero crossing
- log of energy
- autocorrelation coefficient
- first linear predictor coefficient

S = silence
UNS = Unvoiced non silence
VNS = voiced non silence
Sleep Apnea

Cessation of airflow to the lungs that lasts at least for 10s and is associated with at least 4% drop in blood’s oxygen saturation level (SaO2).
Log of energy of the tracheal sound signal is a good indicator of air flow.

Sleep Apnea Detection with Audio and SaO2 monitoring
Sleep Apnea Detection

• Classification of the sound in snore, breathing and noise
  • Use of energy and duration (a mixture)

• Spectrogram shows the snoring appearing in deep colours
Results

• AHI: apnea–hypopnea index
Sleep Stages Classification with Audio

• During sleep (in contrast to wakefulness) there is an increase of upper airway resistance due to decreased activity of the pharyngeal dilator muscles, which is reflected by amplification of air-pressure oscillations during breathing. These air-pressure oscillations are perceived as breathing sounds during sleep.

• REM (rapid-eye movement), N(on)REM, and wakefulness are associated with lack of, some, and considerable body movement.

• Breathing pattern is more periodic and consistent in deep NREM sleep compared to REM and wakefulness.

Audio

• Microphone on the bed: (Edirol R-4 pro, Bellingham, WA, USA) with a directional microphone (RØDE, NTG-1, Silverwater, NSW, Australia) was placed at a distance of one meter above the subject’s head and used for acquiring the audio signals.

• Polisomnography (PSG) for ground truth
Detection of Macro Sleep Stages (MSS)
Raw sound

Preprocessed

Spectrogram

Inhalation (blue), Exhalation (red), body movement (pink) and other (black)
Within Breathing Features

- During sleep, airways resistance is higher than during wakefulness, hence breathing efforts become greater, which translates into several factors including **louder breathing** sounds, prolonged **breathing duration**, and **different vocal sounds** (snores).

<table>
<thead>
<tr>
<th>A. Within breathing features (WB)</th>
<th>Feature code</th>
<th>count</th>
<th>importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection score of inspiration (\mu,\sigma)</td>
<td>WB_DI</td>
<td>2</td>
<td>0.093</td>
</tr>
<tr>
<td>Detection score of expiration (\mu,\sigma)</td>
<td>WB_DE</td>
<td>2</td>
<td>0.048</td>
</tr>
<tr>
<td>Detection score of respiration (\mu,\sigma)</td>
<td>WB_DR</td>
<td>2</td>
<td>0.037</td>
</tr>
<tr>
<td>Duration inspiration (\mu,\sigma)</td>
<td>WB_DurI</td>
<td>2</td>
<td>0.075</td>
</tr>
<tr>
<td>Duration expiration (\mu,\sigma)</td>
<td>WB_DurE</td>
<td>2</td>
<td>0.024</td>
</tr>
<tr>
<td>Stationarity inspiration (\mu,\sigma)</td>
<td>WB_SI</td>
<td>2</td>
<td>0.013</td>
</tr>
<tr>
<td>Stationarity expiration (\mu,\sigma)</td>
<td>WB_SE</td>
<td>2</td>
<td>0.009</td>
</tr>
<tr>
<td>Sound intensity inspiration (\mu,\sigma)</td>
<td>WB_SI1</td>
<td>2</td>
<td>0.044</td>
</tr>
<tr>
<td>Sound intensity expiration (\mu,\sigma)</td>
<td>WB_SI2</td>
<td>2</td>
<td>0.009</td>
</tr>
<tr>
<td>Sound intensity inspiration top 1% (\mu,\sigma)</td>
<td>WB_SI101</td>
<td>2</td>
<td>0.027</td>
</tr>
<tr>
<td>Sound intensity expiration top 1% (\mu,\sigma)</td>
<td>WB_SI201</td>
<td>2</td>
<td>0.053</td>
</tr>
<tr>
<td>Entropy inspiration (\mu,\sigma)</td>
<td>WB_EI</td>
<td>2</td>
<td>0.045</td>
</tr>
<tr>
<td>Entropy expiration (\mu,\sigma)</td>
<td>WB_EE</td>
<td>2</td>
<td>0.008</td>
</tr>
<tr>
<td>Frequency centroid inspiration (\mu,\sigma)</td>
<td>WB_FCI</td>
<td>2</td>
<td>0.031</td>
</tr>
<tr>
<td>Frequency centroid expiration (\mu,\sigma)</td>
<td>WB_FCE</td>
<td>2</td>
<td>0.036</td>
</tr>
<tr>
<td>Frequency bandwidth (resp., inspir., expir.)</td>
<td>WB_FB</td>
<td>3</td>
<td>0.009</td>
</tr>
</tbody>
</table>
Between Breathing Features

• Alternations in ventilation may affect fundamental respiration factors such as respiratory cycle period, respiratory duty cycle, and respiration consistency, and can be measured using sound analysis. These respiration factors are most likely to have more substantial variability during REM as opposed to NREM.

<table>
<thead>
<tr>
<th>B. Between breathing features (BB)</th>
<th>12</th>
<th>0.267</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiration duty cycle</td>
<td>BB_DCR</td>
<td>1</td>
</tr>
<tr>
<td>Inspiration duty cycle</td>
<td>BB_DCI</td>
<td>1</td>
</tr>
<tr>
<td>Expiration duty cycle</td>
<td>BB_DCE</td>
<td>1</td>
</tr>
<tr>
<td>Respiration cycle period ($\mu,\sigma$)</td>
<td>BB_RCP</td>
<td>2</td>
</tr>
<tr>
<td>Respiration cycle period consistency</td>
<td>BB_RCPC</td>
<td>1</td>
</tr>
<tr>
<td>Respiration cycle periods fourth-order curve</td>
<td>BB_RCPfit</td>
<td>5</td>
</tr>
<tr>
<td>Breathing Count</td>
<td>BB_BC</td>
<td>1</td>
</tr>
</tbody>
</table>
Body Movement Features

• Wakefulness is accompanied by relatively greater body movement, compared to NREM, while during REM sleep body movement should be absent by definition.

<table>
<thead>
<tr>
<th>C. Body movement features (BM)</th>
<th>10</th>
<th>0.054</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body movement average score</td>
<td>BM_AS</td>
<td>1</td>
</tr>
<tr>
<td>Body movement overall score percentiles</td>
<td>BM_OS</td>
<td>7</td>
</tr>
<tr>
<td>Sound intensity body movement (all curve)</td>
<td>BM_SI</td>
<td>1</td>
</tr>
<tr>
<td>Sound intensity body movement 10% (all curve)</td>
<td>BM_SI01</td>
<td>1</td>
</tr>
</tbody>
</table>
Real time Classification

![Diagram of a real-time classifier](image)
Results

One Subject

Blue = wake
Orange = REM sleep
Red = Non-REM sleep
Questions