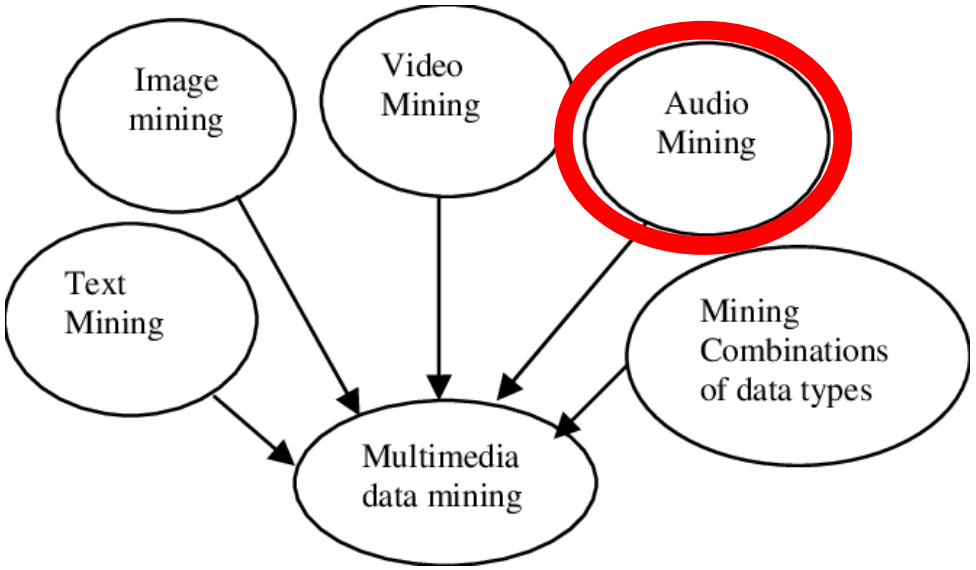
An Introduction to Audio Processing and Machine Learning

Data Mining II

Week 5

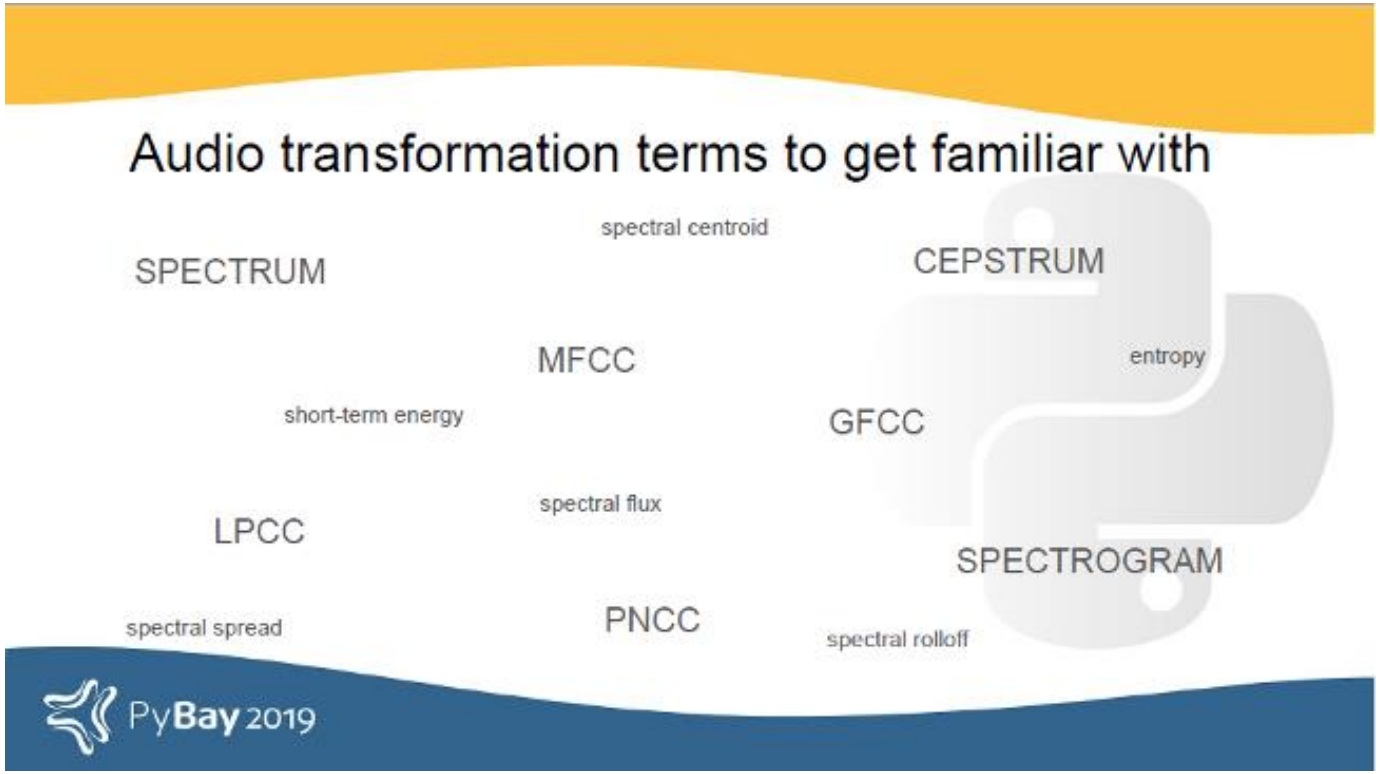


What are audio signals?

- Sinyal audio adalah sinyal yang bergetar dalam rentang frekuensi yang dapat didengar.
- Ketika seseorang berbicara, hal ini menghasilkan sinyal tekanan udara; telinga menerima perbedaan tekanan udara ini dan berkomunikasi dengan otak. Begitulah cara otak membantu seseorang mengenali bahwa sinyalnya adalah ucapan dan memahami apa yang dikatakan seseorang.

What are audio signals?

Sebelum kita masuk ke beberapa alat yang dapat digunakan untuk memproses sinyal audio dengan Python, mari kita periksa beberapa fitur audio yang berlaku untuk pemrosesan audio dan pembelajaran mesin.

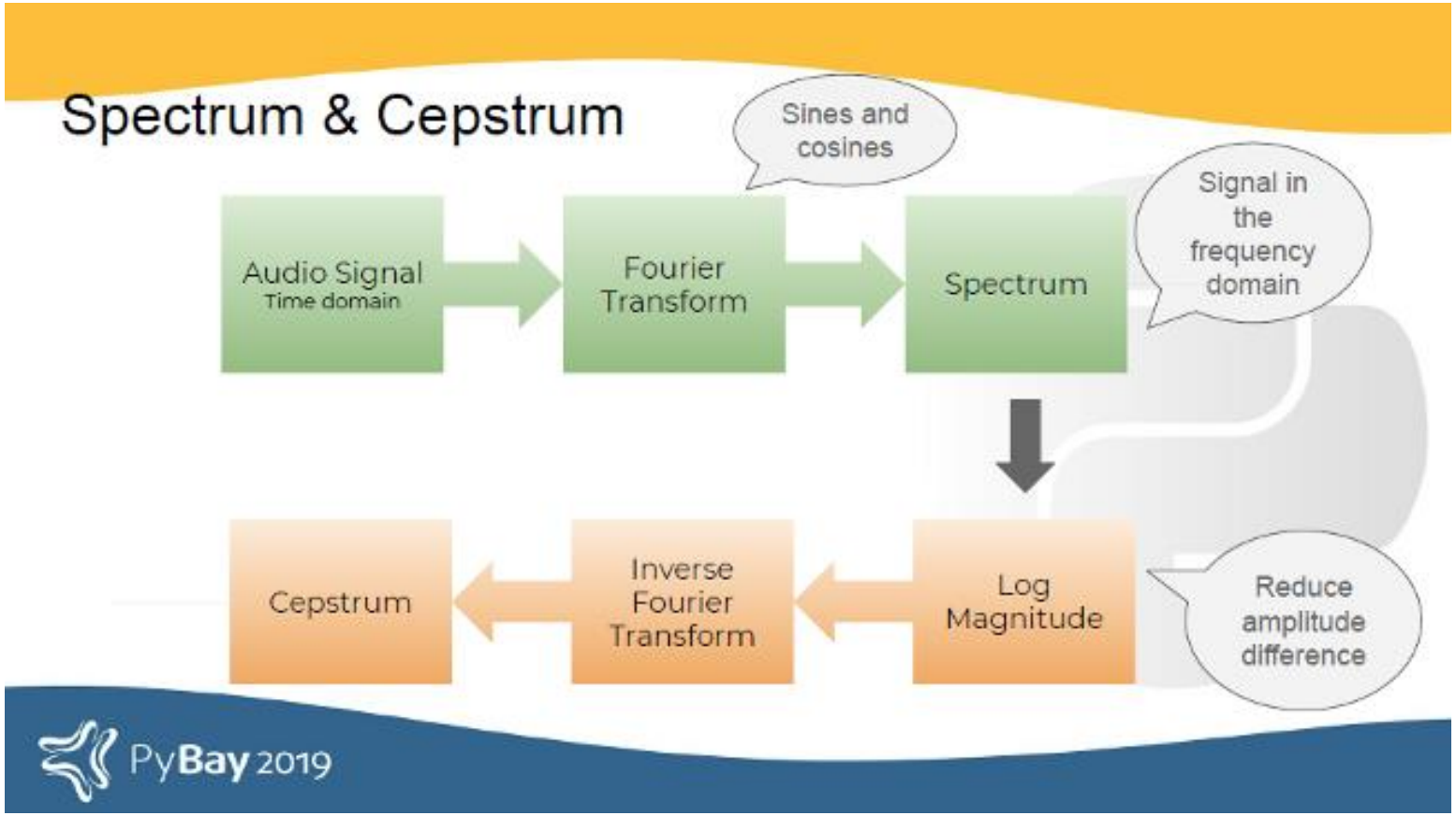


What are audio signals?

- Beberapa fitur dan transformasi data yang penting dalam pemrosesan *speech* dan audio adalah *Mel-frequency cepstral coefficients (MFCCs), Gammatone-frequency cepstral coefficients (GFCCs), Linear-prediction cepstral coefficients (LFCCs), Bark-frequency cepstral coefficients (BFCCs), Power-normalized cepstral coefficients (PNCCs), spectrum, cepstrum, spectrogram*, dan banyak lagi.
- Beberapa fitur ini dapat digunakan secara langsung dan mengekstrak fitur dari beberapa fitur lainnya, seperti spektrum, untuk melatih *(train)* model *machine learning*.

What are spectrum and cepstrum?

Spektrum dan cepstrum adalah dua fitur yang sangat penting dalam pemrosesan audio.



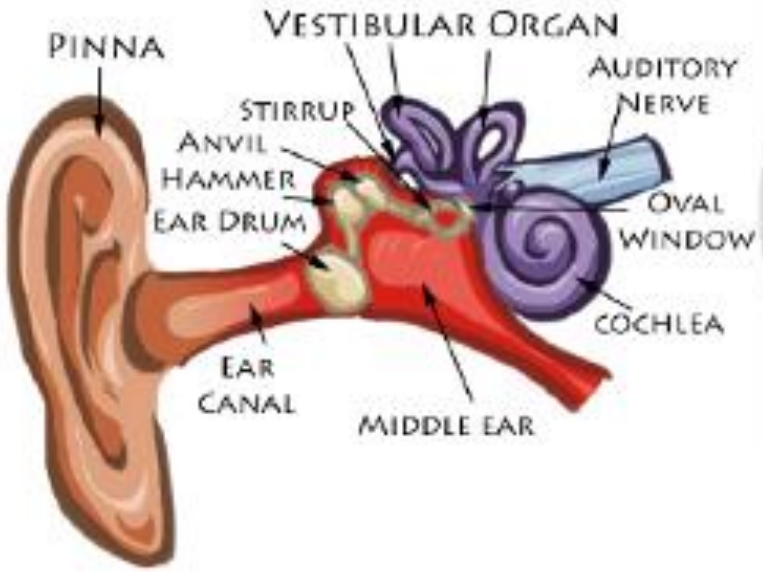
What are spectrum and cepstrum?

- Secara matematis, spektrum adalah transformasi Fourier dari sebuah sinyal. Transformasi Fourier mengubah sinyal domain waktu menjadi domain frekuensi. Dengan kata lain, spektrum adalah representasi domain frekuensi dari input sinyal audio domain waktu.
- Sebuah cepstrum dibentuk dengan mengambil log magnitudo spektrum diikuti dengan transformasi *inverse Fourier*. Hal ini menghasilkan sinyal yang tidak berada dalam domain frekuensi (karena mengambil transformasi *inverse Fourier*) atau dalam domain waktu (karena mengambil besaran log sebelum transformasi *inverse Fourier*). Domain dari sinyal yang dihasilkan disebut *quefrency*.

Apa hubungannya dengan pendengaran?

- Alasan kita memperhatikan sinyal dalam domain frekuensi berkaitan dengan kondisi biologis telinga.
- Banyak hal yang harus terjadi sebelum kita dapat memproses dan menginterpretasikan sebuah suara.
- Salah satunya terjadi di koklea, bagian telinga yang berisi cairan dengan ribuan rambut kecil yang terhubung ke saraf. Beberapa rambut pendek, dan beberapa relatif lebih panjang. Rambut yang lebih pendek beresonansi dengan frekuensi suara yang lebih tinggi, dan rambut yang lebih panjang beresonansi dengan frekuensi suara yang lebih rendah.
- Oleh karena itu, telinga seperti penganalisis transformasi Fourier alami (*a natural Fourier transform analyzer*)

How do we hear?



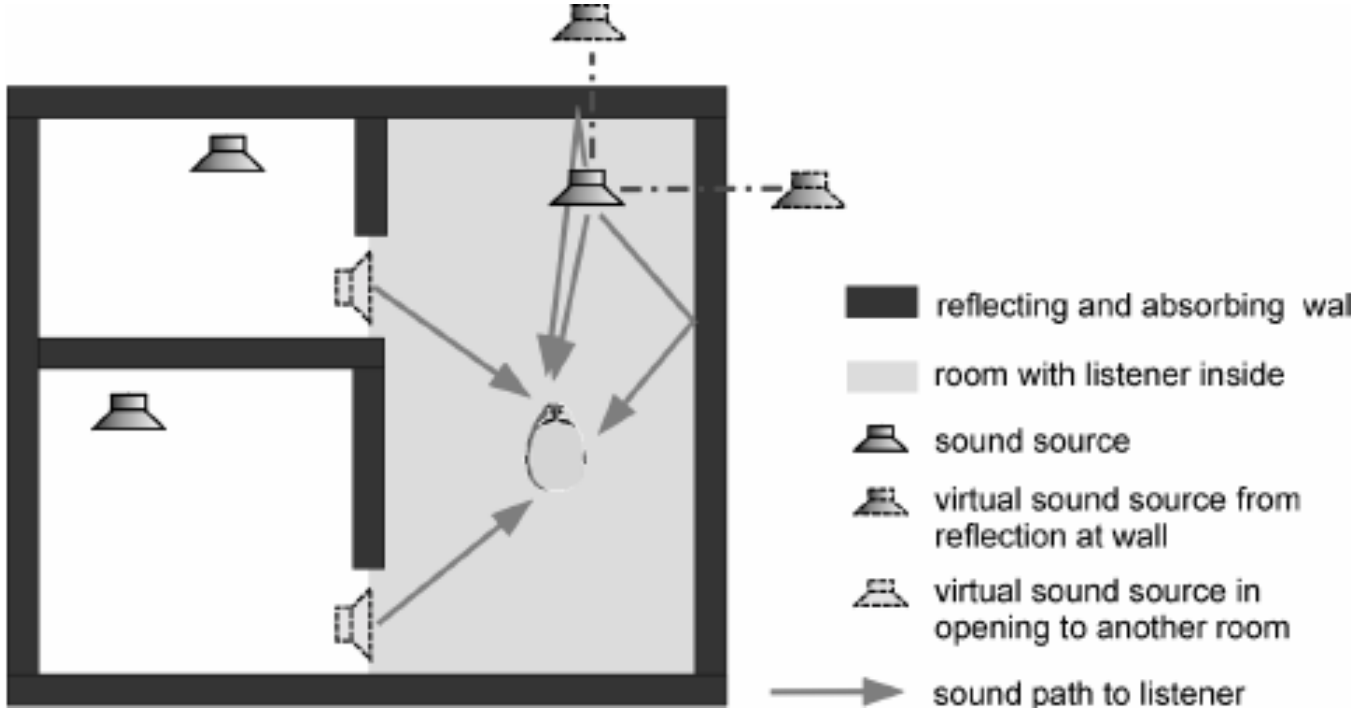
Spiral of tissue with liquid and thousands of tiny hairs that gradually get smaller.

- Each hair is connected to a nerve.
- Longer hair resonate with lower frequencies.
- Shorter hair resonate with higher frequencies.
- Thus the time-domain air pressure signal is transformed into frequency spectrum, which is then processed by the brain.

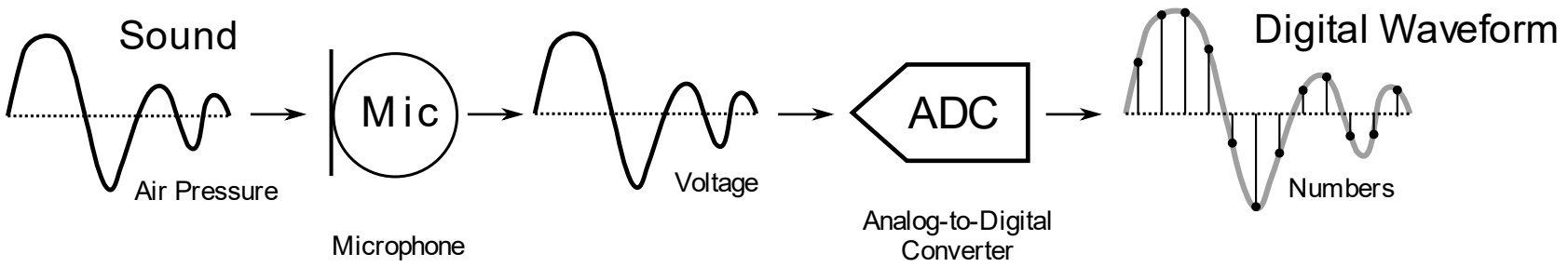
Our ear is a natural fourier transform analyzer!

PyBay 2019

- Hal pertama yang penting adalah bahwa suara hampir selalu, atau pada dasarnya selalu, merupakan campuran.
- Karena suara ini akan bergerak di tikungan tempat, tidak seperti gambar misalnya.
- Sehingga pendengar akan selalu memiliki suara yang berasal dari banyak tempat. Suara ini juga akan dibawa ke dalam tanah dan dipantulkan oleh dinding.
- Semua hal ini membuat pendengar selalu memiliki banyak sumber suara: sumber yang menarik dan kemudian selalu sumber suara lainnya (noise).

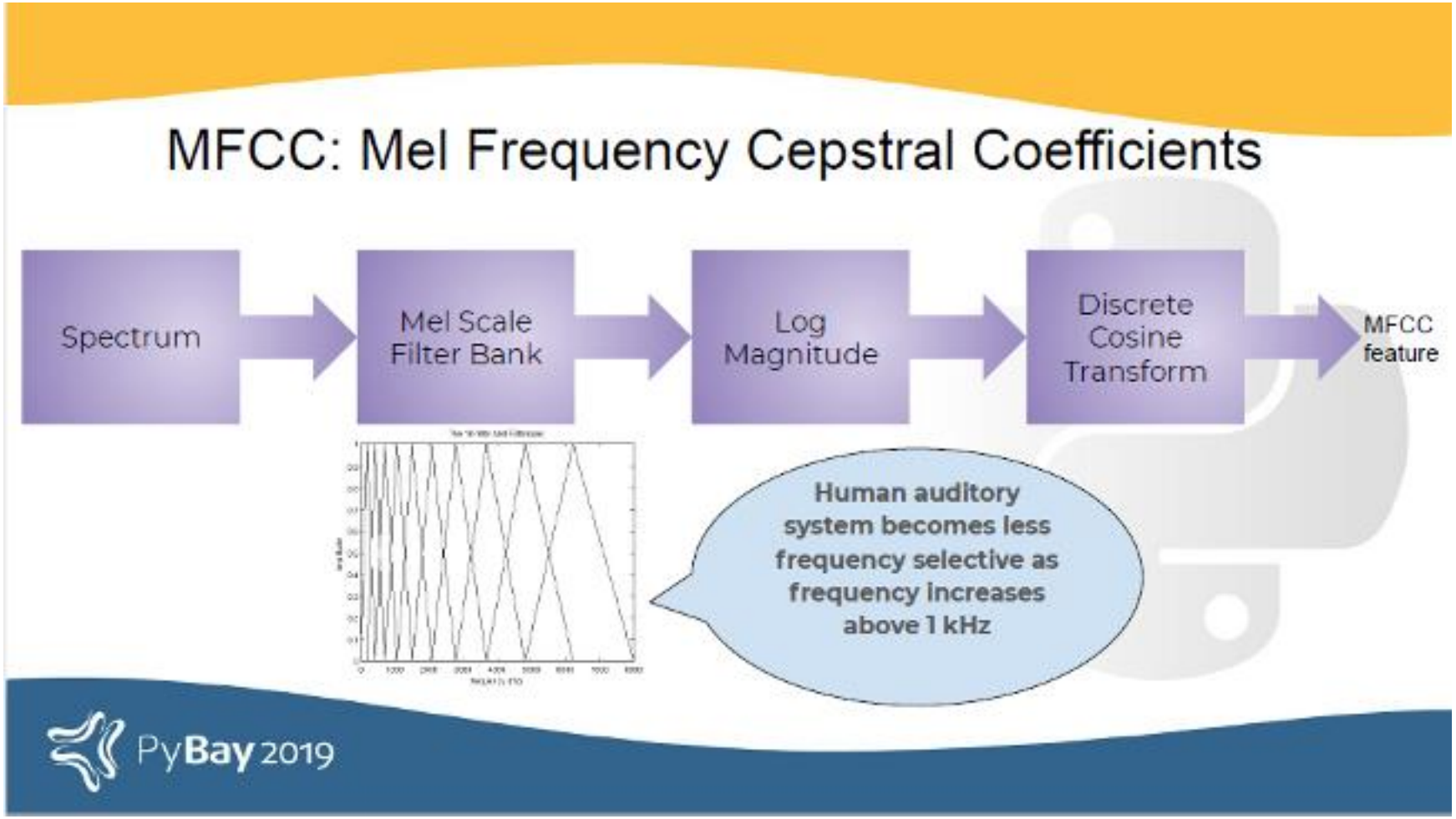


Audio Aquisition



- Secara fisik kita memiliki suara sebagai variasi tekanan udara.
- Kita menggunakan mikrofon untuk mengubah suara menjadi tegangan listrik, sebuah *Analog-to-Digital converted* (ADC) dan kemudian kita memiliki bentuk gelombang digital *(digital waveform).*
- Sebagai audio digital, hal itu *quantized* dalam waktu misalnya dengan *sampling rate* dan amplitudo.
- Kita biasanya berfokus pada *mono primarily* dengan *one channel* ketika melakukan Klasifikasi Audio *(Audio Classification).*
- Ada beberapa metode seputar stereo tetapi tidak digunakan secara luas, dan juga lebih banyak saluran *(channel).*
- Kita biasanya menggunakan format *uncompressed* karena yang paling aman.
- Meskipun dalam kondisi nyata kita mungkin juga memiliki data terkompresi, yang mungkin akan memengaruhi model.
- Jadi setelah kita memiliki *waveform*, kita dapat mengubahnya menjadi spektogram.

Fakta lain tentang pendengaran manusia adalah ketika frekuensi suara meningkat di atas 1kHz, telinga kita mulai kurang selektif terhadap frekuensi. Kondisi ini sesuai dengan sesuatu yang disebut bank filter Mel *(Mel filter bank).*

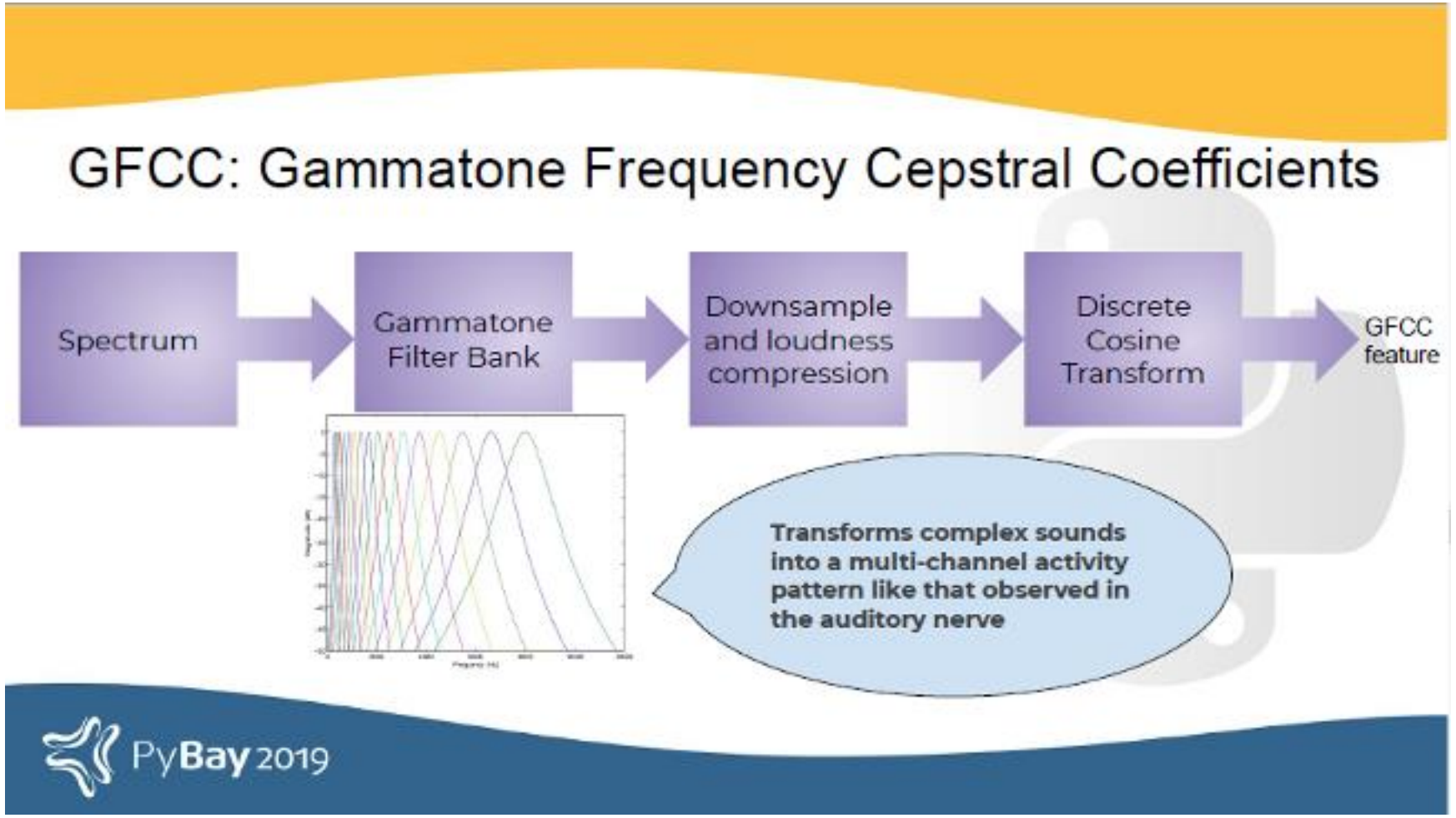


Melewati spektrum melalui bank filter Mel *(Mel filter bank)*, diikuti dengan mengambil besaran log dan *discrete cosine transform* (DCT) menghasilkan Mel cepstrum.

DCT mengekstrak informasi utama dan puncak *(peak)* sinyal. Hal ini juga banyak digunakan dalam kompresi JPEG dan MPEG.

Puncak *(peak)* adalah inti dari informasi audio. Biasanya, 13 koefisien pertama yang diekstraksi dari Mel cepstrum disebut MFCC. Hal ini menyimpan informasi yang sangat berguna tentang audio dan sering digunakan untuk melatih model *machine learning*.

Filter lain yang terinspirasi oleh pendengaran manusia adalah bank filter Gammatone *(Gammatone filter bank).* Filter bank ini digunakan sebagai simulasi front-end koklea. Dengan demikian, ia memiliki banyak aplikasi dalam *speech processing* karena bertujuan untuk meniru cara kita mendengar.



GFCC dibentuk dengan melewatkan spektrum melalui bank filter Gammatone, diikuti oleh kompresi kenyaringan dan DCT. Yang pertama (kurang lebih) 22 fitur disebut GFCC. GFCC memiliki sejumlah aplikasi dalam *speech processing*, seperti *speaker identification*.

Fitur lain yang berguna dalam tugas pemrosesan audio (*especially speech*) termasuk LPCC, BFCC, PNCC, and spectral features like spectral flux, entropy, roll off, centroid, spread, and energy entropy.

Implementasi

- Ada beberapa subbidang audio yang sangat dikenal, pengenalan suara *(speech recognition)* adalah salah satunya.
- Untuk tugas klasifikasi, misalnya pencarian kata kunci, jadi: "Hi Siri" atau "OK Google".
- Dalam analisis musik contohnya adalah klasifikasi genre.
- Dalam ekoakustik, misalnya yaitu menganalisis migrasi burung menggunakan data sensor untuk melihat polanya.
- Mendeteksi pemburu liar di kawasan lindung untuk memastikan bahwa tidak ada orang yang benar-benar berkeliling menembak di tempat yang seharusnya tidak ada tembakan.
- Kontrol kualitas di bidang manufaktur. Petugas *quality control* tidak harus masuk ke dalam peralatan atau produk yang sedang diuji, cukup mendengarkannya dari luar.
- Menguji kursi mobil listrik yaitu dengan memeriksa apakah semua motor berjalan dengan benar.
- Dalam keamanan digunakan untuk membantu memantau CCTV dalam jumlah besar dengan menganalisis audio.
- Dan dalam medis misalnya untuk mendeteksi murmur jantung (suara darah yang mengalir dalam jantung) yang bisa menjadi indikasi kondisi jantung.

Referensi

- https://towardsdatascience.com/machine-learning-on-sound-and-audio-data-3ae03bcf5095
- https://www.jonnor.com/2021/12/audio-classification-with-machine-learning-europython-2019/
- https://opensource.com/article/19/9/audio-processing-machine-learning-python

Terima kasih