IPUT TOKYO

画像音声認識-21,22

- 音声認識の特徴抽出

6/26,29/2023

上條浩一

東京国際工科専門職大学
International Professional University of Technology in Tokyo

授業後
公開

小テストの答え-1

問1(3点)：以下の5つの物質/媒体中を伝わる音の速さを、速い順に、数字をカンマで区切って答えよ。解答例：1,2,3,4,5

1. 氷
2. 空気
3. 鉄
4. 真空
5. 密度 100kg/m^3, 弾性率1.6x10^9の物質

答え：5は音速=$\sqrt{1.6 \times 10^9/100} = 4000$(m/sec)

3.鉄(5290)>5.(4000)> 1.氷(3940)> 2.空気(341)>4.真空(0)

授業後
公開

小テストの答え-2

問2(3点)：以下より、音の３大要素を３つ選べ

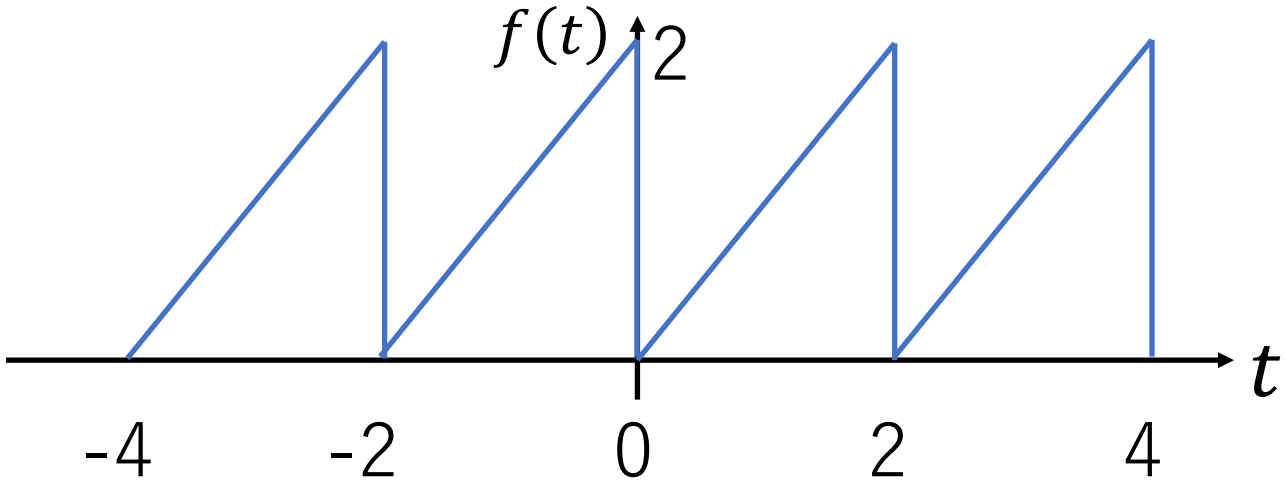
1. 強さ（dB)
2. 高さ（Hz)
3. 速さ（m/sec)
4. 音素
5. 音色
6. 長さ（sec)

答え：1,2,5

授業後
公開

小テストの答え-3

問3(4点)：以下の、周期性を持つ関数(f(t))を、ガウス記号[ ] を使って１つの式で表せ。但し、[x]はxを超えない最大の整数。
ヒント： $4 \leq t < 6$の時 $f(t) = t - 4, 6 \leq t < 8$の時 $f(t) = t - 6$
先週の資料p26

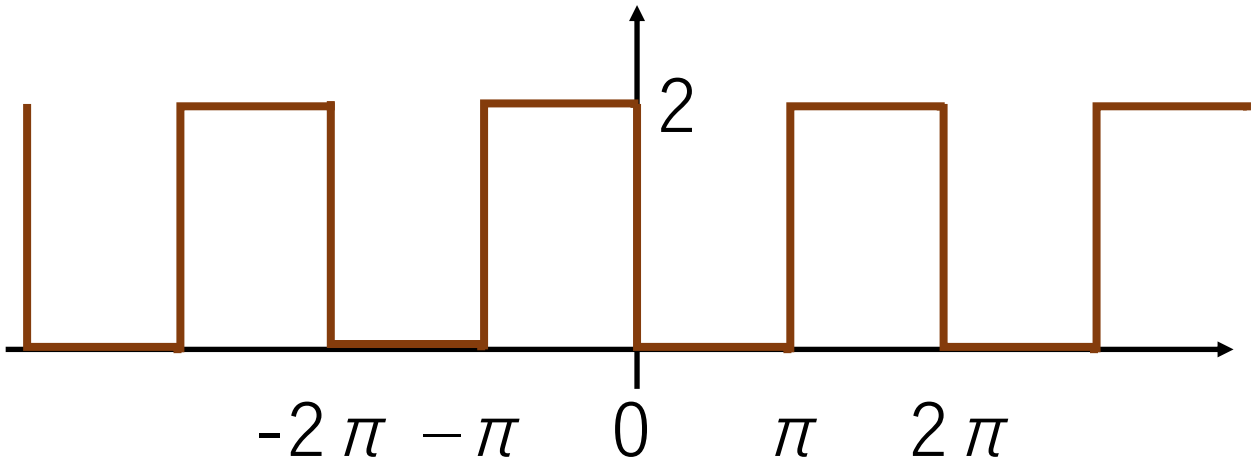


$$f(t) = t\ (0 \leq t < 2),$$
$$f(t+2) = f(t)$$

$$f(t) = t - 2[t/2]$$

宿題10

以下の矩形波をフーリエ級数展開し、 $f(t) = a_0 + \sum_{n=1}^{\infty}(a_n cos(nt) + b_n sin(nt))$ としたときの、a_n b_nを
(1)a_0, (2)a_n (n>0), (3)b_n (n=奇数), (4)b_n(n=偶数)
に分けて答えよ



$$f(t) = \begin{cases} 2 & (-\pi \le t < 0) \\ 0 & (0 \le t < \pi) \end{cases},$$
$$f(t + 2\pi) = f(t)$$

締切：B:6/24(土), A:6/26(月), どちらも9:00

授業後
公開

宿題10 答え

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)dt = 1$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} f(t)\cos nt\,dt = 0$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} f(t)\sin nt\,dt = \begin{cases} -\dfrac{4}{n\pi} & (odd) \\ 0 & (even) \end{cases}$$

$$f(t) = 1 - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)\,t}{2k+1}$$

授業後公開

宿題10 別解

- 例題の関数に１を足して、位相をπずらせばよいので、例題の関数をg(t)とすると、演習19-1は、

$$f(t) = (g(t+\pi)+1) = \left(\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)(t+\pi)}{2k+1}+1\right)$$

$$\sin(a+b) = \sin a \cos b + \sin b \cos a$$

を使うと、

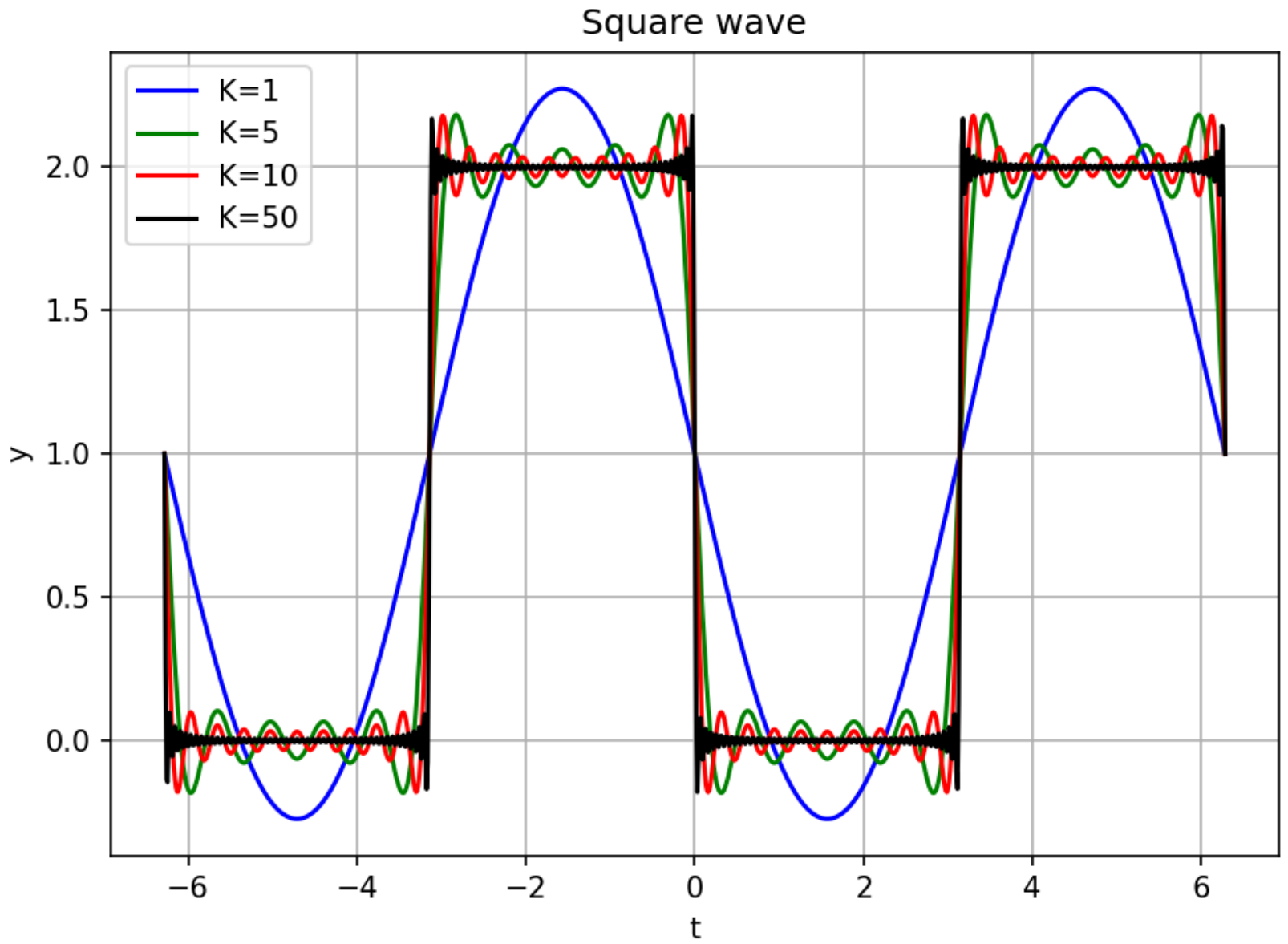
$$\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)(t+\pi)}{2k+1}+1 = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)t\,\cos(2k+1)\pi}{2k+1}+1$$

-1

$$= 1 - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)t}{2k+1}$$

授業後
公開

宿題10 グラフ



授業後公開

宿題10 良くある間違い

$$a_0 = 2$$

- 2π でなく π で割っている。
- Net上に、$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)dt$　となっているものがあるが、この場合、

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos(nt) + b_n sin\,(nt))$$

となっている。（例：Wikipedia, https://ja.wikipedia.org/wiki/%E3%83%95%E3%83%BC%E3%83%AA%E3%82%A8%E7%B4%9A%E6%95%B0 ）

例えば、 $f(t) = 1$ を考えれば解る。

前回の復習

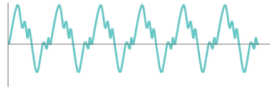
- 音とは
  - ➢音の３大要素
- 音声とは
  - ➢音素
  - ➢スペクトル
- 音声認識とは
  - ➢音声認識の歴史
- フーリエ級数

本日のゴール
音声認識の前処理、特徴量抽出

- 音声処理の全体の流れ
- 音声処理の歴史
- 前処理
  - ➢デジタル化
  - ➢高域強調
  - ➢量子化
- オイラーの公式
- 特徴量抽出
  - ➢スペクトル分析
  - ➢ケプストラム分析
  - ➢MFCC
- DPマッチング
- 音響モデル（前半）
  - ➢特徴量から単語列の抽出
  - ➢ベイズの定理

音声認識全体の流れ

音声波形



ディジタル化 → 高域強調 → 量子化

連続値→離散値　　高い音を強く　　（例）16bitデータに

前処理

スペクトラム分析 → ケプストラム分析

周波数成分を抽出　　ノイズなどを除去

特徴量抽出

X

音声のベクトル抽出

X:音声特徴量（ベクトル）

W:単語列

音響モデル → 言語モデル → 文章

p(X|W)

p(W)

これはペンです
This is a pen

音声認識の歴史の変遷

表 1　音声認識の方法論の変遷

第 1 世代	1950～1960 年代	ヒューリスティック(経験則)
第 2 世代	1960～1980 年代	テンプレート　(DP マッチング，オートマトン)
第 3 世代	1980～1990 年代	統計モデル　(GMM-HMM, N-gram)
3.5 世代	1990～2000 年代	統計モデルの識別学習
第 4 世代	2010 年代	ニューラルネット　(DNN-HMM, RNN)
4.5 世代	2015 年～	ニューラルネットによる End-to-End

河原達也. "音声認識技術の変遷と最先端――深層学習による End-to-End モデル――." *日本音響学会誌* 74.7 (2018): 381-386.

http://sap.ist.i.kyoto-u.ac.jp/lab/project/paper/ASJ18-7.pdf

ニューラルネットによるEnd-to-End

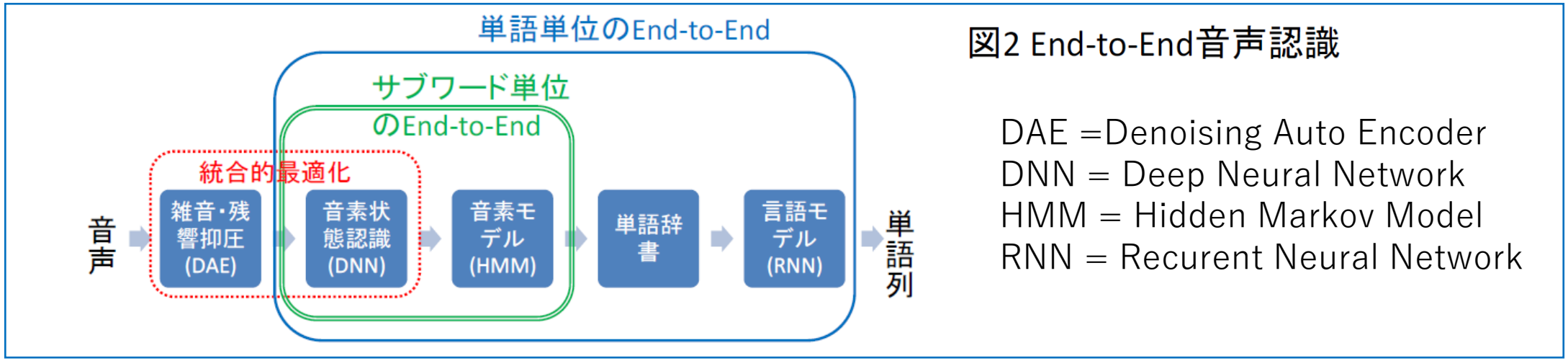


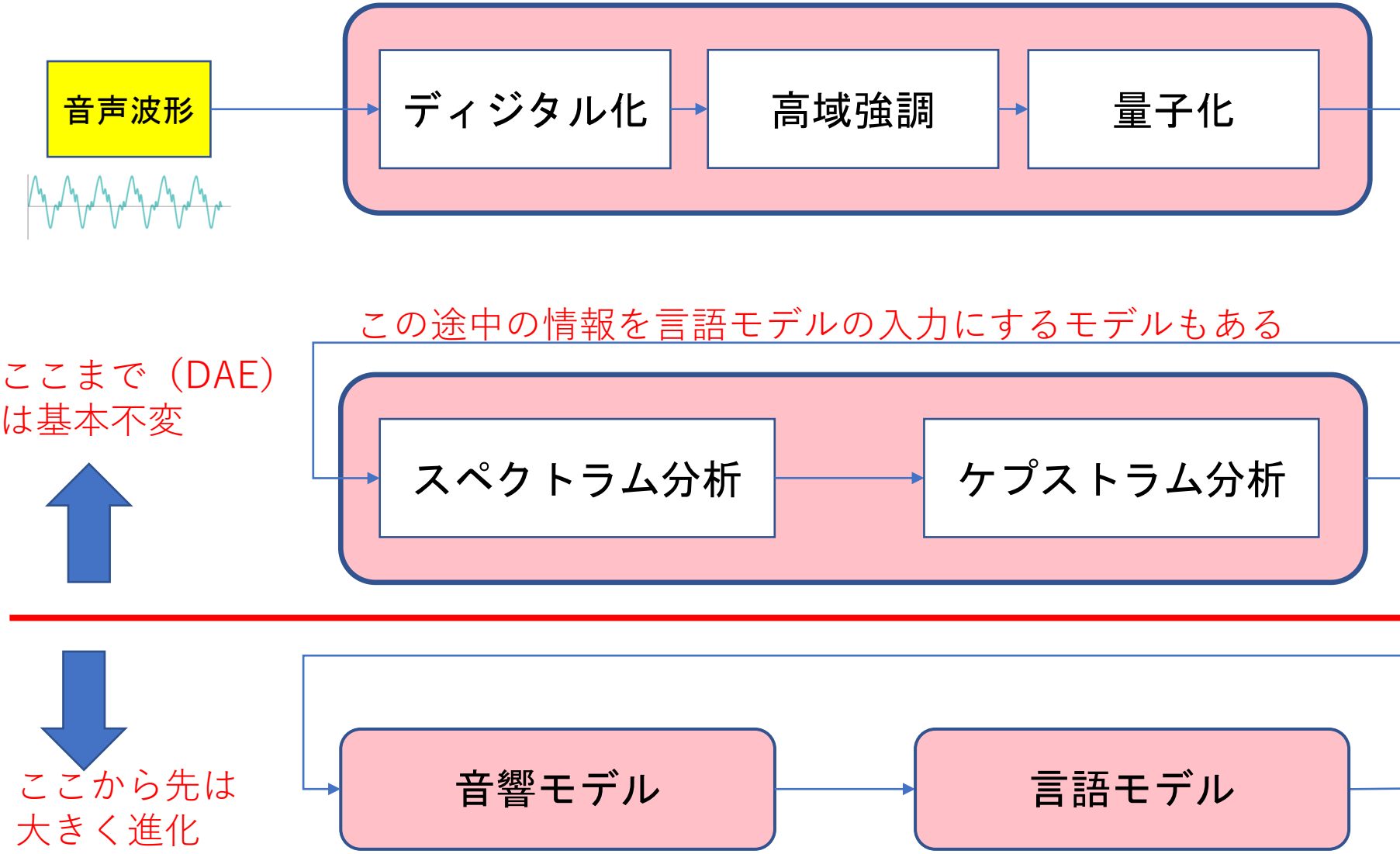
図2 End-to-End音声認識

DAE =Denoising Auto Encoder
DNN = Deep Neural Network
HMM = Hidden Markov Model
RNN = Recurent Neural Network

河原達也. "音声認識技術の変遷と最先端――深層学習による End-to-End モデル――." *日本音響学会誌* 74.7 (2018): 381-386.

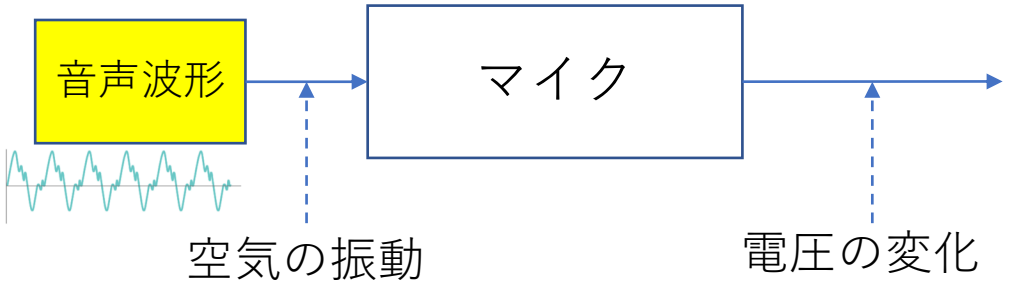
http://sap.ist.i.kyoto-u.ac.jp/lab/project/paper/ASJ18-7.pdf

音声認識全体の流れ（再掲）



前処理

音声のディジタル化-1



- アナログ（連続）値のため、この状態では計算機では扱えない
- ディジタルにする必要がある

アナログディジタル変換（AD変換）

高域強調

- 音声のパワーは高域（高周波数）になるほど減衰
- それを補償するため、約6dB/octの高域強調を行う
  - ➢oct = 2倍。例：オクターブ
  - ➢10Hz →100Hz, $6\log_2 10 = 19.9$dB
  - ➢$H(z) = 1 - \alpha z^{-1}, z = \exp(jw)$
  - ➢ α =0.97がよく使われる

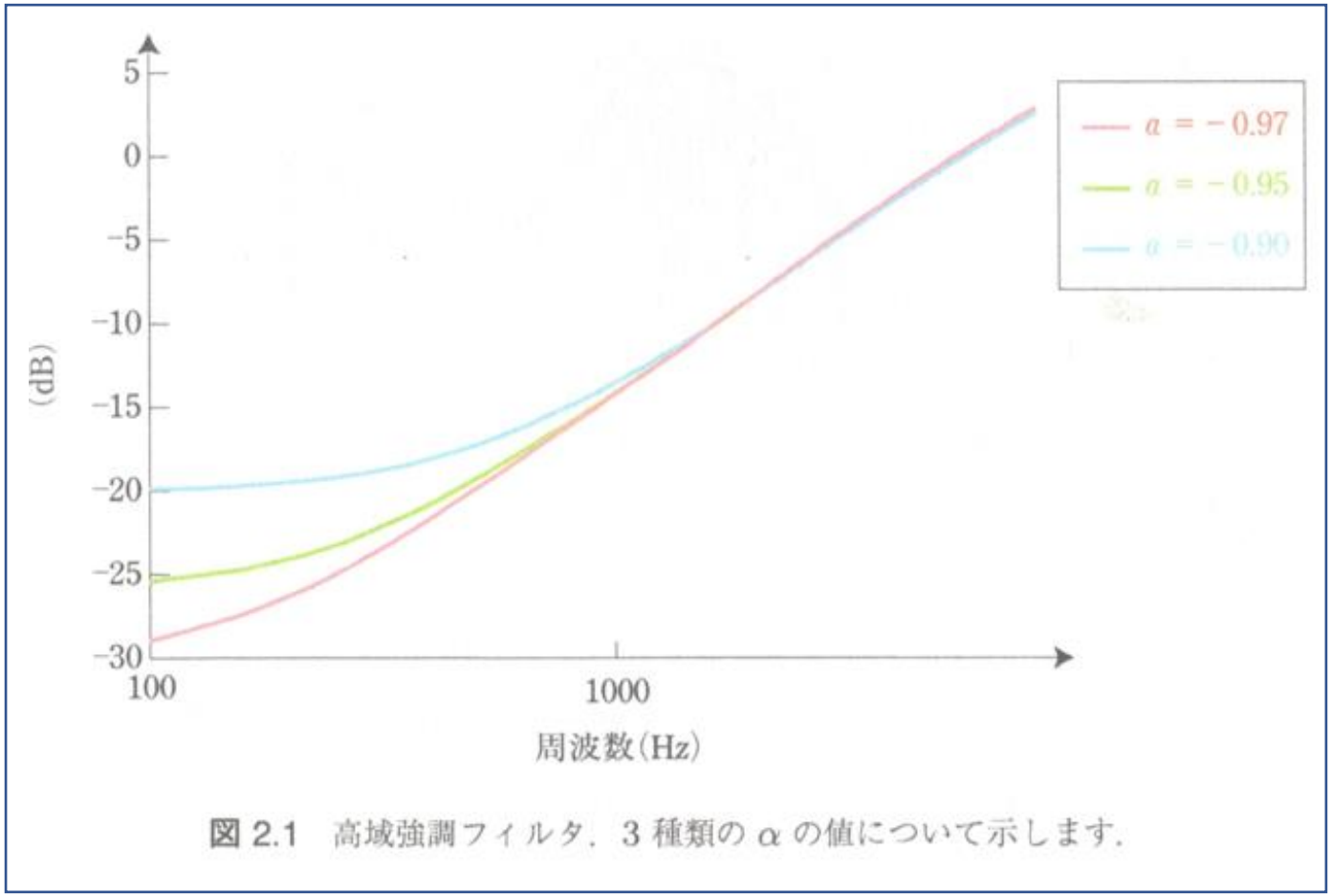


図 2.1　高域強調フィルタ．3 種類の α の値について示します．

篠田浩一、音声処理、p13

数学の知識-オイラーの公式

- オイラーの公式: $i=\sqrt{-1}$ として、 $e^{i\theta}=i\sin\theta+\cos\theta$
- 信号処理でよく使われる。 $j=\sqrt{-1}$ とする場合もある
- 例

$$e^{i\pi/3}=\cos\left(\frac{\pi}{3}\right)+i\sin\left(\frac{\pi}{3}\right)=\frac{1}{2}+\frac{\sqrt{3}}{2}i$$

$$|e^{i\theta}|^2=\sin^2\theta+\cos^2\theta=1$$

絶対値→大きさ

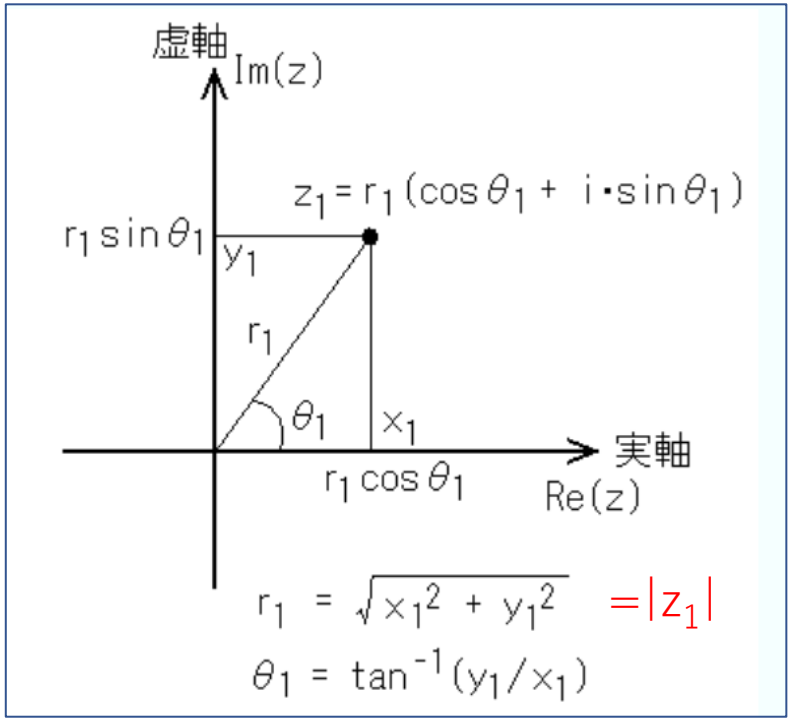
i の成分とそれ以外の成分の2乗和

応用

$$(e^b)^c=e^{bc}$$

ここは1

$$|e^{ia+b}|^c=|e^{ia}e^b|^c=e^{bc}$$



実軸（x)と虚軸(y)上の長さ

Θは、位相のずれ

http://dbkids.co.jp/popimaging/seminar/complex/complexplane.htm

演習21-1 (LMS提出)

- 以下の値を求め、カンマ(,)で分けて、LMSに回答せよ。 π はpi、e^aは exp(a) or e^a, $\sqrt{a}$はsqrt(a)で大丈夫です

1. $e^{i\pi/6}$

2. $|e^{(i/4+3)\pi}|^2$

演習21-1 解答

- 以下の値を求め、カンマ(,)で分けて、LMSに回答せよ。π はpi、e^aは exp(a), $\sqrt{a}$はsqrt(a)で大丈夫です

1. $e^{i\pi/6} = \cos\left(\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2} + \frac{i}{2}$

2. $|e^{(i/4+3)\pi}|^2 = |e^{i\pi/4}e^{3\pi}|^2 = e^{6\pi}$

音声信号の表記方法

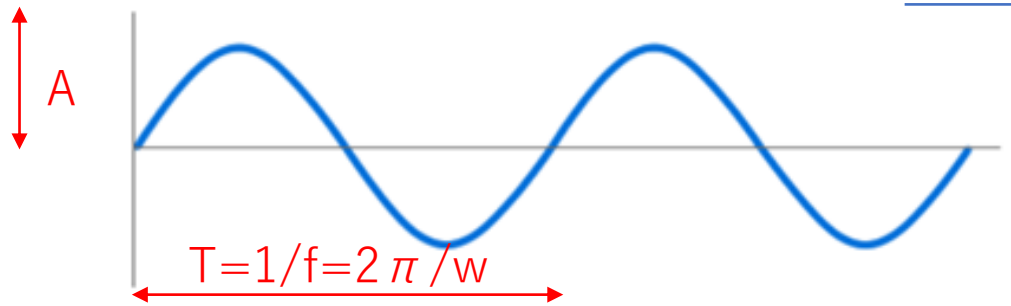
例えば、強さA、周期Tの音声信号 f(t)は、sin, cosは使わずに、シンプルに

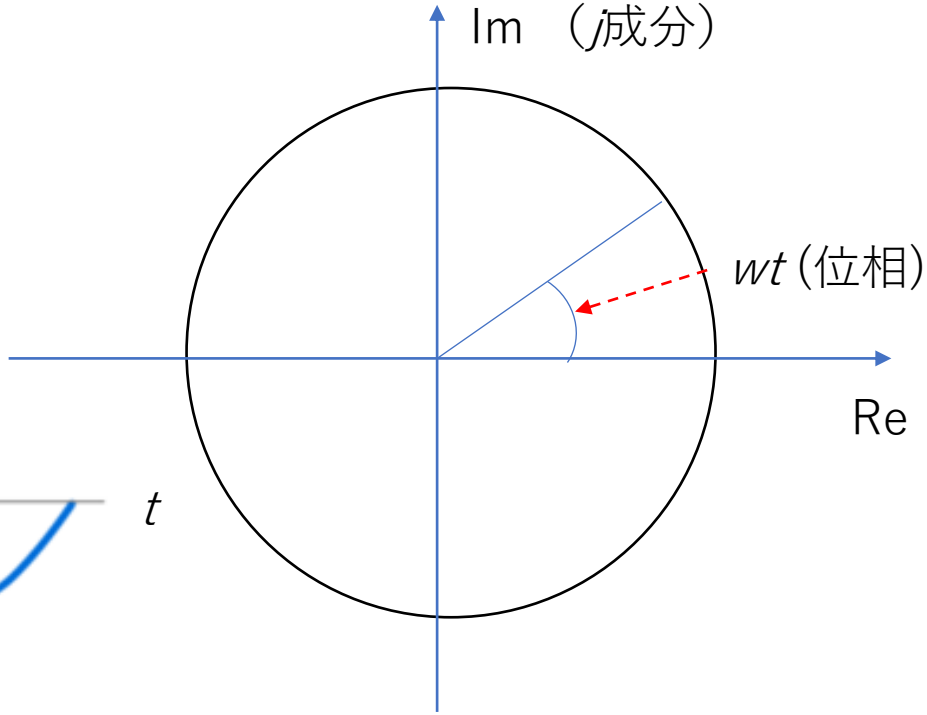
$$f(t) = A\exp(jwt) = Ae^{jwt} \ (= A(\cos(wt) + j\sin(wt)))$$

と表記することが多い $(w = 2\pi f, f = 1/T)$

この場合、パワースペクトル（信号強度）は、

$$|Ae^{jwt}| = A\sqrt{\cos wt^2 + \sin wt^2} = A$$





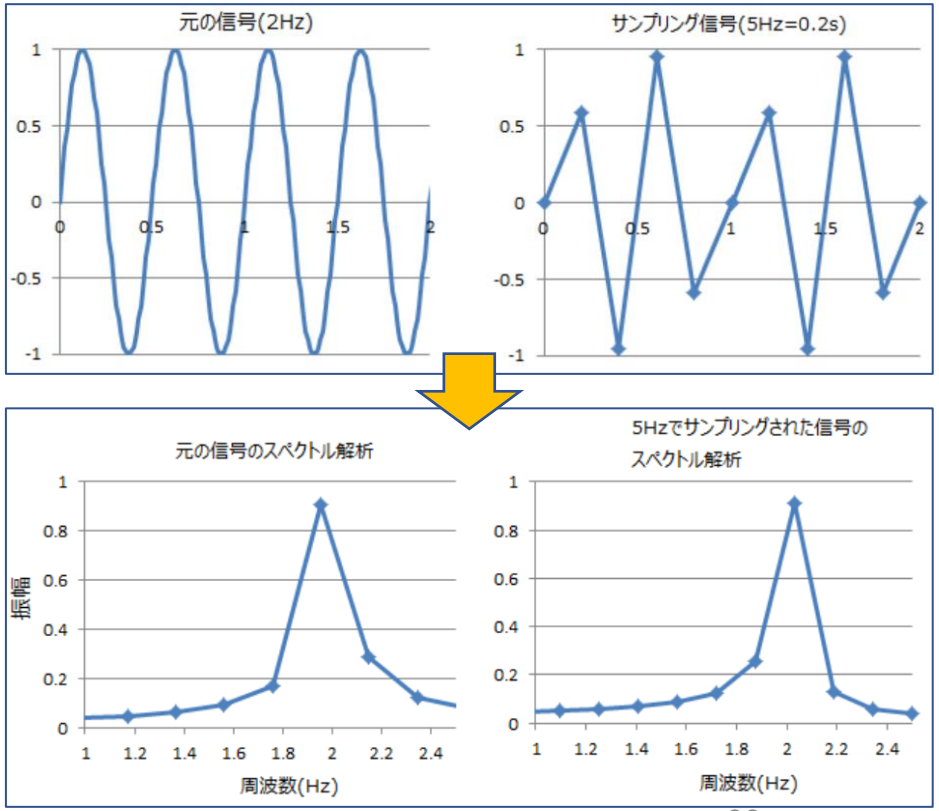
標本化(サンプリング)定理

- 音の時間変化x(t)が、W(Hz)未満の帯域に制限されているとき、その倍の周波数（つまり、2WHz → $T \leq 1/(2W)$)で標本化（サンプリング）すれば、完全に復元ができる➔**標本化（サンプリング）定理**
  - ➢証明：https://ja.wikipedia.org/wiki/%E6%A8%99%E6%9C%AC%E5%8C%96%E5%AE%9A%E7%90%86 等

定義 2.1（サンプリング定理）

$x(t)$ が 0 (Hz) 以上，W (Hz) 未満の帯域に制限されているとき，$x(t)$ を $T \leq 1/(2W)$ (s) ごとに標本化すれば，次式を用いて，標本値系列からもとの波形が完全に再現できる．

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin(\frac{\pi}{T}(t-nT))}{\frac{\pi}{T}(t-nT)}$$



- Wを**ナイキスト周波数**　という

篠田浩一、音声処理、p13

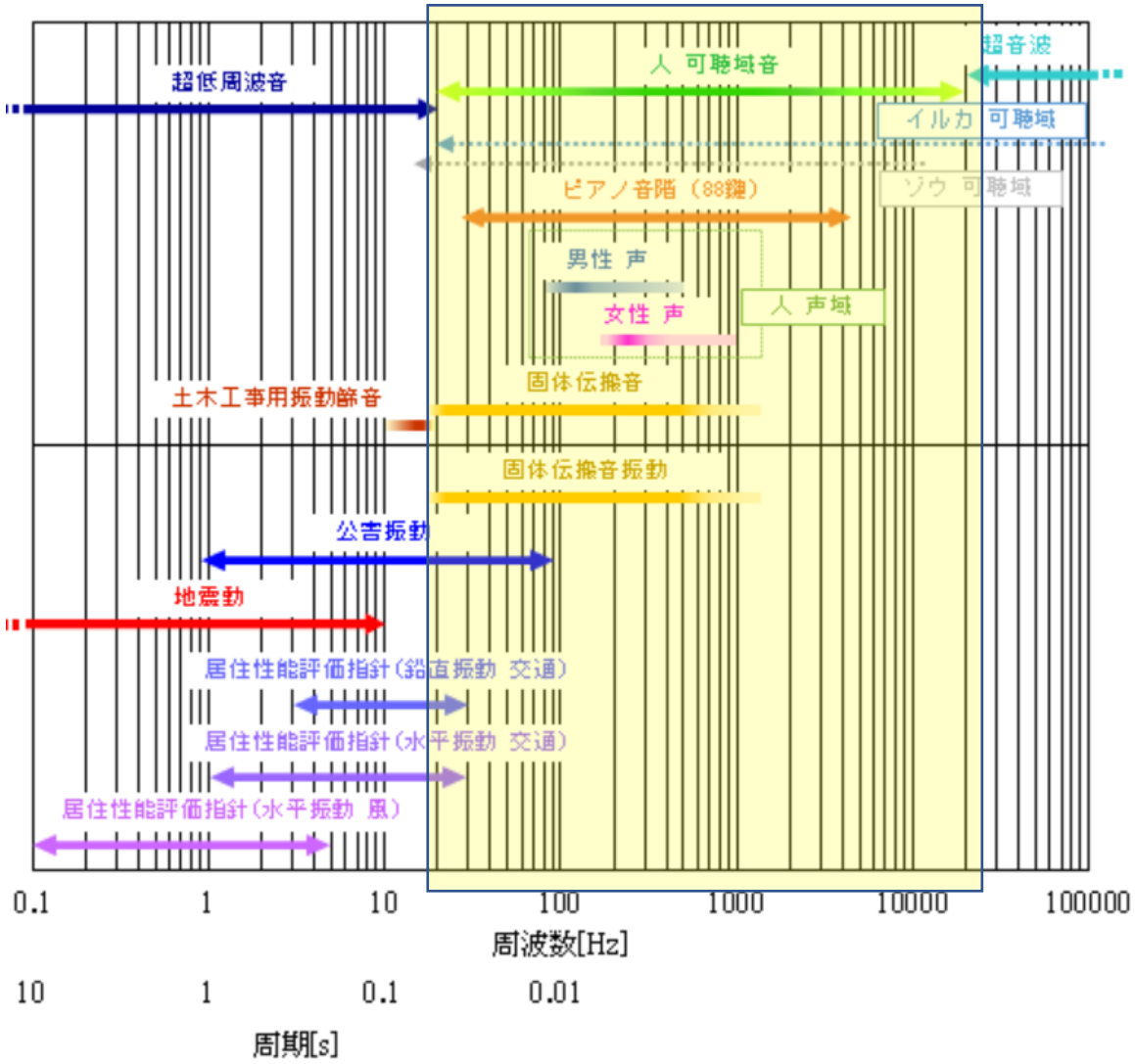
http://arduinopid.web.fc2.com/M34.html

どのぐらいでサンプリングすればよいか？

- 人は20Hz~20kHzが可聴範囲

- CD：サンプリング周波数=44.1kHz
  - ➢22.05kHzで十分可聴域

- 音声：8kHz以下にほとんどの音韻が含まれている
  - ➢16kHzの精度で十分



https://hochouki.soudan-anshin.com/cont/sound/

https://www.yacmo.co.jp/catlog/tabid/420/Default.aspx

http://arduinopid.web.fc2.com/M34.html

量子化

- 標本化：連続した時間を離散化
- **量子化**：連続した振幅幅を離散化
- $G(dB) = 20\log_{10}\frac{p}{20}$ (p:音圧 uPa),
- 0dB(p=20)を最低レベルとして、16bit→p=20*2^{16} →96dBまでカバー

騒音レベル[dB]		音の大きさのめやす	
極めてうるさい	140	ジェットエンジンの近く	聴覚機能に異常をきたす
	130	肉体的な苦痛を感じる限界	
	120	飛行機のプロペラエンジンの直前・近くの雷鳴	
	110	ヘリコプターの近く・自動車のクラクションの直前	
	100	電車が通る時のガード下・自動車のクラクション	
	90	大声・犬の鳴き声・大声による独唱・騒々しい工場内	極めてうるさい
	80	地下鉄の車内（窓を開けたとき）・ピアノの音　聴力障害の限界	
うるさい	70	掃除機・騒々しい街頭・キータイプの音	うるさい
	60	普通の会話・チャイム・時速40キロで走る自動車の内部	
普通	50	エアコンの室外機・静かな事務所	日常生活で望ましい範囲
	40	静かな住宅地・深夜の市内・図書館	
静か	30	ささやき声・深夜の郊外	静か
	20	ささやき・木の葉のふれあう音	

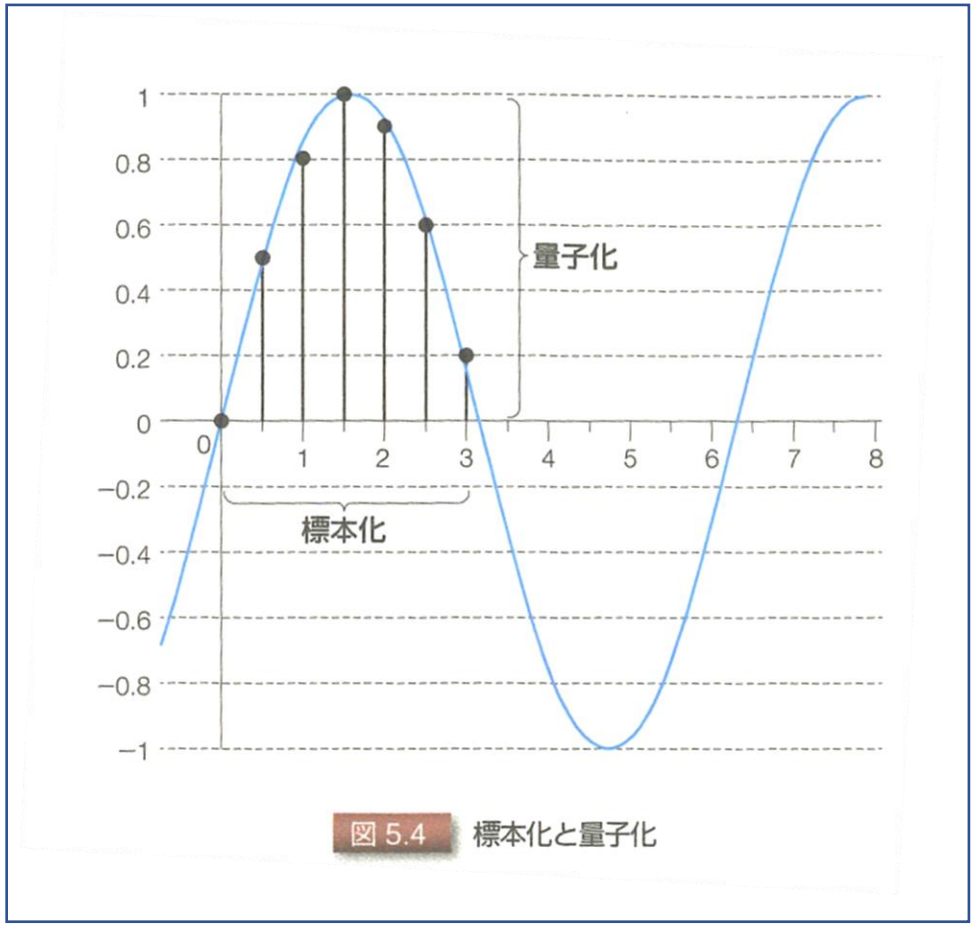


図 5.4　標本化と量子化

教科書p63 図5.4

https://bto-pc.jp/select/silent-pc-decibel-datums.html

特徴量抽出

スペクトル分析 (Spectrum analysis)

- デジタル化された音声信号に対して、一定の長さの信号を取り出し、その中に含まれる波の成分を分析
  - ➢単純に波をぶつ切りにすると、両端の所でいきなり0に変化することになり、もとの波にはない性質が出てきてしまうので、分析したい幅よりも広い範囲で信号を切り出し、端になるほど減衰するフィルタをかける
  - ➢フレームとして切り出した音声信号にフーリエ変換を行い周波数成分を求める

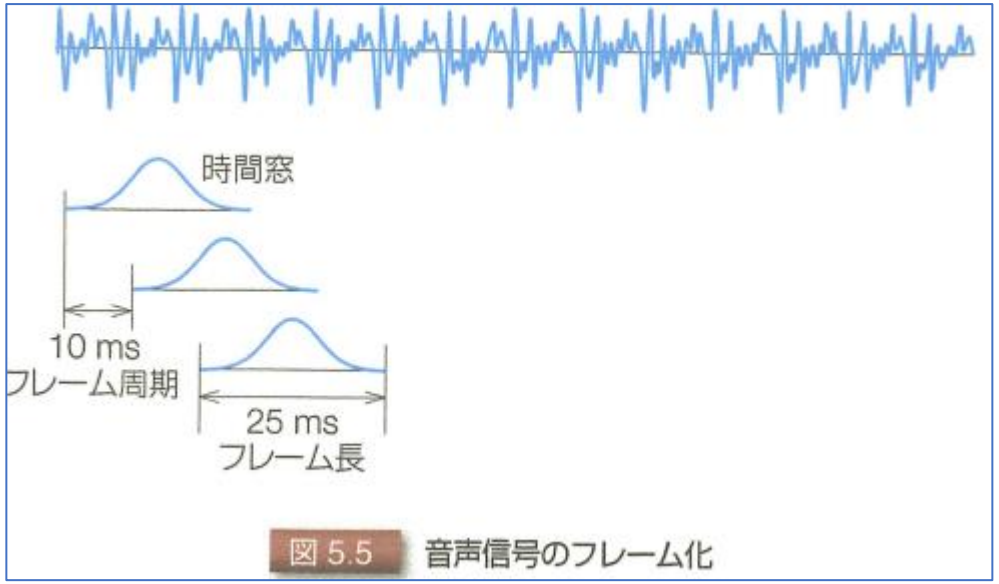


図 5.5 音声信号のフレーム化

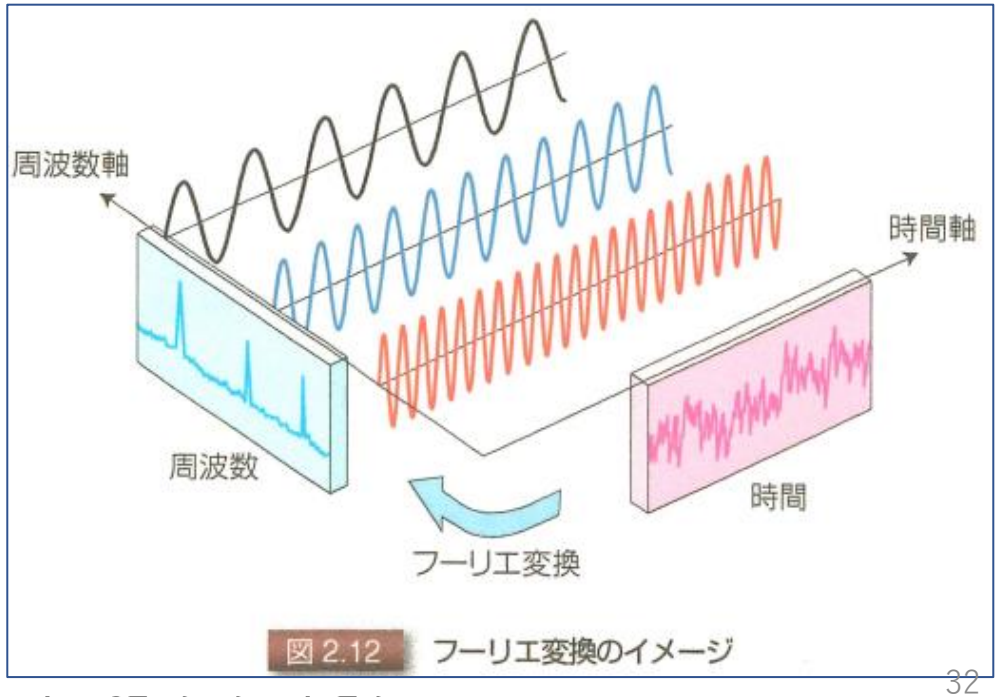


図 2.12 フーリエ変換のイメージ

教科書p65 図5.5, p24 図2.12

メルフィルタバンク (Mel filter bank)

- 図5.6:フーリエ変換によるパワースペクトル例(実部と虚部の√二乗和)
  - ➢実際に求めたいのは、人間が音声を聴取しているメカニズムに近い情報
  - ➢人間は、蝸牛内の基底膜の振動で周波数を感じているが、その振動を感知する有毛細胞は、特定の周波数帯域の音の大きさを感知している。

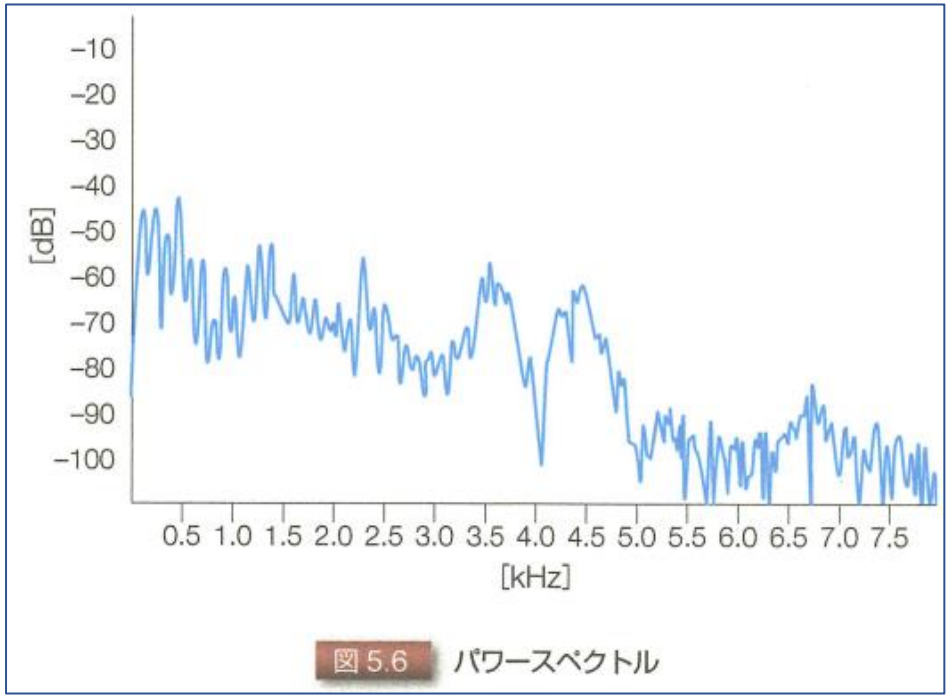


図 5.6 パワースペクトル

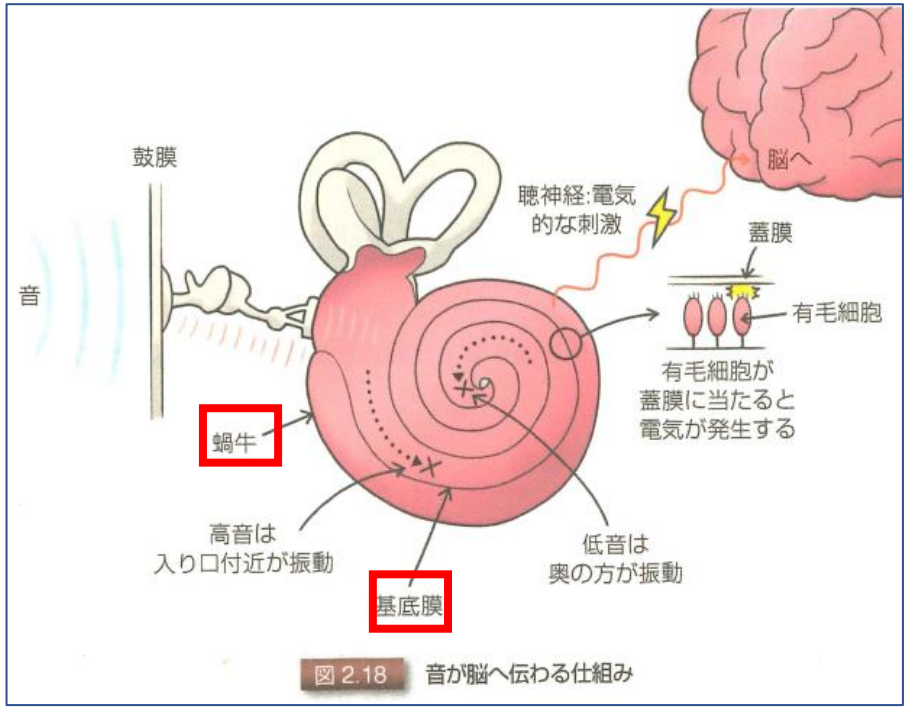


図 2.18 音が脳へ伝わる仕組み

メルフィルタバンク (Mel filter bank)(2)

- 周波数が低いほど、少しの音の違いで分かる
- 人間が感じる音の変化を一定にしたもの**→メル尺度**
- L個(通常L=24)の三角窓関数(図5.7)で積分したスペクトル**→メル帯域スペクトル**

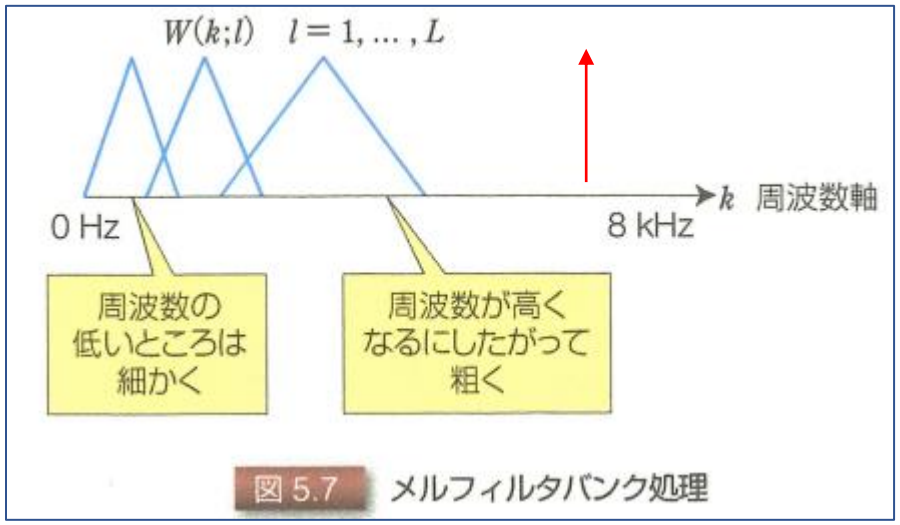


図 5.7 メルフィルタバンク処理

教科書p66 図5.6, p28 図2.18, p67 図5.7

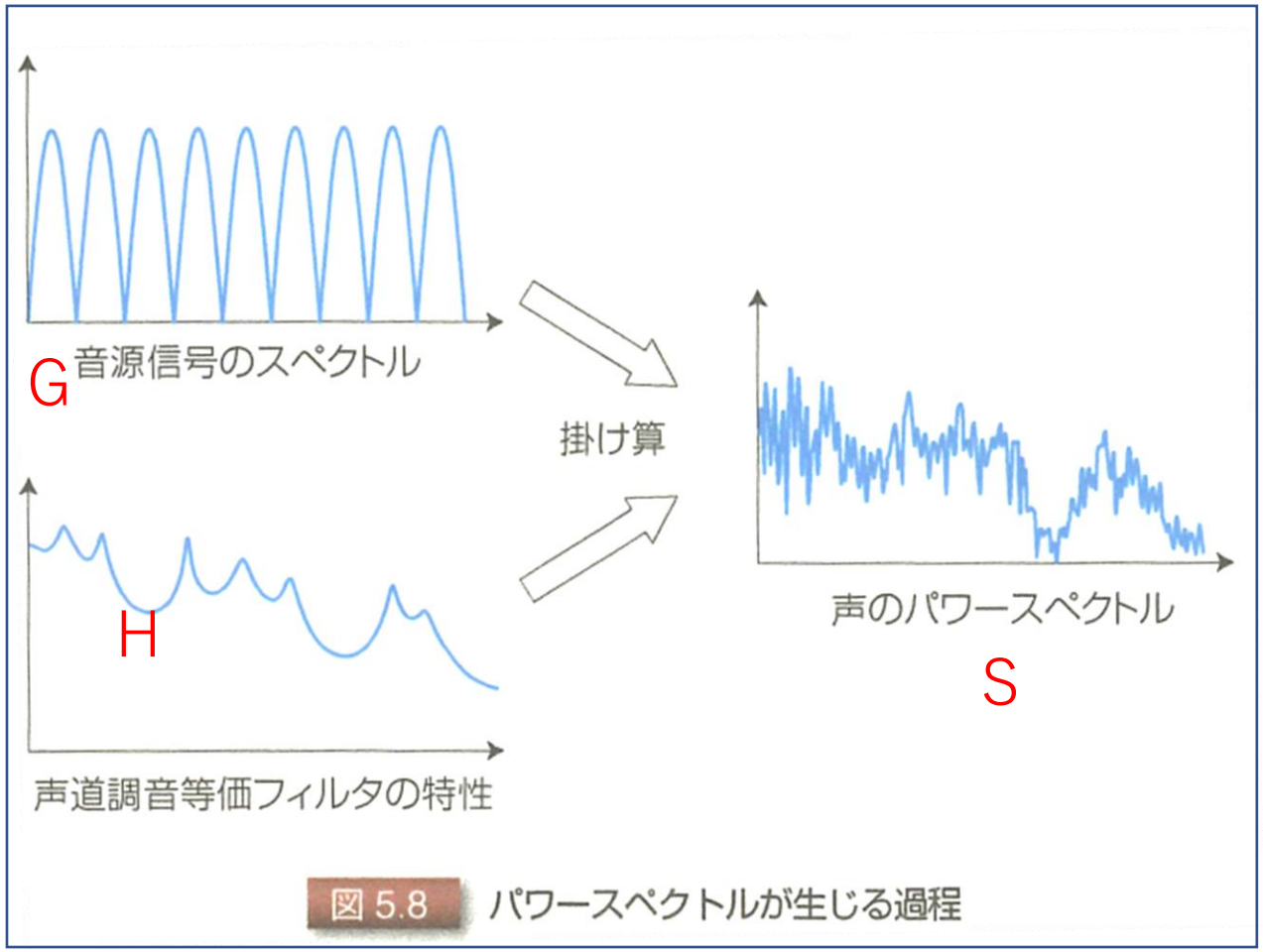
タバンクを計算します．ここで，$m(l)$ は l 個目のメルフィルタバンクの値，$W(k;l)$ は図 5.7 に示すような三角窓関数，$|S(k)|$ はパワースペクトルを示します．k は周波数です．

$$m(l) = \sum_k W(k;l) \cdot |S(k)| \quad (l = 1, \ldots, L) \qquad (5.3)$$

- 以上で、人間の聴覚に対応する方法で、音の周波数成分の情報を取得
- しかし、この周波数成分には、音源情報と、声道情報が混在

ケプストラム分析(Cepstrum Analysis)-1

- 声のパワースペクトルは、音源信号のスペクトル(G)と、声道調音透過フィルタ特性(H)の畳み込み
  - ➢パワースペクトルの微細なギザギザは音源信号(G)によって生じる
  - ➢大まかな形の変化は声道調音透過フィルタの伝達特性(H)によって生じる
- ほしいのはHで、SからHのみを取り出したい



教科書p68 図5.8

声帯と声紋

S

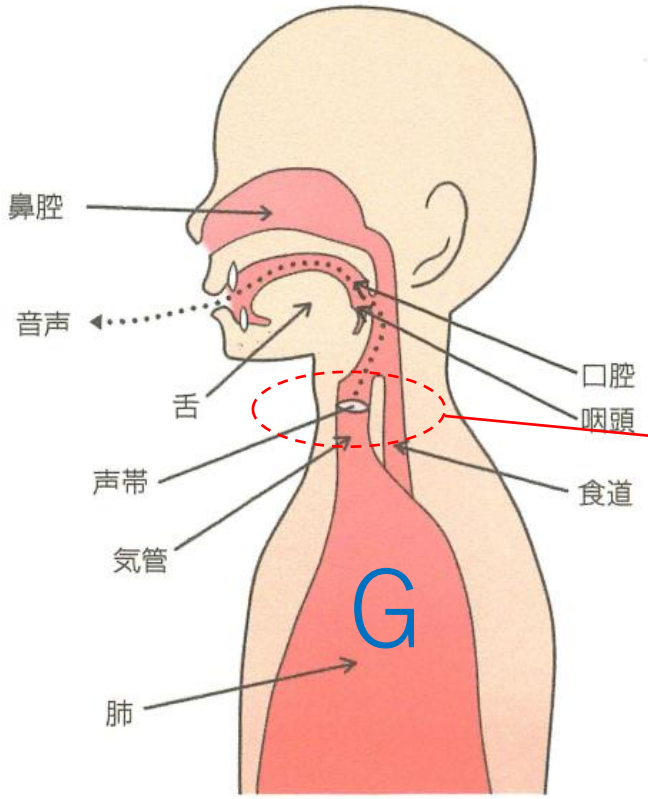


図 2.2 発声器官の構造

特性H

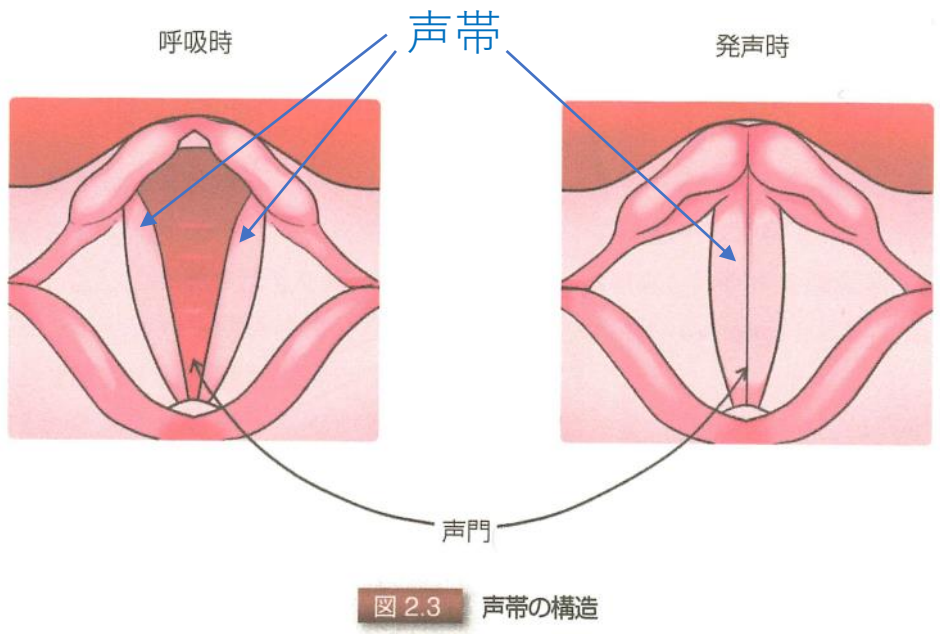


図 2.3 声帯の構造

教科書p15 図2.2, p16 図2.3

ケプストラム分析 - 2

- 声のパワースペクトルをフーリエ変換をするとGとHの畳み込み演算は掛け算になり、それらの対数を取ると、足し算になる
- メルフィルタバンクの対数を、離散コサイン変換→更にその周波数成分に分解→これで得られた情報を、**ケプストラム**（spectrumの最初の４文字を逆にしたもの）と呼ぶ

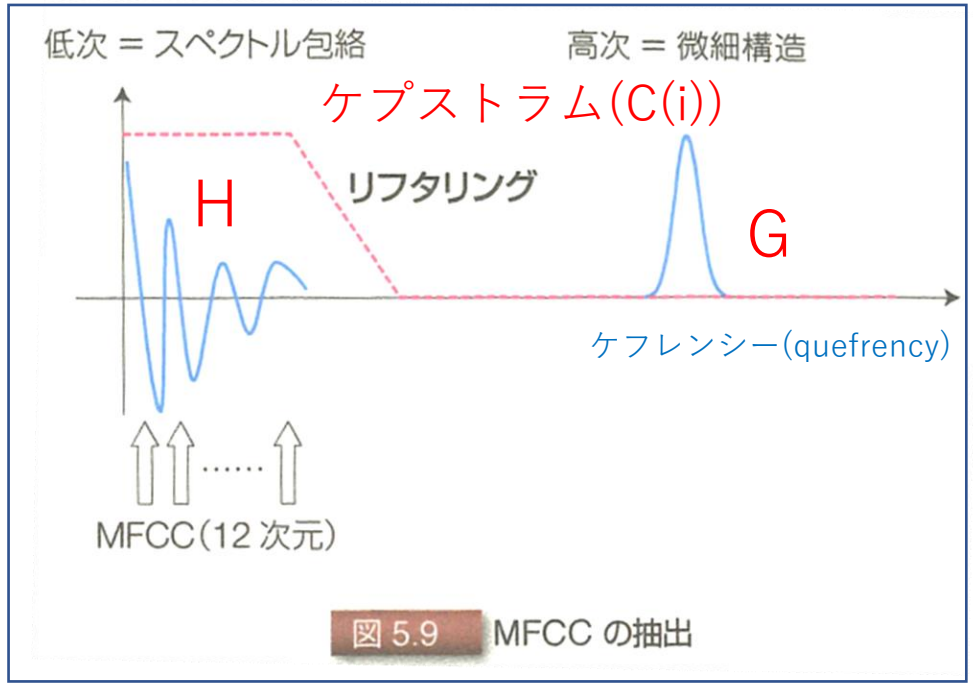
$$C(i) = \sqrt{\frac{2}{L}} \sum_{l=1}^{L} \log m(l) \cdot \cos\{(l - \frac{1}{2}) \frac{i\pi}{L}\} \qquad (5.4)$$

24

- 更に、高周波を削って、音声に関係する周波数を12個とる

→**MFCC** (Mel Frequency Cepstral Coefficient)

- MFCC自体周波数成分ではないが、これをそのまま特徴量として使っても特に問題ない



教科書p69 図5.9

ケプストラム分析 - 3

音声信号 $S = G * H$

↓ フーリエ変換

スペクトル $S = G \cdot H$

↓ 絶対値・メル帯域化・対数化

$\log|S| = \log|G| + \log|H|$

↓ 離散コサイン変換(DCT)

ケプストラム $\mathrm{DCT}\log|S| = \mathrm{DCT}\log|G| + \mathrm{DCT}\log|H|$

人間の感覚上の音の高さを反映（高周波ほど鈍感）

ケプストラム
※音のスペクトルを信号と見なしてフーリエ変換したもの

教科書p61 図5.2

図 5.2 ケプストラムを求める手順

ケプストラム分析による特徴量の抽出

- 母音：MFCCそのもの
- 子音：MFCCの変化分（=△MFCC)、その変化分（=△△MFCC)
- パワー： △パワーと△△パワー　　（パワー＝音声信号の強さ）

➔これら38次元(12x3+2)を、10msec毎に取得し、音声の特徴量とする

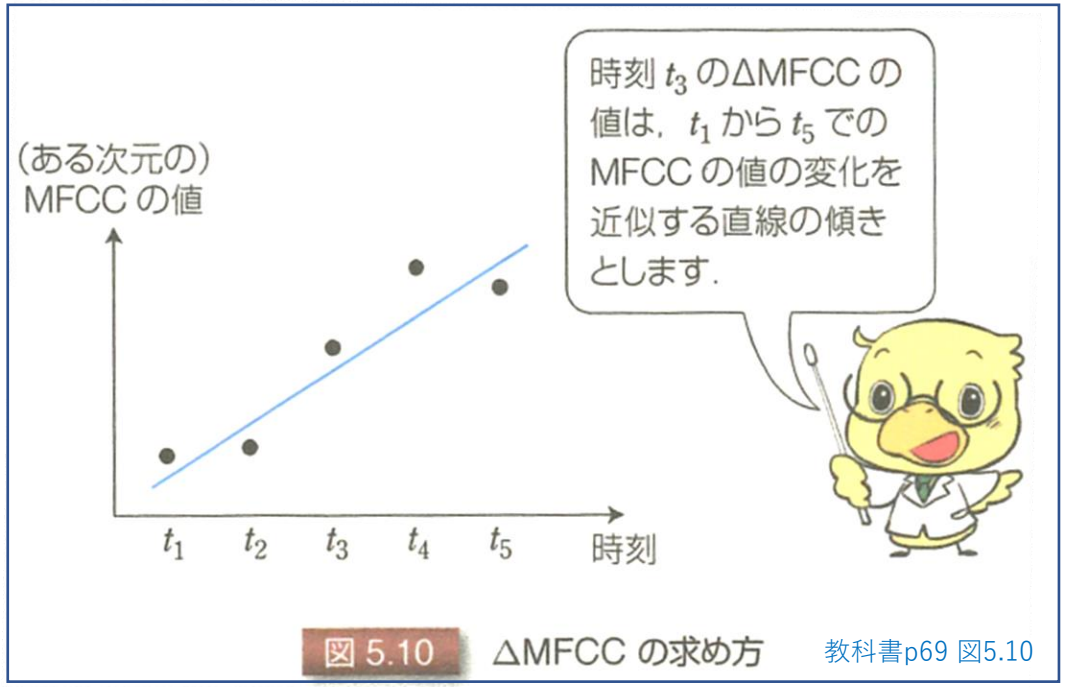
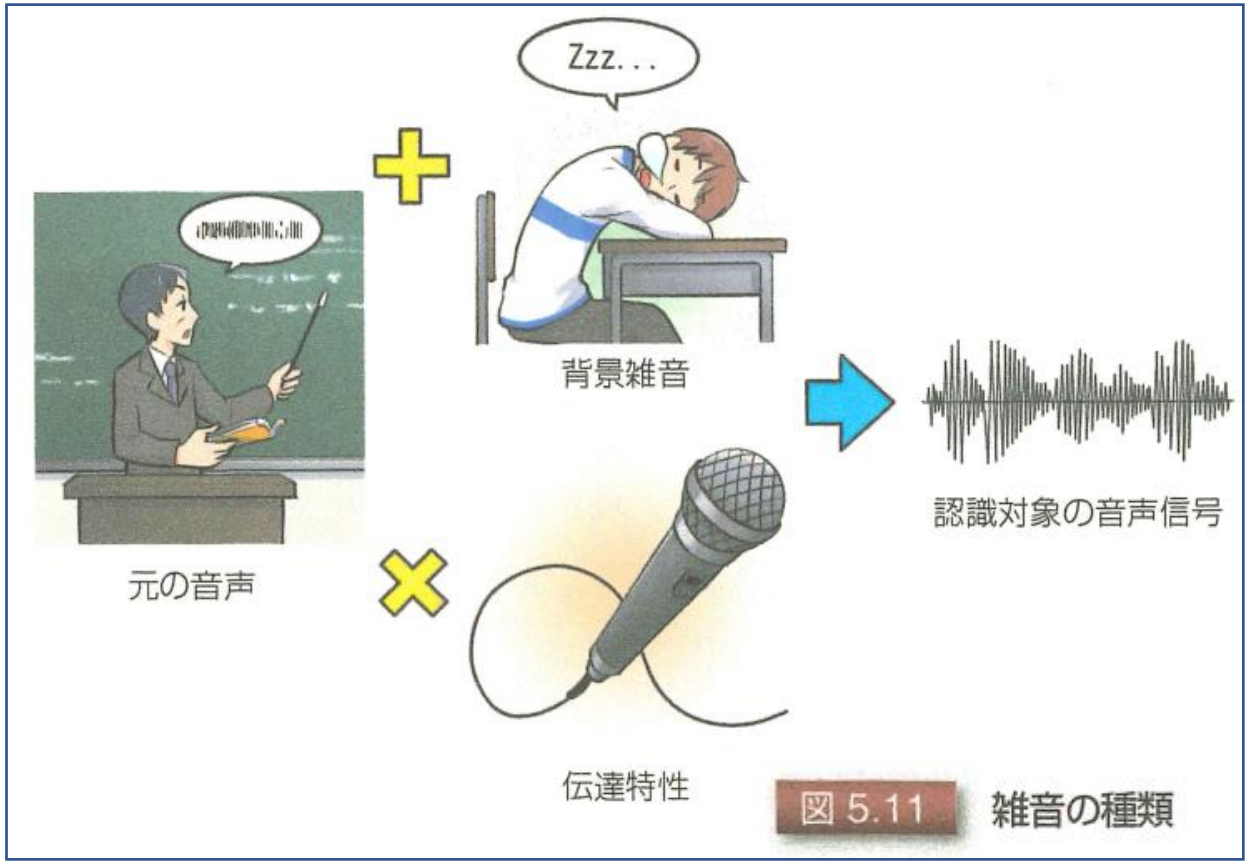


図 5.10　△MFCC の求め方　　教科書p69 図5.10

雑音の除去

- 音声信号には通常何らかの雑音が入っている
  - ➢パワースペクトルを求めた状態で雑音除去
- 加法性雑音：音声信号に重なる形
  - ➢例:背景雑音
  - ➢周波数空間での引き算で除去(spectral subtraction)
- 乗法性雑音：音声信号を歪ませる形
  - ➢例:マイクの伝達特性
  - ➢パワースペクトルの対数(掛け算が足し算に)引き算で除去



教科書p70 図5.11

- CMS(cepstrum mean subtraction, ケプストラム平均除去):発話全体のケプストラム平均を、各フレームのケプストラムから引く

対数による情報成分の除去

- 音源 S にノイズ Nが掛け算で乗っている場合、対数を取ることにより、Nの成分を除去

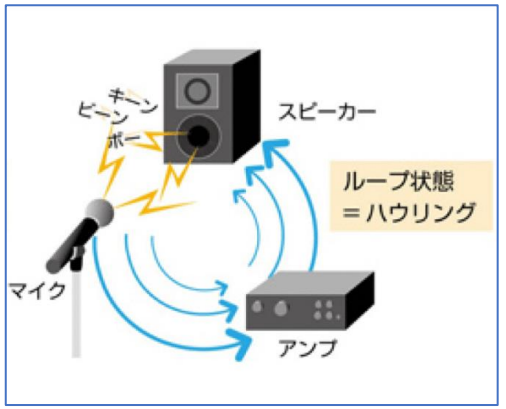
$$\log(S \times N) = \log(S) + \log(N)$$

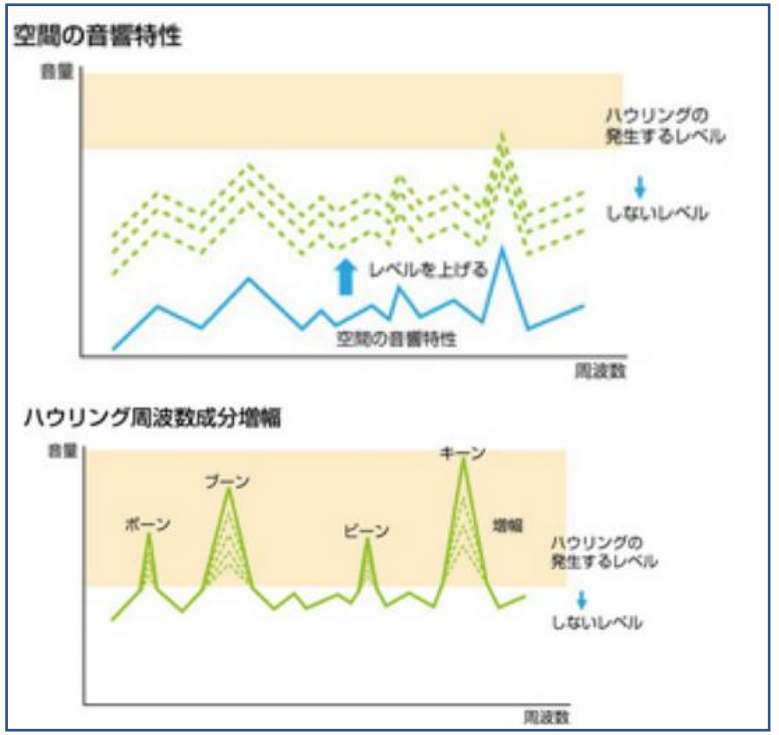
観測される音 ＝ SxN

この成分を除去する

(参考)エコー(echo)、ハウリング (howling)

- エコー：スピーカーから流れた音声をマイクが拾い相手のスピーカーから流れることで発生
- ハウリング：Zoomや会議室などで耳にする、「キ～ン」「ボー」という、耳をつんざくような音を発生する現象
  - ➢英語の原義：犬の遠吠え
  - ➢原因：マイクがスピーカーの音を拾って、その音をまたスピーカーから出力してしまう、という音のループ
  - ➢その空間によって、空間特有の音響特性がある
- 対策：
  1. マイクとスピーカーの距離を離す
  2. マイクを切る、スピーカーの音量を下げる
  3. イコライザーやFBS(Feedback Suppressor)を利用して､ハウリングを起こす周波数のカットを行う



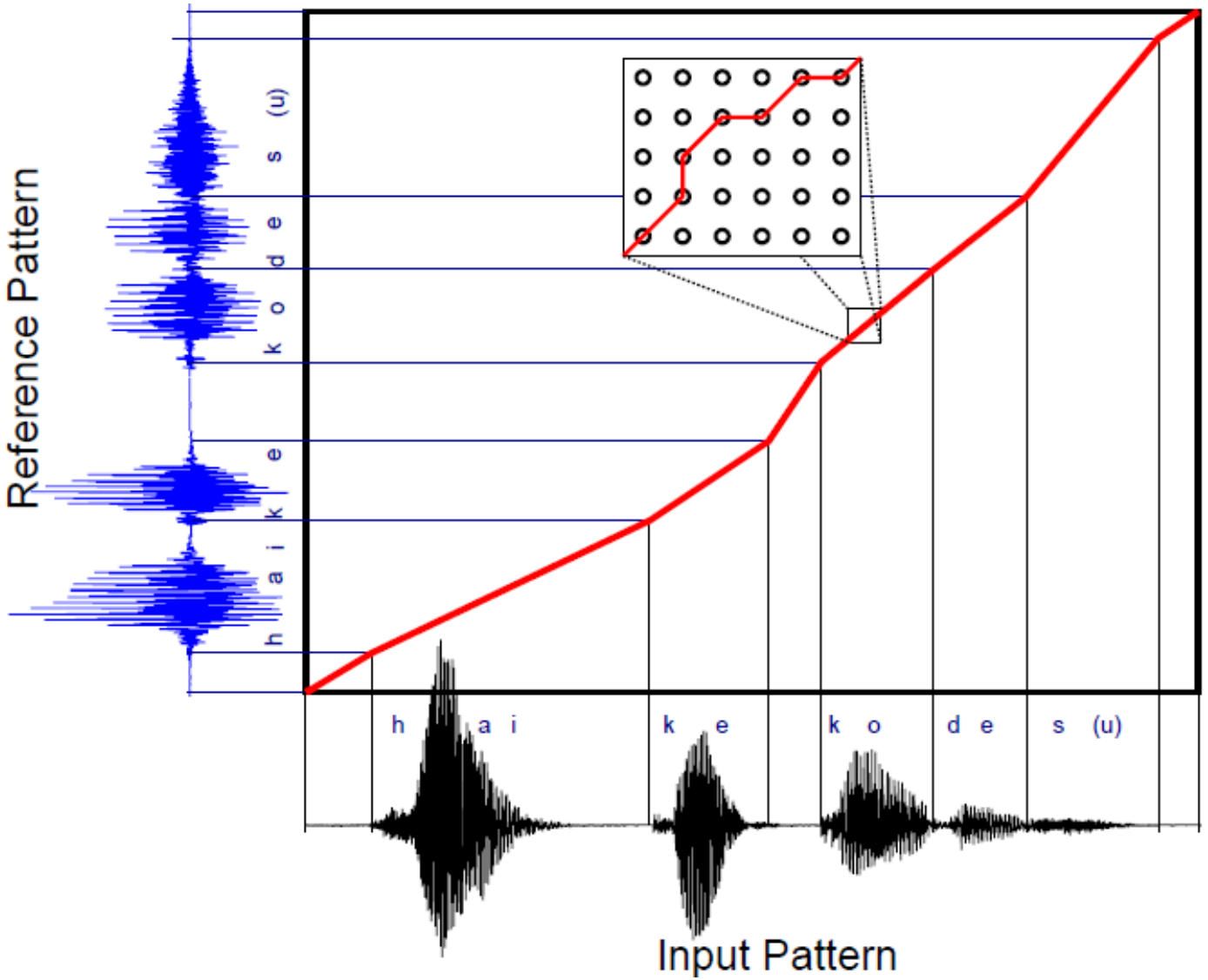


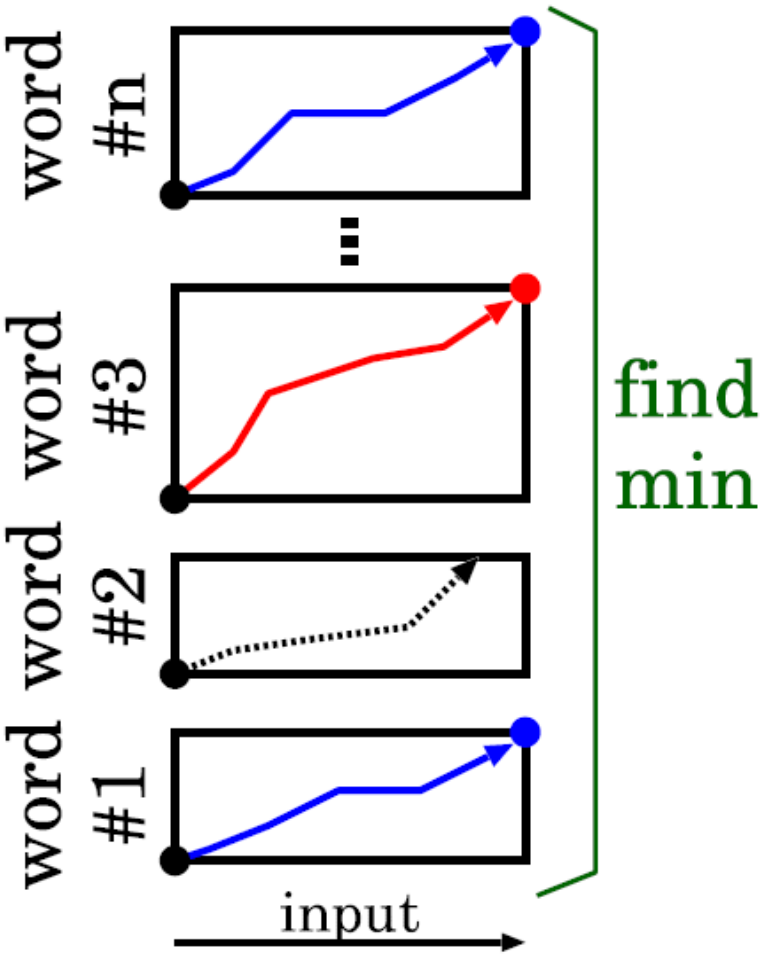
https://www.toa.co.jp/otokukan/otolabo/theme1/ss6.htm

DPマッチング

- 特徴量（ケプストラム）は得られ、そこから音声認識を行うが、その前に、1970～80年代に主に使われていた音声マッチングの方法を理解
- DP=dynamic programming
- ２つのパターン（ある人の音声のパターンと、標準的な音声のパターン）のマッチング
  - ➢例：音素（“あ”、“い”、子音、等）のパターンと、音声のマッチング
- 問題：音声は長さがわからないので、マッチングは簡単ではない
- 最近は、殆ど使われない
  - ➢統計的手法、Deep learning等が主流
- しかし、信号処理を理解する上では、理解は必須

音声におけるDPマッチング例





https://ocw.u-tokyo.ac.jp/lecture_files/engin_01/5/notes/ja/E1-DPmatching.pdf

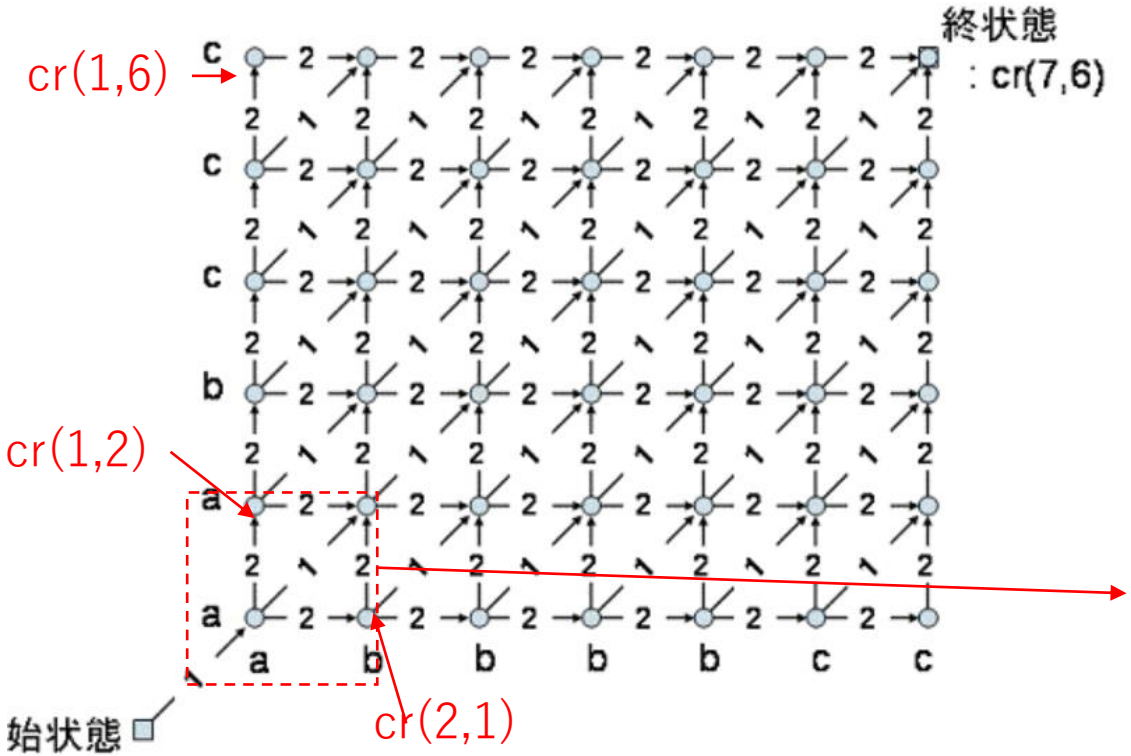
DPマッチング（前頁）

- “はい下戸です”と言っているらしいが、それを、各音素の標準パターン（y軸)と、実際の音声（x軸）でDPマッチングし、距離を求める

- それ以外の音声の候補（右のword#1～word#n）と音声の距離も同様に求め、一番距離の近いことばをしゃべっている、と判断する

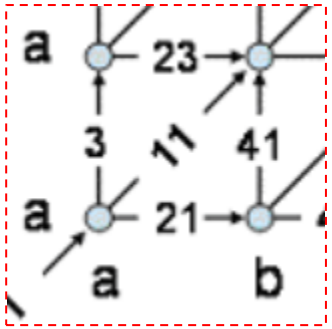
DPマッチング：例

1. P1 = {a,b,b,b,b,c,c}とP2 = {a,a,b,c,c,c}の距離を求める
2. x軸方向にP1, y軸方向にP2を並べ、上下に移動したら2点、斜めに移動したら1点とし、始状態から終状態まで進む
3. 移動した先のx軸とy軸の値が違う場合、ペナルティとして、移動の点数を10倍



4. 縦、横、斜めから来る中での**最終的な（終状態での）**最低点を採用する。

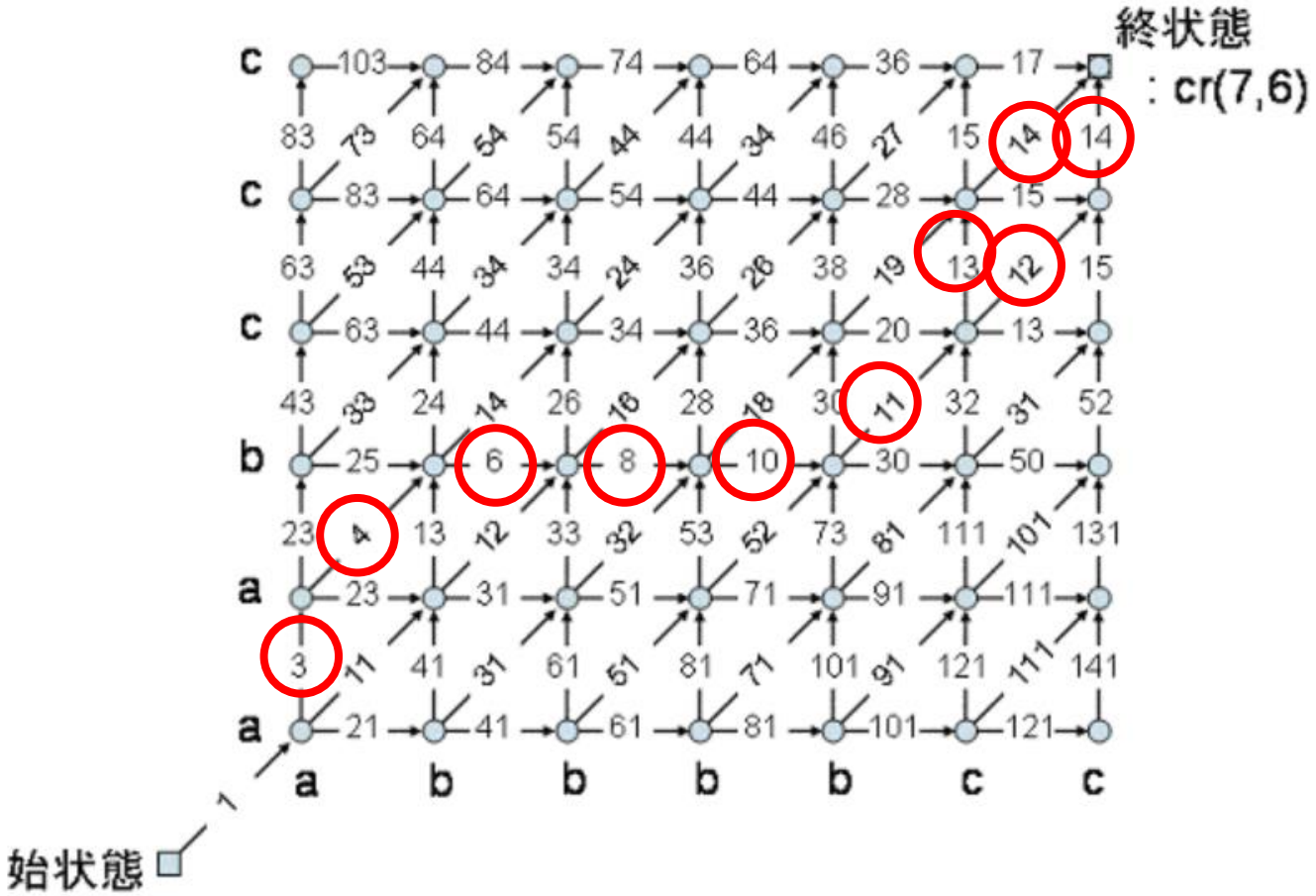
例：cr(2,2)では、斜め（cr(1,1)）から来る点数が、1+1*10=11で最低

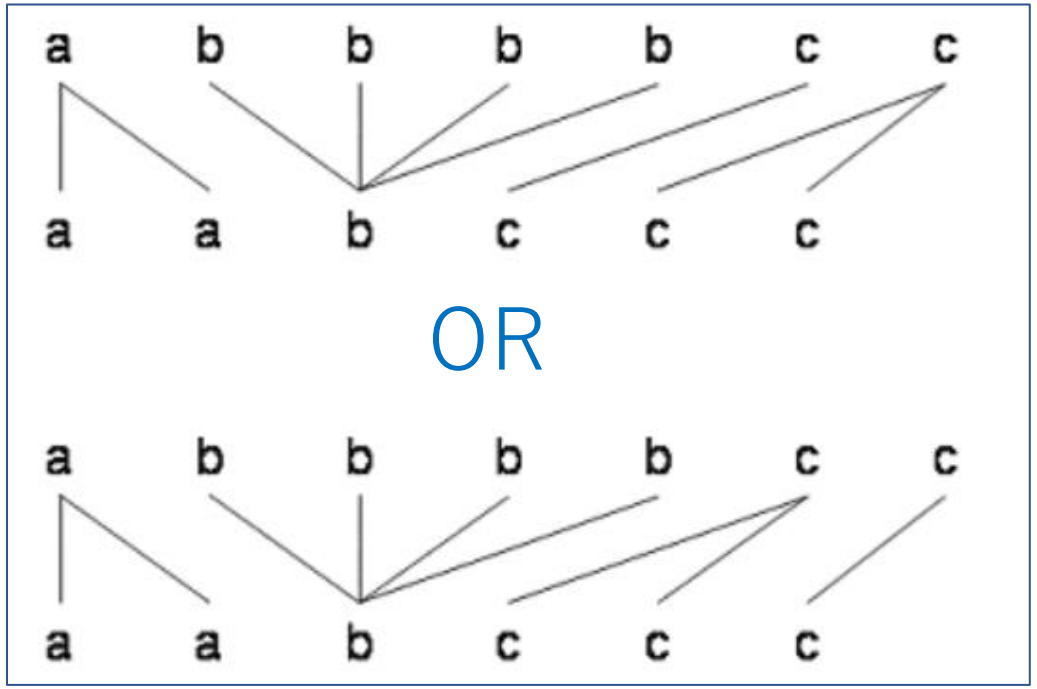


http://web.tuat.ac.jp/~tuatmcc/contents/monthly/200207/DP.html

DPマッチング：例－続き

5. これを繰り返し、終了状態までの経路の最低点になる経路を逆に辿る

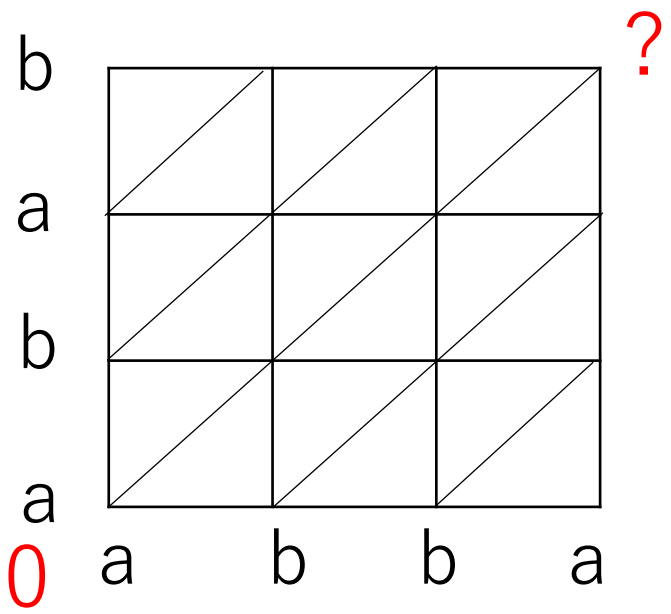




http://web.tuat.ac.jp/~tuatmcc/contents/monthly/200207/DP.html

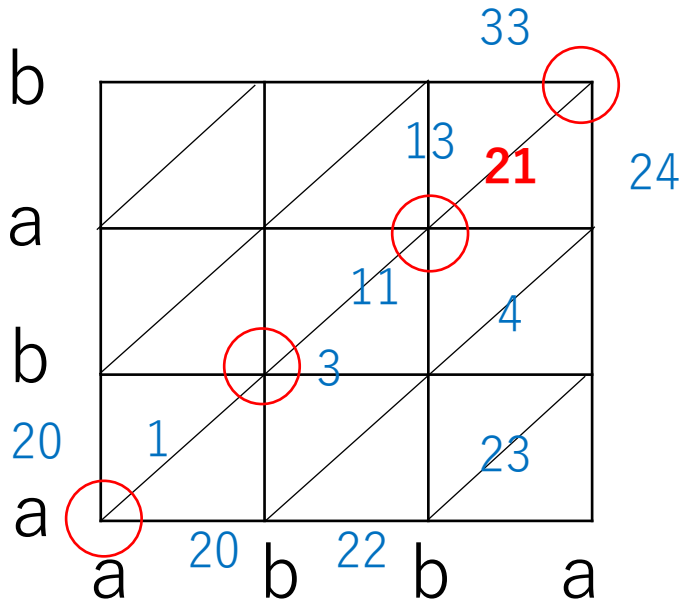
演習21-2 （LMS提出）

- 以下の{a,b,b,a}と{a,b,a,b}をDPマッチングした際の点数を求めよ
- (ただし、左下は0からstart)
- 計算方法は、前2頁の2.~5.と同じ
- (後回しで行う予定)



授業後
公開

演習21-2 答え



DPマッチングから隠れマルコフモデルへ

- DPマッチングの問題点
  - ➢比較的単純で、機械学習とは呼べない
  - ➢人間の音声には様々な揺らぎ（話者の違い、周辺雑音）の影響を受ける

- 隠れマルコフモデル（来週説明予定）
  - ➢大量のデータを用いて学習し、確率的に音声認識を行う
  - ➢1980年代から使われ始めた
  - ➢最近も使われているところもあるが、DNNに押され気味

音響モデル

特徴量から単語列の抽出

- ケプストラムで、音声の特徴量（X）は抽出できたので、音声の特徴ベクトル（X）より、尤もらしい単語列（W）を求める

$$\tilde{W} = \underset{W}{\operatorname{argmax}} P(W|X)$$

を計算したい

- しかし、XからWを求めるのは容易ではない
  - Xは膨大なvariation
  - それに比べ、Wは（依然として大量だが）、Xに比べれば限定的

argmax

$\underset{x}{\operatorname{argmax}}\, f(x)$：f(x)を最大にするx

（もちろん、変数は2つ以上でもOK）

例：$\underset{x}{\operatorname{argmax}}\,(-x^2+4x+5)$ ➔ $x = 2$（x=2の時、最大値9を取る）

演習21-3 (LMS提出)

$\underset{\theta}{\text{argmax}}\ \sin\theta \quad ,0 \leq \theta < \pi$ を求め、LMSに記述せよ

π はpi でいいです

授業後
公開

演習21-3 答え

$\pi/2$（最大値1を取る）

特徴量から単語列の抽出-続き

- 音声認識においては、ベイズの定理に基づき、以下の式を用いる
- 求めたいのは、一番可能性の高いW

X（音声特徴量）からW（単語列）の予測は大変

W（単語列）からX（音声特徴量）予測はまだ楽

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)}$$

$$\tilde{W} = \underset{W}{\operatorname{argmax}} \frac{P(X|W)P(W)}{P(X)} = \underset{W}{\operatorname{argmax}} \underbrace{P(X|W)}_{\text{音響モデル}} \underbrace{P(W)}_{\text{言語モデル}}$$

$P(X)$はWの最大値に無関係なので、除外

ベイズの定理
Bayesian principle

- W:雨が降る
- X:傘を持っている

- 雨が降る確率：P(W) = 0.3
- 傘を持っている確率 P(X) = 0.4　　（天気に関係なく）
- 雨が降っているときに傘を持っている確率P(X|W) = 0.9
- 傘を持っているときに雨が降っている確率P(W|X)は?

- 雨が降っている条件下で傘を持っている確率と、傘を持っている条件下で雨が降っている確率は、どちらも、雨がふっておりかつ傘を持っている確率P(W,X)と同じ

- $P(W,X) = P(X|W)P(W) = P(W|X)P(X)$ → $P(W|X) = \frac{P(X|W)P(W)}{P(X)} (=0.675)$

$P(W \cap X)$でも良い

宿題11：周波数成分の測定

- 各自、スマホ、またはPCに、周波数成分測定アプリをインストールし、何か声、音を入力し、インストールしたアプリ名、何の音or声を入力したか、その時の周波数の波形、をLMSに提出せよ。
  - 例：Spectroid、Sonic Visualiser、Room EQ Wizard(REW) 、iSmartESA(有料)
  - https://www.appbank.net/app-rank/life/interior/measurement/frequency-measurement/も参考になります

- **締切：B:7/1(土), A:7/3(月), どちらも9:00**

本日のまとめ

- 音声処理の全体の流れ
- 音声処理の歴史
- 前処理
  - デジタル化
  - 高域強調
  - 量子化
- オイラーの公式
- 特徴量抽出
  - スペクトル分析
  - ケプストラム分析
  - MFCC
- DPマッチング
- 音響モデル（前半）
  - 特徴量から単語列の抽出
  - ベイズの定理