

東京大学 工学部 4年生 夏学期 501130

応用音響学

導入：科目の紹介

嵯峨山 茂樹

東京大学 工学部 計数工学科

<sagayama@hil.t.u-tokyo.ac.jp>

資料所在 <http://hil.t.u-tokyo.ac.jp/~sagayama/applied-acoustics/>

応用音響学の狙い

■ 狙いと性格

- 「信号処理論 I, II」に続く「信号処理論 III」と捉えても良い
- 統計的信号処理、パターン処理、非定常時系列モデル
- 確率モデル、統計的学習

■ 対象領域

- 音声学と音声分析
- 音声符号化
- 音声認識
- 音声合成

■ 講義重点

- 基本アルゴリズムの理解
- 統計的信号処理の基礎

■ 前提知識

- 線形系理論とフーリエ解析
- 数理統計学 (分布、推定)

前提知識、成績評価、一般的注意

■ 前提とする知識

- 信号処理 (フーリエ解析、スペクトル推定、サンプリング、etc.)
- 数理統計学 (確率変数、分布、推定、etc.)

■ 成績評価

- 受講態度評価 (熱意、出席回数、着席位置、質問応答、私語)
- 期末試験の成績

■ 一般的注意

- 私語禁止
- 前方に着席
- 服装、態度

工学部4年生講義「応用音響学」で扱う内容 (予定)

■ 音声とは

- 日本語の音声学 (音韻論、英語と日本語、母音と子音、音節、韻律、日本人はなぜ英語が下手なのか?)

■ 短時間スペクトル分析

- サンプリング定理、量子化雑音
- 高域強調、波形窓
- 短時間自己相関関数
- 短時間スペクトル解析、ピッチ構造
- 短時間ケプストラム解析

■ 全極型モデル

- 線形モデル
- 線形予測分析 (LPC)
- 残差信号、ピッチ抽出
- 偏自己相関分析 (PARCOR)
- 線スペクトル対分析 (LSP)

■ スペクトル距離尺度

- 板倉・齋藤距離
- Euclid 距離、Mahalanobis 距離

■ クラスタリング解析

- k -means クラスタリング
- スカラー量子化とベクトル量子化
- 混合ガウス分布とEMアルゴリズム

■ 非線形時間伸縮

- 動的時間伸縮 (DTW)、DP マッチング
- One Pass DP アルゴリズム

■ 音声の確率モデル

- 多次元正規分布、多次元混合正規分布
- マルコフ過程
- 隠れマルコフモデル (HMM)
- Viterbi アルゴリズム

■ 確率モデルの学習

- Viterbi 学習 (時間軸クラスタリング)
- EM アルゴリズムとHMMの学習*
- 連結学習*

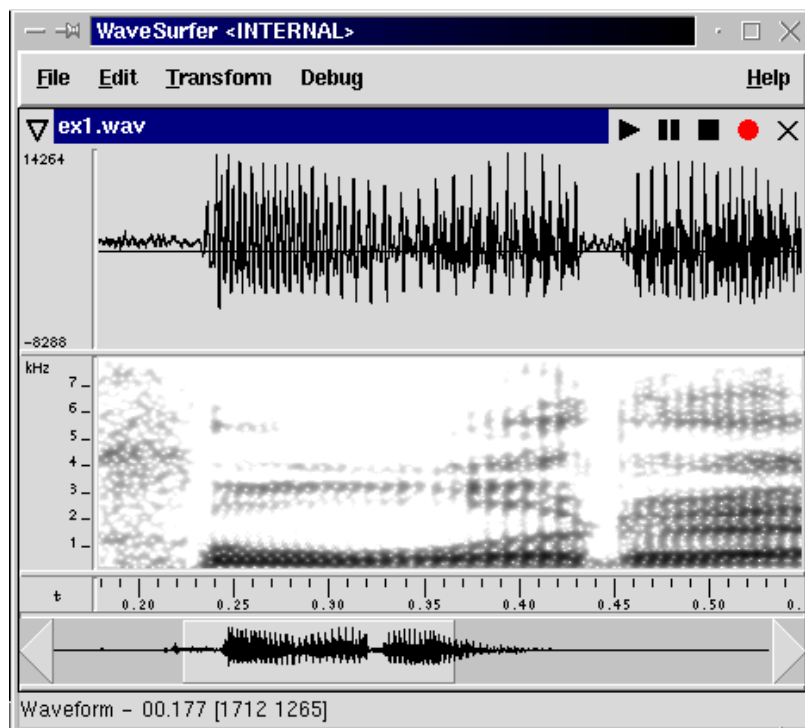
■ 音声認識システム

- 単語音声認識
- 大語彙音声認識、連続音声認識*

例1: 音声を解析する便利ツール

■ wavesurfer (KTH)

<http://www.speech.kth.se/wavesurfer/>

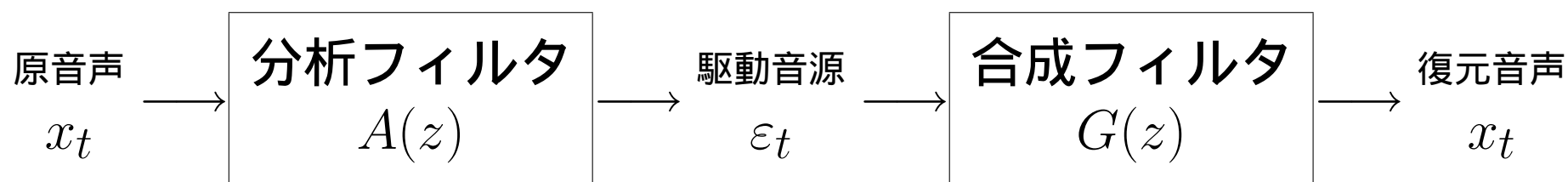


Wavesurfer の画面
Linux, Windows で動く。
(デモ)。

■ spwave (名古屋大学板倉研究室板野氏(現・和歌山大学)作成)

<http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/spLibs/spwave/index-j.html>

例2: LPC 音声分析合成モデル



- 分析フィルタ: 全零型分析フィルタ $A(z) = \sum_{i=0}^p a_i z^{-i}$

- インパルス応答: $\{1, a_1, a_2, a_3, \dots, a_p, 0, 0, \dots, \}$

- 合成フィルタ: 全極型合成フィルタ $G(z) = 1/A(z) = \frac{1}{\sum_{i=0}^p a_i z^{-i}}$

- 音声生成モデルのフィルタ

- 駆動音源信号 $\varepsilon(t)$

- 音源 $\varepsilon(t)$ を LPC 分析で得られたまま用いる — 音声の復元
 - 音源 $\varepsilon(t)$ を白色雑音(無声音)/パルス列(有声音) — 音声分析合成
 - 音源 $\varepsilon(t)$ を複数のパルス列で近似 — マルチパルス符号化
 - 音源 $\varepsilon(t)$ の残差波形コードブックを作る — CELP (現代の主流)
 - 音源 $\varepsilon(t)$ を他の信号で置き換える — 音声加工 (例1, 例2)

例3: ベクトル量子化 – 高効率符号化の基礎理論

音声(以外も)の情報圧縮では、ベクトル量を符号に置き換える技術が必要。そのような符号化アルゴリズムは「ベクトル量子化(vector quantization)」と呼ばれ、クラスタリングアルゴリズムに基づく学習アルゴリズムを利用する。

- k -means clustering アルゴリズム
- LBG アルゴリズム
- Segmental k -means アルゴリズム

例4: 音声認識 – 非線形時間伸縮マッチング

音声認識では、時間軸の変動とスペクトルの変動を如何に学習し、それをモデル化してマッチングするかが問題。代表的な2方法は：

- DP(動的計画法) マッチング アルゴリズム
- HMM(隠れマルコフモデル) とは何か

キーワード：必ず理解すべき事項一覧 1/3

■ A: 音声とは

- 音声学と音韻論の違い、母音(3要素、日本語母音)、子音(3要素、日本語の子音)
- Phonotactic Constraints、音節、モーラ、日本語の音節構造(促音、撥音、長音)、開音節
- 連濁、無声化、音素変形の環境依存性
- 日本語のアクセント、 m モーラ n 型とは、アクセント結合、アクセント句
- 日本人はなぜ英語が下手なのか？(日本語と英語の音声学的相違)

■ B: 短時間スペクトル分析

- サンプリング定理、Nyquist 周波数
- 量子化雑音、そのパワー $\frac{q^2}{12}$
- 高域強調、波形窓
- 短時間自己相関関数、正定値 Toeplitz 行列、Wiener-Khinchine の定理、Herzfeldt の定理
- ピリオドグラム、短時間スペクトル解析、ピッチ構造、スペクトル包絡
- ケプストラム、短時間ケプストラム解析、ピッチ推定(ケプストラム法)

キーワード：必ず理解すべき事項一覧 2/3

■ C: 全極型モデル(LPC, PARCOR, LSP)

- 音声生成の線形系モデル、音声分析合成の原理
- 全極型モデル、巡回型デジタルフィルタの安定性、Schur-Cohnの条件
- 線形予測分析(LPC)、正規方程式(Yule-Walker方程式)
- 残差信号、ピッチ抽出(残差相関法)
- 偏自己相関分析(PARCOR)、PARCOR係数
- Levinson-Durbin アルゴリズム
- 格子型フィルタ、全帯域通過デジタルフィルタ、音響管モデル
- 線スペクトル対分析(LSP)、LSP周波数
- 音声スペクトルを重みとした直交多項式系

■ D: スペクトル距離尺度とクラスタリング解析

- 特徴ベクトル空間、Euclid 距離、Mahalanobis 距離、板倉・齋藤距離
- クラスタリング、 k -means アルゴリズム、(極小値への)収束性
- スカラー量子化とベクトル量子化、音声符号化・圧縮・伝送
- 正規分布、多次元正規分布、多次元混合正規分布
- 尤度関数、Kullback-Leibler ダイバージェンス
- 不完全データ問題、EM アルゴリズム

キーワード：必ず理解すべき事項一覧 3/3

■ E: 非線形時間伸縮整合

- 動的時間伸縮 (DTW)、DP マッチング、整合経路、整合窓、トレースバック
- 二段DP マッチング、Level Building アルゴリズム、One-Pass DP アルゴリズム
- セグメンタル k -means アルゴリズム

■ F: 音声の確率モデルと学習

- 確率的定常信号源、多次元混合正規分布
- マルコフ過程
- 隠れマルコフモデル (HMM)、Mealy 型 / Moore 型、trellis
- Viterbi アルゴリズム、Viterbi 経路、Viterbi 分割、Viterbi 学習 (時間軸のクラスタリング)
- Forward アルゴリズム、Baum-Welch アルゴリズム、EM アルゴリズム
- HMM の学習、連結学習

■ G: 音声認識システム

- 単語音声認識
- 言語モデル (ネットワーク文法、 n -gram 文法)
- 大語彙音声認識、連続音声認識

参考書リスト (未完)

特に教科書として推薦できるものに 印を付す。本講義に関連している内容項目を [] で示す。

■ 音声学関係

- 町田健編、猪塚元、猪塚恵美子「日本語音声学のしくみ」(シリーズ・日本語のしくみを探る 2), 研究社, 2003, ISBN4-327-38302-3. 2000 円.
- 窪園晴夫「日本語の音声」(現代言語学入門2), 岩波書店, 2001, ISBN4-00-006692-7 C3380, 3400 円.

■ 音声情報処理関係

- 守谷健弘「音声符号化」, 電子情報通信学会, 1998, ISBN4-88552-156-4, 3000 円.
- 安藤彰男「リアルタイム音声認識」, 電子情報通信学会, 2003, ISBN4-88552-195-5, 3600 円.
- 鹿野, 伊藤, 河原, 武田, 山本「音声認識システム」(情報処理学会 IT text), オーム社, 2001 ISBN4-274-13228-5, 3500 円.
- 鹿野, 中村, 伊勢「音声・音情報のデジタル信号処理」, 昭晃堂, 1997.
- 春日正男, 船田哲男, 林伸二, 武田一哉「音声情報処理」(映像情報メディア基幹技術シリーズ 1), コロナ社, 2001, ISBN4-339-01261-0, C3355, 3500 円.

■ 信号処理関係

- 添田喬、中溝高好、大松繁「信号処理の基礎と応用」(理工学基礎シリーズ), 日進出版, 1979, ISBN4-8173-00106-6 C0041 ¥3000.
- 金井浩「音・振動のスペクトル解析」(日本音響学会編 音響テクノロジーシリーズ 5), コロナ社, 1999, ISBN4-339-01105-3 C3355 ¥5000. [距離、 \approx 変換、DFT、LPC、PARCOR]
- 萩原将文「デジタル信号処理」(電子情報通信工学シリーズ), 森北出版, 2001, ISBN4-627-70131-4 C3355, 2000円.
- 今井聖「信号処理工学 -信号・システムの理論と処理技術-」, (映像情報メディア学会編 テレビジョン学会教科書シリーズ 8), コロナ社, 1993, ISBN4-339-01058 C3355, 2800円.
- 浜田望「よくわかる信号処理」, (セメスタ学習シリーズ), オーム社, 2000, ISBN4-274-12990-X C3055, 2400円.
- 金城繁徳、尾知博「例題で学ぶデジタル信号処理」, コロナ社, 1997, ISBN4-339-00678-5 C3055, 2400円.
- 辻井重男、久保田一「わかりやすいデジタル信号処理」, オーム社, 1993, ISBN4-274-12939-X C3055, 2500円.

■ 統計数理関係

- 渡辺美智子、山口和範 編著「EMアルゴリズムと不完全データの諸問題」多賀出版, 2000, ISBN4-8115-5701-8 C1033 ¥6600.

[参考] 大学院講義「信号処理特論」の内容 (ここから選択)

- 学部講義の復習
 - ケプストラム MFCC
 - HMM
 - 単語音声認識
 - n -gram 言語モデル
 - 大語彙音声認識、連続音声認識
- 基本アルゴリズム
 - 学習: EM-algorithm
 - 探索: A* 探索
- 音素モデル
 - GMM
 - Tying [Young?]
 - 環境依存音素モデル [K-F Lee]
- 言語モデル
 - ネットワーク文法
 - 文脈自由文法
 - n -gram 統計文法 [??]
 - Latent Semantic Analysis [Belegarda]
- 探索問題
 - N-best algorithm (Tree-Trellis) [Lee & Soong]
 - Stack Decoder
- チャネル適応
 - CMN [Acero & Stern]
- 話者適応
 - VFS [Sagayama]
 - MAP [Lee]
 - MLLR [Legetter]
 - EigenVoice [Kuhn]
- 雑音適応
 - SS [Noll]
 - Varga-Moore の方法 [Varga]
 - PMC/NOVO [Gales]
 - JA [Sagayama]
 - Array Microphone [??]
- 音響モデル学習規範
 - ML, MAP, MMI, MCE
- 音声合成、音声対話
 - 音声合成、HMM合成、韻律モデル
 - 擬人化対話エージェント
- 音声認識の今後の問題
 - 自由発話
 - 翻訳電話

計数工学科と音声研究

- 音声研究者には計数工学科出身者が多い。
- 物理情報の工学（計測、数理）
Cf. 現在の「情報工学」：離散情報、記号情報が主体
- 「つぶしが効く」計数出身者

なぜ音声研究？ — 嵯峨山の場合

■ 中学時代

- 英語: 魅惑の発音記号 [æ, ʌ, ɔ, θ, ð, ʃ, ə, ø, ε, ...], 国語: 口語文法

■ 高校時代

- 合唱: ラテン語 (r, r, j, m, R, B, ...)
- ミュージカル映画 “My Fair Lady” 中 Higgins 教授の 魅惑の発音記号

■ 大学時代

- (音声学)
- 学部: 文字認識
- 大学院: 音響計測 (M 系列変調相関法), 信号処理

■ 電電公社通研 (のちNTTヒューマンインタフェース研究所)

- 数理的な手法
- 活発な研究者 (齋藤收三, 橋本新一郎, 板倉文忠, 好田正紀, 佐藤大和, 古井貞熙, 中津良平, 北脇信彦, 鹿野清宏, 東倉洋一)

■ 自動翻訳電話研究所 (音声情報処理研究室長)

- 音声認識合成技術の最先端開拓

■ 「初めに言葉があった。言葉は神であった。」(聖書・ヨハネ伝)

- 実は、文字のない頃、「言葉」は「音声」であった。(ってことは ... ?)