東京大学 工学部 計数工学科/物理工学科

応用音響学 : テキスト音声合成

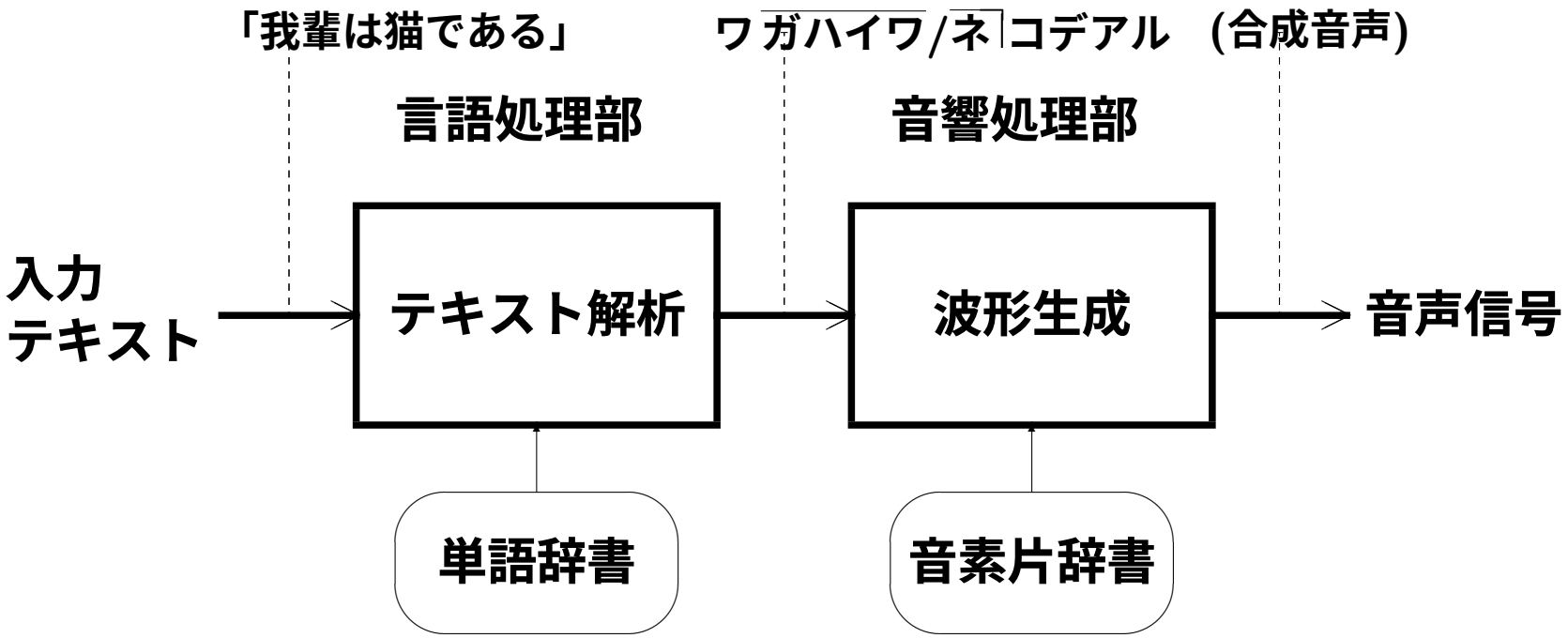
嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>
東京大学 工学部 計数工学科

資料所在 http://hil.t.u-tokyo.ac.jp/~sagayama/applied_acoustics/

音声合成の要素

- 合成単位
  - 録音編集合成型
  - 波形編集合成型
  - パラメータ編集合成型
- 韻律制御 (アクセント、アクセント結合、話調成分)
- テキスト音声合成 (テキスト解析)

テキスト音声変換のモデル

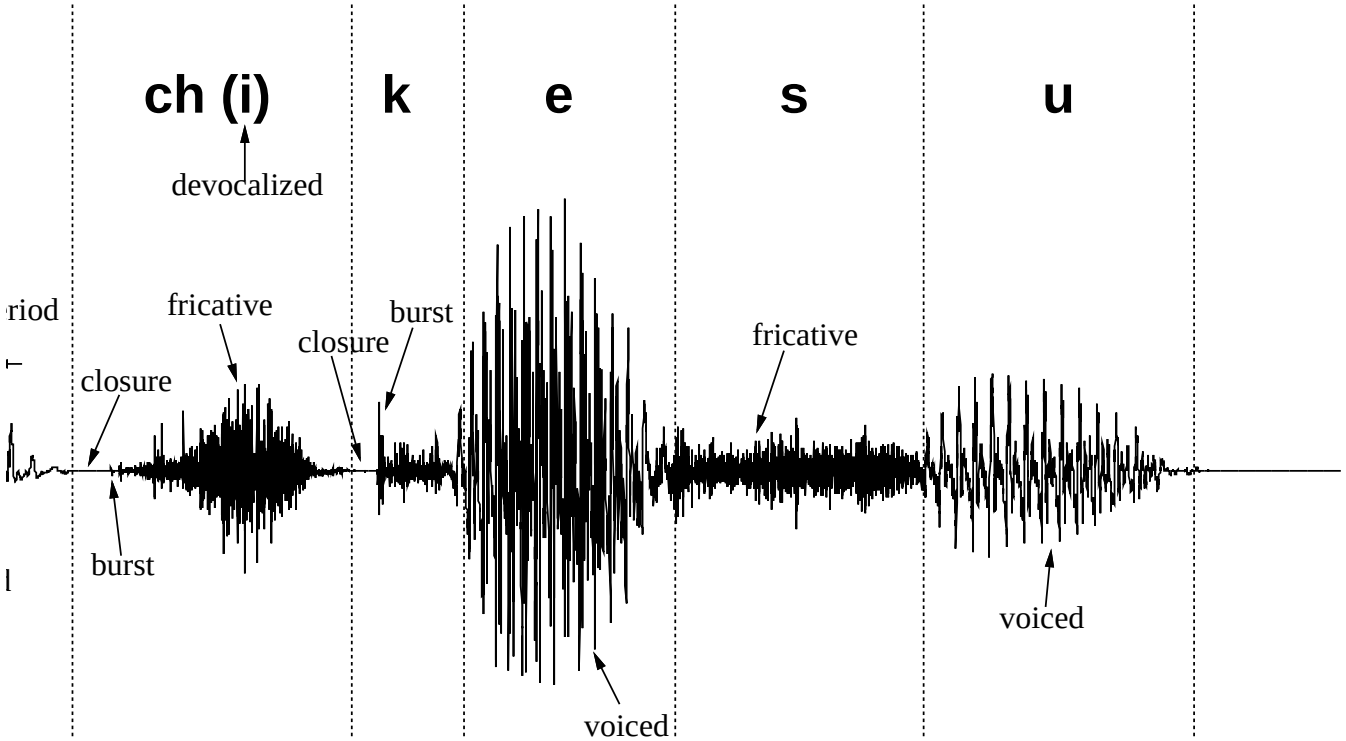


- 読み付与
- 呼気段落決定
- アクセント決定
- ポーズ付与

- ピッチパターン生成
- 素片選択，連結
- ピッチ変換
- 時間制御

音声波形

「打ち消す」の音声波形を示す。「ち」は無声化して、母音が脱落している。有声音 (ここでは u, e, u) には、周期性が見られる。これが、ピッチ (声の高さ) として感じられる。



Speech Waveform /uchikesu/

図 1. 音声波形の例「打ち消す」

嵯峨山茂樹 (東京大学大学院情報理工学系研究科システム情報学専攻)

音声合成単位と波形生成手法

- 合成単位
  - 長い単位: VCV単位(PARCOR-VCV 1978)，CV単位 (LSP-CV 1980)，CVC単位 (LSP-CVC 1984)
  - 環境依存音素パターンクラスタリング (LSP-COC方式 1986)
  - 可変長単位 (ATR ν-talk 1989?, CHATR 1995?)
- 素片表現
  - フォルマント
  - PARCOR, LSP パラメータ
  - 波形 (最小セット，巨大データベース)
- ピッチ制御
  - 合成フィルタのパルス音源周波数制御
  - 残差波形繰り返し周期制御
  - PSOLA (Pitch-Synchronous OverLap and Add)

可変長単位による音声合成方式の例： ATR νTalk

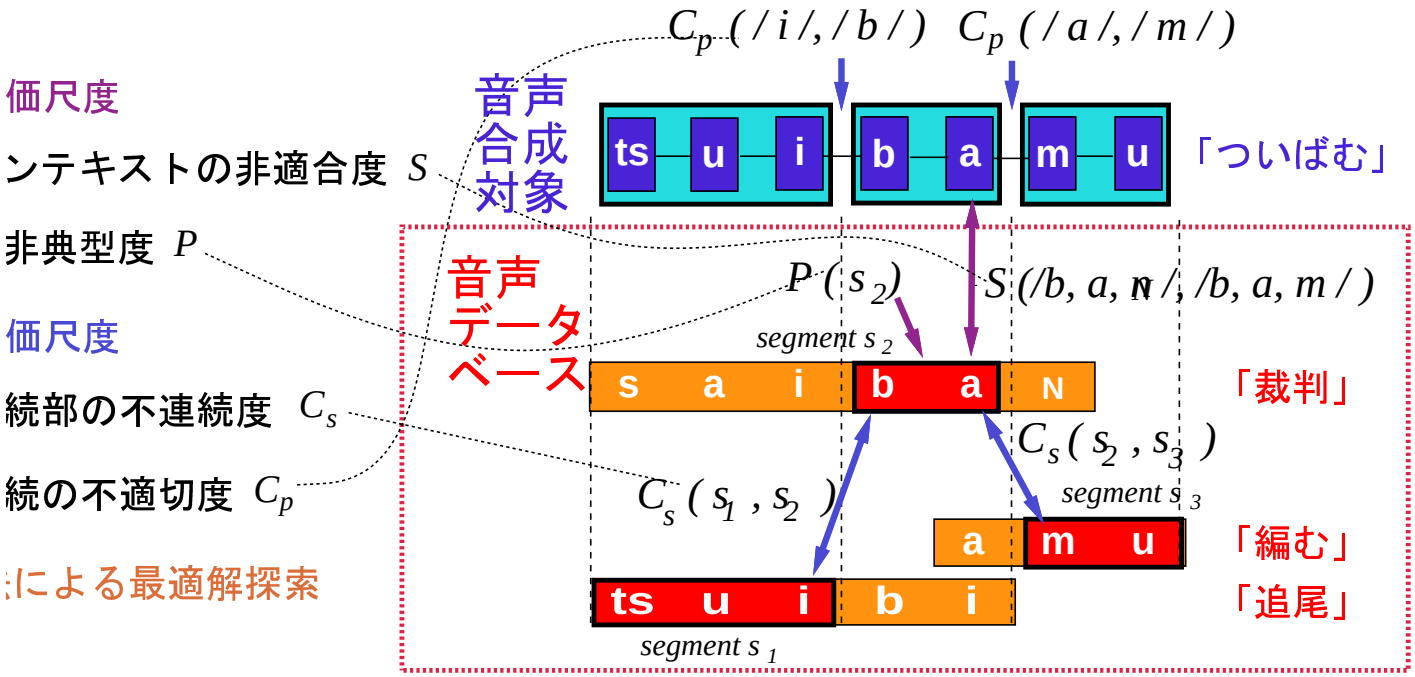
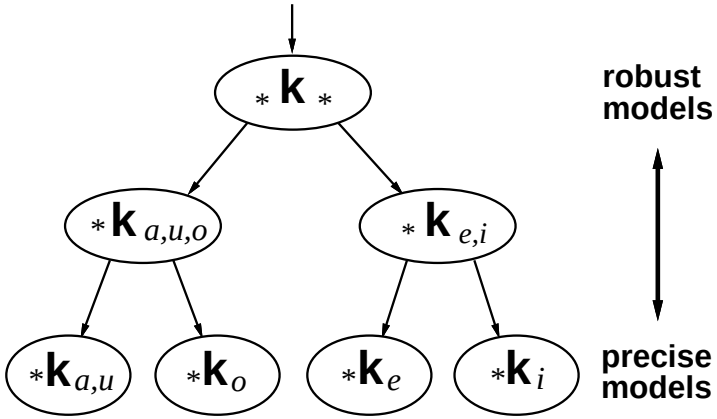


図2. 可変長単位による音声合成方式 — 4種類の歪み尺度を定義し、歪み最小化問題を動的計画法により解くことにより、音声データベースから最適な音声単位を選択する

音素環境クラスタリング (COC，PEC) (1987)

- 音素の環境依存性により自動分類



Hierarchy of Allophone Clusters
Utilized in Hierarchical Smoothing

図 3. 階層的に詳細なモデルとなる

- クラスタ分散が小さくなるように分割
- クラスタ平均パターン (セントロイド) を用いる
- 音声合成ボード「しゃべりん坊」(NTT data)

嵯峨山茂樹 (東京大学大学院情報理工学系研究科システム情報学専攻)

音素の環境依存性 – 異音 (allophone) の例

- 先行音素，後続音素など phoneme context の影響
  - 子音の例: 音素 /h/ は後続母音により別の allophone
    - は /ha/ [ha, xa], ひ /hi/ [hi, ɕi], ふ /hu/ [hɯ, ɸɯ], へ /he/ [hɛ, xɛ], ほ /ho/ [ho, xo]
  - /i/ に後続する子音は 「硬口蓋化」する．
    - き /ki/, ぴ /pi/ は 摩擦性
  - 「撥音」の例: 撥音 /ɴ/ は後続音素により別の調音
    - 後続音素が /a, i, u, e, o, k, h, j, w, g/ の場合 [ŋ]
    - 後続音素が /t, ʧ, ts, d, n, r, ʒ, dʒ/ の場合 [n]
    - 後続音素が /m, b, p/ の場合 [m]
    - 後続音素が /s, ʃ/ の場合 [ĩ]
- 「促音」の例: 撥音 /Q/ は後続音素が 1 モーラ長くなる．
  - 後続音素により別の調音
  - 後続音素が摩擦音 /s, ʃ, z, ʒ/ の場合，[s, ŋ, z, ʒ]
  - 後続音素が無声破裂音 /k, t, p, ʧ, ts/ の場合，無音
  - 後続音素が有声破裂音 /b, d, g/ の場合 (外来語)，buzz bar

日本語のアクセント

モーラ数とアクセント型

- 高低アクセント
  - ピッチアクセント: 日本語、セルボ・クロアチア語、リトアニア語、スェーデン語、ノルウェー語、古代ギリシャ語(?)
  - ストレスアクセント: 英語、ドイツ語、スペイン語、ロシア語、アフリカ東・北部
  - アクセントなし: フランス語、インドネシア語、韓国ソウル方言、日本水戸・仙台・熊本・宮崎方言
- m モーラ n 型アクセント (関東方言)
  - 2 モーラの例:
    端(0型) $\overline{\text{hashi}}$ ⇒ 端を(0型)
    箸(1型) $\text{ha}\overline{\text{shi}}$ ⇒ 箸を(1型)
    橋(2型) $\text{hashi}\overline{\ }$ ⇒ 橋を(2型)

嵯峨山茂樹 (東京大学大学院情報理工学系研究科システム情報学専攻)

日本語のアクセント型 (関東方言)

1. m モーラ n 型アクセント

	1モーラ	2モーラ	3モーラ	4モーラ	5モーラ
0型	柄 え	端 はし	形 かたち	情報 じょうほう	当り前 あたりまえ
1型	絵 え	箸 はし	修理 しゅうり	音声 おんせい	アクセント あくせんと
2型		橋 はし	落ちる おちる	明らか あきらか	お母さん おかあさん
3型			話 はなし	補う おぎなう	山桜 やまざくら
4型					帆掛け船 ほかけぶね
5型					お正月 おしょうがつ

外来語のアクセント型 (関東方言)

1. **外来語のアクセント: ③型が多い: m モーラ $(m-2)$ 型アクセント**
   例: ウィンブルドン(4)、シリーズ(2)、スポーツ(2)、カメレオン(3)、etc..
   例外: マニュアル(0,1)、アクセント(1)、etc.

日本語のアクセント

アクセント変化規則

- 大学院 /daigakuin/ ＋ 大学 /daigaku/
  → 大学院大学/daigakuin daigaku/
- 食べる→食べられる→食べられます→食べられません
- 食べる→食べるようだ→食べるようであります
- 「ニワニワニワトリガイル」イントネーション → 意味による
  「庭には鶏が…」ニワニワ/ニワトリガイル
  「庭には二羽鳥が…」ニワニワ/ニワ/トリガイル
  「二羽、庭には鳥が…」ニワ/ニワニワ/トリガイル

韻律制御

イントネーション

- 話調成分
- ポーズ (呼気段落)
- 強調
- 状況
- 自由発話

ピッチ周波数情報の利用

- ピッチ抽出法 — Lag Window 法 (1978)
  - 音声スペクトルを，それをLag Window によりスペクトルを平滑化したもので割って，フーリエ変換する．精度が高い．
- 単語音声認識とピッチパターンの組合せ (1990 高橋)
  - かんびょう vs かんぴょう，びょういん vs びよういん
- 韻律の情報量 (1991 村上)
  - アクセント/ポーズ情報 ∼ かな1字分の情報
- 音素パターンのピッチ周波数依存性 (1991 Singer)
  - ピッチ周波数とスペクトルには相関
- ピッチ周波数パターンによるアクセント句境界推定 (1991 下平)
  - ピッチパターンのクラスタリング，最適セグメンテーション(One-Pass DP)

日本語テキストからの音声合成の難しさ — 読み

- 文字種類が多い
  - ひらがな (50) ＋ かたかな (50) ＋ 漢字 (3000)
  - ローマ字、アルファベット、数字、記号
- ベタ書き (語境界がわかりにくい)
  - 畜産物価格安定法 = 畜産物/価格/安定法 or 畜産/物価/格安/定法
- 漢字の読みが多様 (音、訓)
  - 行 (ギョウ、アン、コウ、いく、おこなう)： 「行った」?
- 同字異音語
  - 今日 (こんにち，きょう)，最中 (さいちゅう，もなか)，etc.
    (逆に、同音異義語も多いので、了解しにくい。)
- 連濁、数詞
  - 株式会社 (かぶしきがいしゃ)
    いっぽん、にほん、さんぼん
- 文字から読みが決まらない
  - こうし (講師、子牛、格子)、えいり (絵入り、営利)、

嵯峨山茂樹 (東京大学大学院情報理工学系研究科システム情報学専攻)

日本語テキストからの音声合成の難しさ — アクセント他

- 文字だけからアクセントが決まらない
  - (最近は平板化傾向: 泉、赤とんぼ、電車、インク、ドラム、ディスク、etc.)
    (Cf. 英語 sénse, sénsor, sénsitive, sensátion, sensátional, etc)
- 句のアクセントの規則が複雑
  - 音声＋合成＋技術＋研究＋会＋定例＋総会＋準備＋委員＋選出＋期間＋中は...
- 「ニワニワニワトリガイル」イントネーション → 意味による
  - 「庭には鶏が…」ニ ワニ ワ/ニ ワトリガイル
    「庭には二羽鳥が…」ニ ワニ ワ/ニ ワ/ト リガイル
    「二羽、庭には鳥が…」ニ ワ/ニ ワニ ワ/ト リガイル
- 語彙、構文、意味、文脈、状況、常識、世界; 文法、音声、音節構造

自動翻訳電話の概念

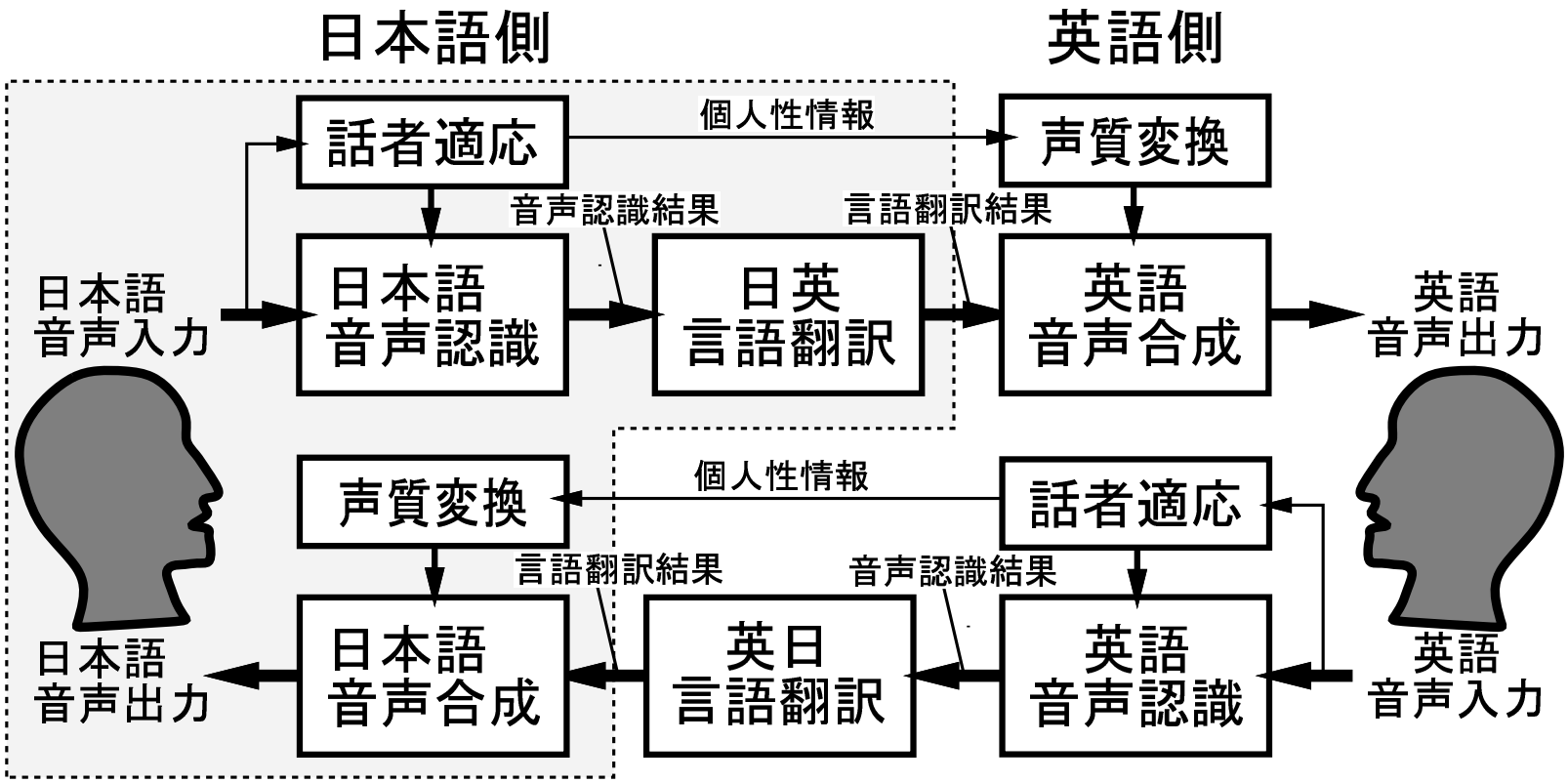


図 4. 自動翻訳システムの全体の概念図。日本語音声認識、日本語音声合成、言語翻訳、話者適応、声質変換、に重点を置いて研究した