韻律の記号化と
記号化された韻律の音声化

～音声合成の立場から～

峯松 信明

（東京大学大学院工学系研究科）

goo.gl/3JICag

テキストを読み上げる機械

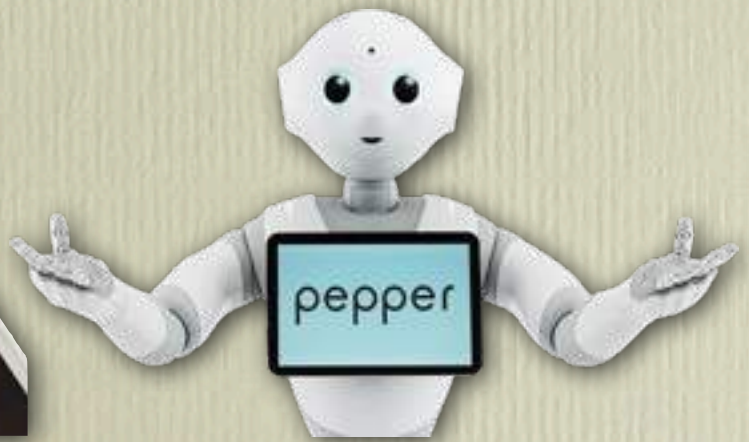
は，テキストを読み上げてない!?

- 多くの場合，テキスト → 韻律記号付きテキスト → 音声
  - **JEITA** IT-4006 日本語テキスト音声合成用記号（2000～）
    designed by 音声入出力方式標準化専門委員会
  - 韻律記号付きテキストの例（テキストからの自動韻律予測）

日本語の勉強は，とても難しいですが，アニメが大好きなので，これからも頑張ります。

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

- NHKアナウンサーで，この韻律拡張原稿を読んでる人がいる!?

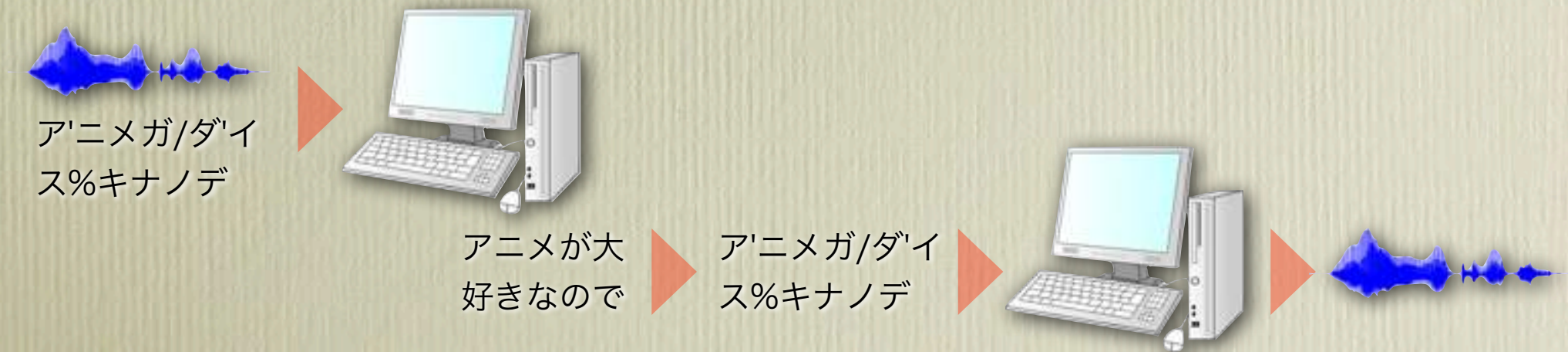


NEWS WEB EASY（イージー）

音声合成装置とJEITAラベリング

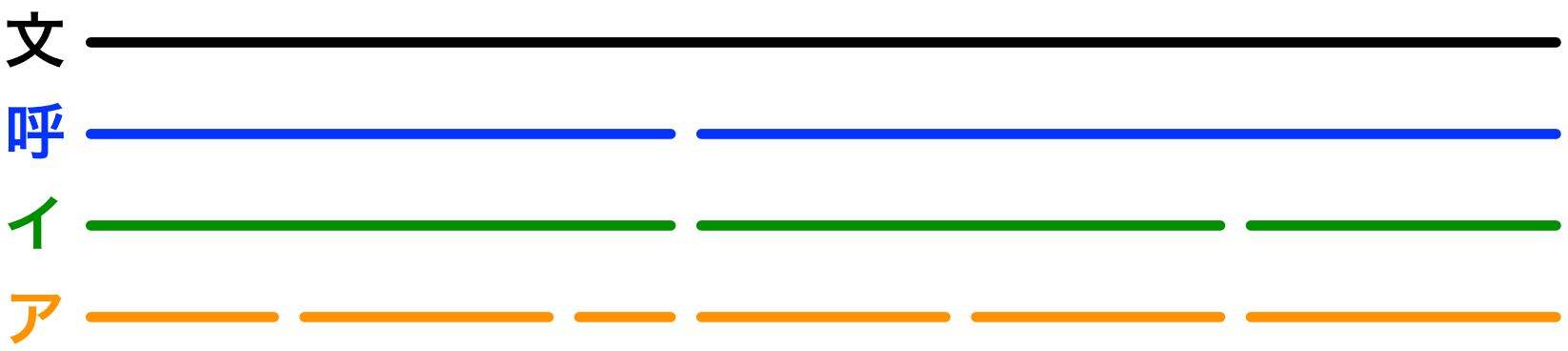
音声合成エンジンの作り方・動かし方

- 作り方
  - ある話者の読み上げ音声を千～数千文用意する（学習コーパス）。
  - 耳の良い人に韻律ラベリング（**JEITA**ラベリング）してもらう。
  - ラベル付けされたコーパスを作って「音のモデル」を作る。
- 動かし方
  - 任意のテキストを**JEITA**フォーマットに自動変換する（含韻律予測）
    - 自動変換器は，上記の**JEITA**ラベル付きコーパスを使って構築
  - **JEITA**表記された個々の音素・モーラに対して「音のモデル」を適切に選択し，**JEITA**テキストを音（音声波形）化する



JEITAラベリングの基本　高低？落ち？

韻律の階層構造を前提に文単位でラベリング



上位層の１韻律単位は，一つまたはそれ以上の下位韻律単位より構成。

呼気段落：明確なポーズで区切られるまとまり

イントネーション句：大局的・漸次的なピッチ下落パターンによるまとまり

アクセント句：高々一つのアクセント核で構成されるアクセント的まとまり

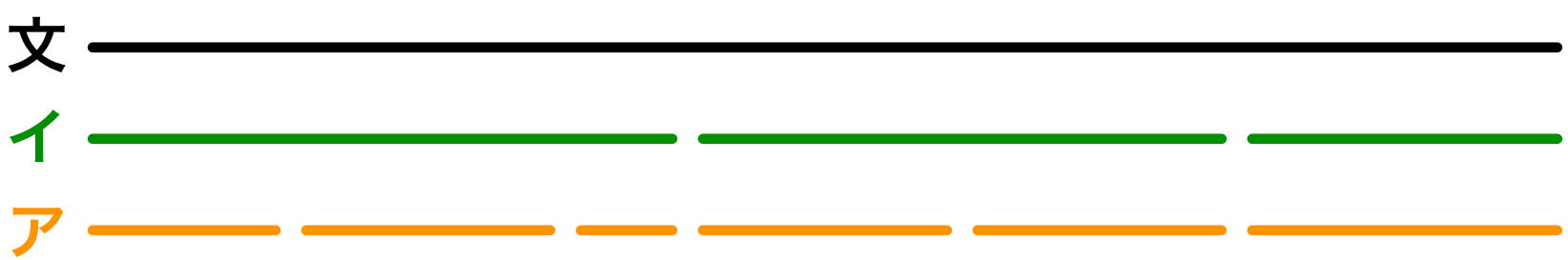
うまれて＋はじめて → うまれて｜はじめて　うまれてはじめて

わたしは＋たべる → わたしは｜たべる　わたしはたべる

以下では，簡単のため，イントネーション句境界には必ずポーズが入るものとする。即ち，呼気段落＝イントネーション句として話を進める。

JEITAラベリングの基本　高低？落ち？

韻律の階層構造を前提に文単位でラベリング



上位層の１韻律単位は，一つまたはそれ以上の下位韻律単位より構成。

イントネーション句：大局的なピッチ下落パターン＆ポーズによるまとまり
アクセント句：高々一つのアクセント核で構成されるアクセント的まとまり

日本語の勉強は，とても難しいですが，アニメが大好きなので，これからも頑張ります。

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

各モーラに対して，H/L は明示的に指定しない。

核が下落であることも明示的には教えない。

アクセント句冒頭２モーラの，ピッチの挙動も教えない。

上位層の１韻律単位は，一つまたはそれ以上の下位韻律単位より構成。

イントネーション句：大局的なピッチ下落パターン＆ポーズによるまとまり
アクセント句：高々一つのアクセント核で構成されるアクセント的まとまり

日本語の勉強は，とても難しいですが，アニメが大好きなので，これからも頑張ります。

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

各レベルの韻律現象に対して範囲を指定する ← 階層構造

JEITAフォーマットから音声へ

JEITAを元に文脈・コンテキスト情報で味付けをする

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

- 「ン」は「ン」だけど，どんな「ン」が欲しいのか？
  - 文中のイ句数は４，最初から４つ目のイ句にある
  - イ句中のア句数は２，最初から２つ目のア句にある
  - ア句中のモーラ数は６，最初から２つ目のモーラ
  - ア句中，アクセント核位置から前へ３つ目のモーラ
  - 直前音素は/a/，その前は/g/，直後音素は/b/，その後は/a/
  - 単語の品詞は動詞，アクセント句は動詞句・・・・・・

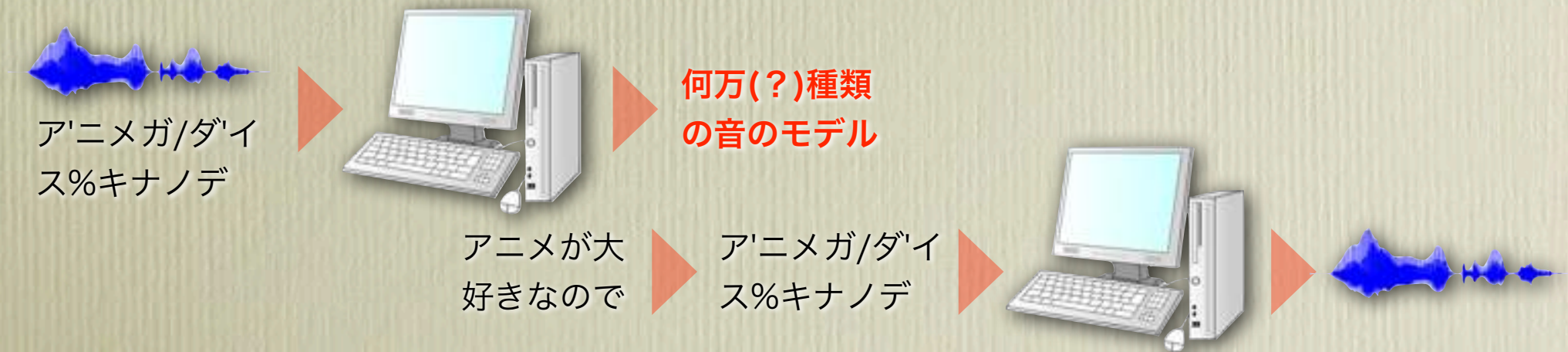
    → コンテキスト（文脈）情報を使って「ン」をより詳細に描写する

    → **コンテキストラベリング**（JEITAの韻律はこのための準備）
  - そういう「ン」を音として生成したい
    - スペクトル（音色）はどういうスペクトルが適しているのか？
    - ピッチはどの程度の高さにすべきなのか？　長さは？　強さは？

音声合成装置とJEITAラベリング

音声合成エンジンの作り方・動かし方

- 作り方
  - ある話者の読み上げ音声を千～数千文用意する（学習コーパス）。
  - 耳の良い人に韻律ラベリング（JEITAラベリング）してもらう。
  - **ラベル付けされたコーパスを作って「音のモデル」を作る。**
- 動かし方
  - 任意のテキストをJEITAフォーマットに自動変換する（含韻律予測）
    - 自動変換器は，上記のJEITAラベル付きコーパスを使って構築
  - JEITA表記された個々の音素・モーラに対して**「音のモデル」を適切に選択し**，JEITAテキストを音（音声波形）化する



核の前のン（かンら’ん しゃ），そこのピッチ

核の後のン（かんら’ン しゃ），そこのピッチ

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

- 「ン」は「ン」だけど，どんな「ン」が欲しいのか？
  - 文中のイ句数は４，最初から４つ目のイ句にある
  - イ句中のア句数は２，最初から２つ目のア句にある
  - ア句中のモーラ数は６，最初から２つ目のモーラ
  - ア句中，アクセント核位置から前へ３つ目のモーラ
  - 直前音素は/a/，その前は/g/，直後音素は/b/，その後は/a/
  - 単語の品詞は動詞，アクセント句は動詞句・・・・・・

  → コンテキスト（文脈）情報を使って「ン」をより詳細に描写する

  → **コンテキストラベリング**（JEITAの韻律はこのための準備）

アクセント句第１モーラ そこのピッチ

アクセント句第２モーラ そこのピッチ

韻律制御における話者の癖として・・

生まれて初めて　　生まれて三ヶ月

- 生まれて初めて → うまれてはじ’めて or うまれて/はじ’めて
- うまれてはじ’めて　　うまれてさんか’げつ
  - 「は」「さ」では全然下げない
- うまれて/はじ’めて　　うまれて/さんか’げつ
  - 「は」「さ」を明確に下げる人，うっすらと下げる人
  - はじ＝LH，さん＝LH，但し，どう実現するのか → 異音的変動あり

あなたは飲み込む

ねんばんがん
m　ng　n

- あな’たは/のみこ’む
  - 「み」で明確に上げる人，うっすらと上げる人
  - ア句の冒頭２モーラ ＋ イ句の１番目のア句 → 上がりは明確
  - ア句の冒頭２モーラ ＋ イ句の２番目のア句 → 上がりは明確or不明確
  - のみ＝LH，但し，どう実現するのか → 異音的変動あり

話者の違いもコンテキストの一部

韻律の記号化と記号化された韻律の音声化

JEITA韻律ラベリング for コンテキストラベリング

- 韻律階層構造を前提としたラベリング
  - 各ラベルの音響的意味は，明示的には機械に教えない。
- 1音素 with コンテキストラベリング → N種類の音として準備
  - 韻律ラベルはコンテキストラベルの中で使われる。

JEITA韻律ラベル vs. NHKラベル

- モーラ単位でのH/L指定はしない，という意味では類似
  - 但し，階層構造（範囲指定）やアクセント句などの差異もある
- コンテキストラベリングを通して音素の音的実体を多数用意
  - その結果，音声提供者の発話スタイル・癖がそのまま反映される。
    - うまれて｜はじめて or うまれてはじめて（テキストのJEITA化）
    - どのようなアクセント核の時に，どのくらい下げるのか？
    - どのようなアクセント句冒頭で，どのくらい上げるのか？
    - その話者のシミュレータを作るだけ。どちらが正しいかは問わない。

NHKアクセント辞典の大きな変化

峯松が抱いていたアクセント辞典のイメージ

孤立発声時のモーラ
単位ピッチレベル

sil■■■■■sil

■=H/L
sil=silence

文脈中でのモーラ
単位ピッチレベル

A□□□□□B
C□□□□□D
P□□□□□Q
X□□□□□Y
：　□=言及なし

それが，こう変わった（峯松の認識）。

文脈独立・非依存
のピッチ特徴

＊□□■■□＊

■=核(H)，■=L
□=言及なし

文脈中での
ピッチ特徴

sil□□■■□sil
A□□■■□B
C□□■■□D
P□□■■□Q
X□□■■□Y
：

- 様々な文脈で比較的安定して観測されるピッチ特徴を示す。
- 孤立発声時のモーラ単位でのピッチレベルすら教えてくれない。

NHK新アクセント辞典

峯松にとってのアクセント辞典

- 各単語を孤立単語として発声する場合に，各モーラのピッチレベル（H/L）を手短に示してくれる情報源

新アクセント辞典（核位置だけの呈示）

- 上記のようなユーザにとっては，ちと困った改訂
  - 「きんてつで'んしゃ」問題　　「ゆでた'まご」問題
- 「文脈非依存の不変な情報（＝核位置）を示す」とは言うが
  - さいたま＋だいがく → さいたまだいがく　（複合語表現）
  - あるく　あるかない　あるきます　（用言の活用）
  - おとこのこ　このおとこのこ　おとこのこむけのほん
- 「アクセント核の位置は文脈で変わる」ことを大前提として
  - つまり，複数の語がアクセント的にまとまることを前提として
  - テキスト読み上げ技術は構築されています。

いわゆる匂坂論文

音声合成技術者が一度は読むはずの論文 [匂坂+'83]

- アクセント句 ＝ 文節＋文節＋・・・
  - アクセント句 ＝ 高々一つの核で構成される，アクセント的まとまり
- 文節 ＝ 単語＋単語＋・・・
- 単語接続によるアクセント核位置変化（文節内で核位置決定）
  - 名詞＋名詞などの複合語表現（接頭語＋名詞なども含む）
  - 用言＋後続語などの，所謂，活用に伴うアクセント変形
- 文節接続によるアクセント核位置変化
  - わたしは ＋たべ'る → わたしは｜たべ'る　わたしはたべ'る
  - あな'たは＋たべ'る → あな'たは｜たべ'る　あな'たはたべる

論　文

UDC 534.782:809.56-148

日本語単語連鎖のアクセント規則

正員　匂坂　芳典†　　正員　佐藤　大和†

Accentuation Rules for Japanese Word Concatenation

Yoshinori SAGISAKA† *and* Hirokazu SATO†, *Regular Members*

(7) 秋永一枝：“共通語のアクセント”，日本放送協会編，日本語アクセント辞典，所載

最後に・・・

文脈独立・非依存のピッチ特徴

＊□□■■□＊

■=核(H)，■=L

□=言及なし

NHK日本語発音アクセント新辞典

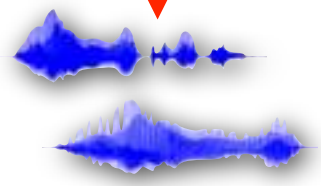
- 文として自然な音調で発音できるようにするための「手がかり」を提供する実践的なツール

OJADの韻律読み上げチュータ・スズキクン

日本語の勉強は，とても難しいですが，アニメが大好きなので，これからも頑張ります。

ニホンゴノ/ベンキョーワ_トテモムズカシ'ーデスガ_ア'ニメガ/ダ'イス%キナノデ_コレカラ'モ/ガンバリマ'ス%.

音声合成器



互換性をもっと大切に

MacBookは USB-C だけを採択

- 従来の USB 機器は接続できなくなりました。
- でも，ちゃんと変換アダプタを発売します。
  - これで従来の USB 機器も接続できます。
  - インタフェースを変えても互換性を担保する。

アクセント情報をどう示すか？

- モーラ単位のピッチレベルを表示するのか？
- 核位置だけを表示するのか？
- 一般読者にとっては，インタフェースの問題なのでは？
- であれば，互換性を保持してあげる必要があるのでは？

紙の辞典では×だが，電子版なら複数方式で呈示可能

- iOS 版などを作成する際には，是非互換性を大事にして欲しい。
  - どの呈示方式が良いのかは，ユーザに選ばせるのが良心的。

ご清聴ありがとうございました

goo.gl/3JlCag