

デジタル情報の処理と認識（'22）

第6回 音声の認識と合成

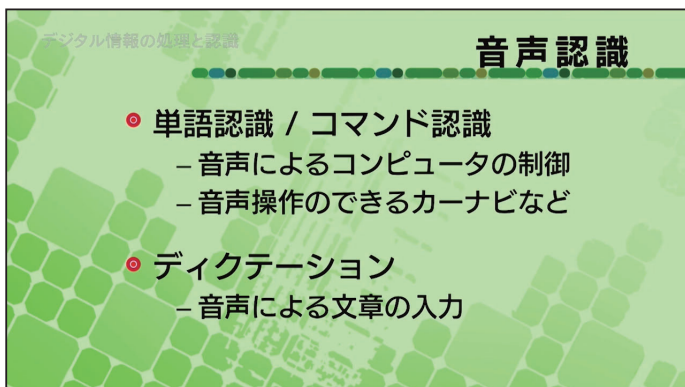
1



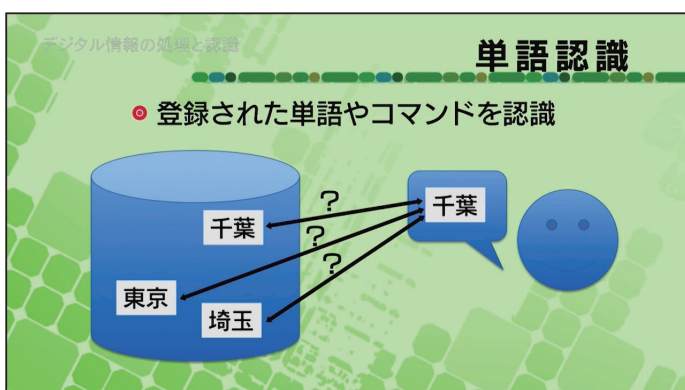
2



3



4



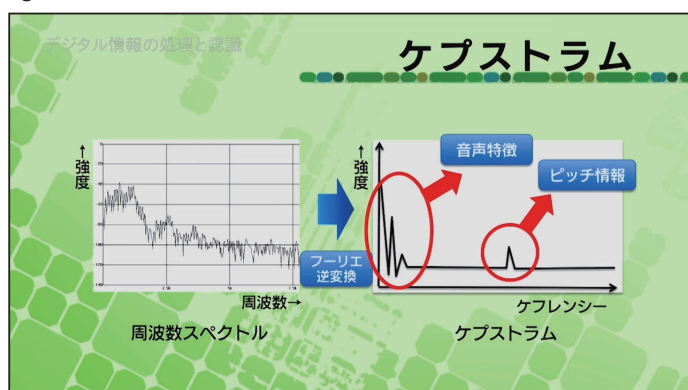
デジタル情報の処理と認識 ('22)

第6回 音声の認識と合成

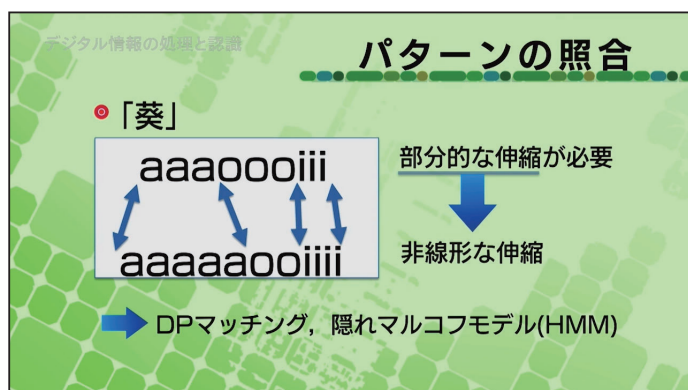
5



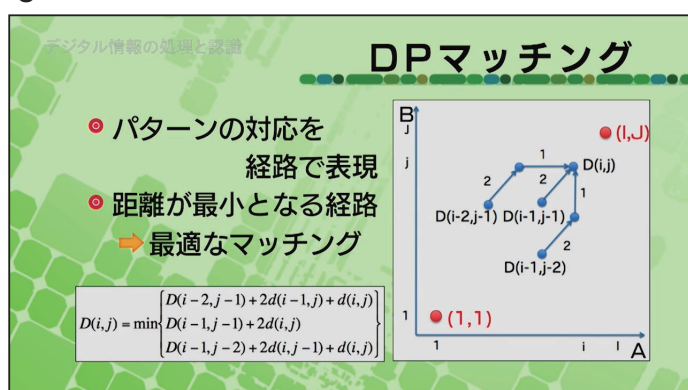
6



7

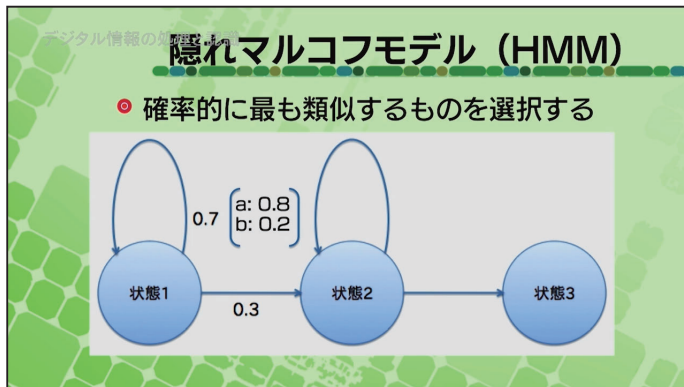


8



第6回 音声の認識と合成

9



10



11

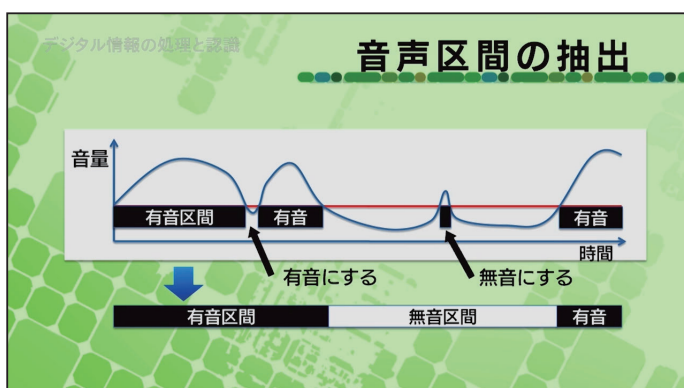
デジタル情報の処理と認識

ディクテーション

- 制限のない自由文を連続的に認識
- 話している区間はどこか?
- 認識の際の曖昧性が高い
 - 単語の制限がない
 - 単語間の区切りも必ずしも明確ではない

➡ 経験則や統計的制約などにより
認識結果の曖昧性を減らし 認識精度を向上

12



デジタル情報の処理と認識 ('22)

第6回 音声の認識と合成

13

デジタル情報の処理と認識

辞書による制限

- 辞書の中で最も類似している単語を探し
認識結果とする
ex. とおきょう → とうきょう
- 類似した発音の単語が多い場合には
誤認識が起こりやすい
- 辞書に登録されていない未知語を
どう検出するかが問題

5

14

デジタル情報の処理と認識

文法的な制約

- 2つの単語が接続するか否かによる制限
形容詞の後に名詞は続くが
助詞は接続しない等の制約を利用
- 文全体の文法的な制約による語彙の制限
⇒ 文全体として 文法的により自然な認識結果

10

15

15

デジタル情報の処理と認識

統計的な制約/話者適応

- N-gram
 - (N-1) 個の連続する単語の後に どのような
単語が続きやすいかによる語彙の制限
“I am are” × “I am a” ○
 - 2つの場合bigram 3つの場合trigram
- 話者適応
 - 特定話者 ↔ 不特定話者

20

25

16

デジタル情報の処理と認識

これは音声認識のレイです。このように音声用入力
することができます。コマンドの利用することがで
きます。

30

35

第6回 音声の認識と合成

17



18

デジタル情報の処理と認識

音声合成

- 時報案内や駅のアナウンスなどで実用化
- 「〇〇はどこ？」とスマートフォンに向かって話しかけると音声で案内
- 話した内容を翻訳して音声で読み上げ
- 歌声の合成

19

デジタル情報の処理と認識

録音編集方式

- あらかじめ単語や定型文などを録音し簡単な編集を行って音声を合成

時報案内

- ー 「10時」, 「15分」, 「です」等を録音
- ー 繋ぎ合わせて「10時15分です」を合成

- 比較的音質が良いが 決まった内容の音声以外は合成できない

20

デジタル情報の処理と認識

規則合成方式

- 基本的な音声単位の結合方法や韻律情報等の規則に基づき音声を合成
- テキストに基づき音声を合成する方式
 - ➡ テキスト音声合成 (Text-To-Speech)
- 合成できる音声の自由度が高いが合成音声が機械的で不自然な場合もある

第6回 音声の認識と合成

21

デジタル情報の処理と認識

合成の単位

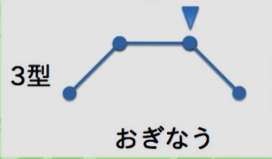
- CV (Cは子音 Vは母音)
「あ」と「か」 → 「あか」
- VCV (母音+子音+母音)
usa と agi → うさぎ(usagi)
- CVC (子音+母音+子音)
mik と kan → みかん (mikan)

22

デジタル情報の処理と認識

韻律情報の付加

- 日本語のアクセントは音の高低
(↔ 英語は強弱)
- モーラ：おおむね仮名1文字に相当
- 下がる1つ前が
アクセントの位置

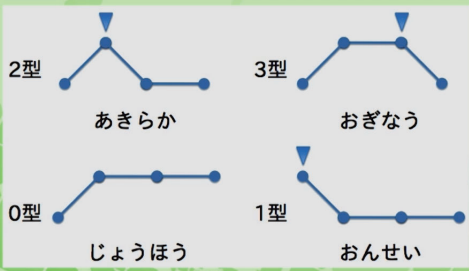


3型
おぎなう

23

デジタル情報の処理と認識

韻律情報の付加



2型
あきらか

3型
おぎなう

0型
じょうほう

1型
おんせい

24

デジタル情報の処理と認識

韻律情報の付加

- 単語のつながりによっても変化



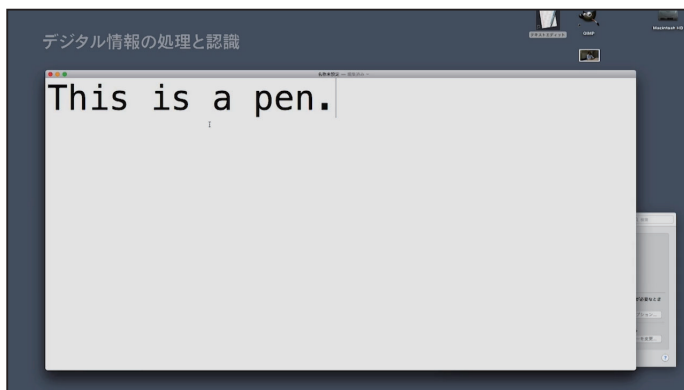
歩く せる → 歩かせる

- 発音のし始めはピッチが高く
徐々にピッチが低くなる

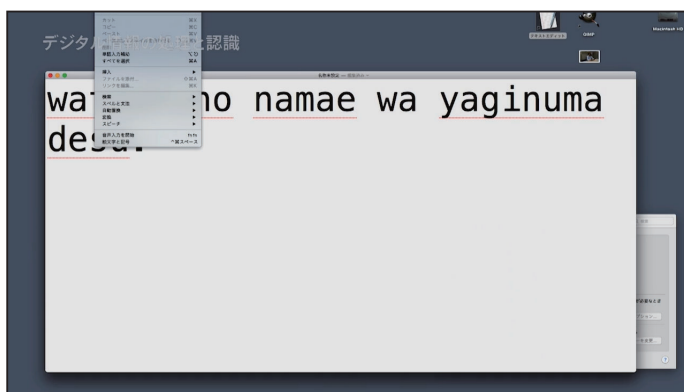
デジタル情報の処理と認識（'22）

第6回 音声の認識と合成

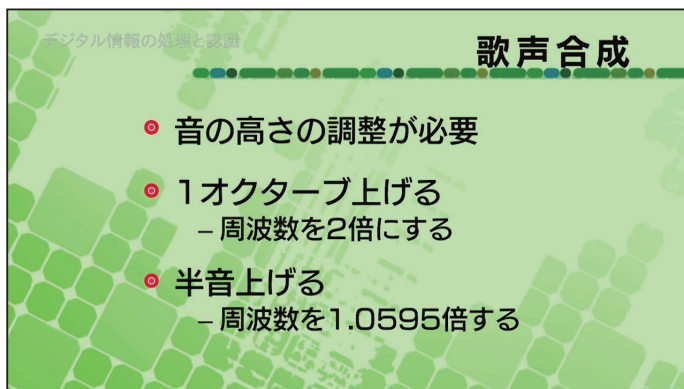
25



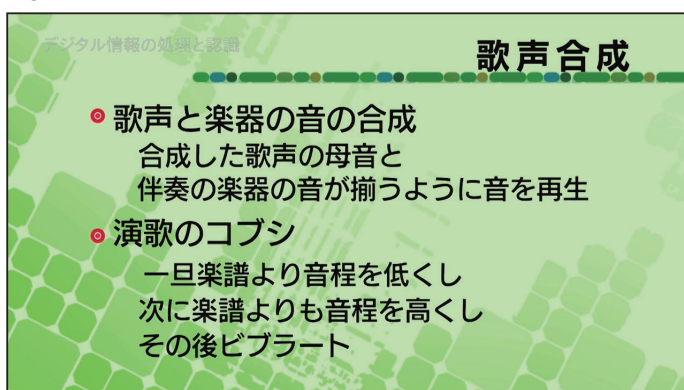
26



27



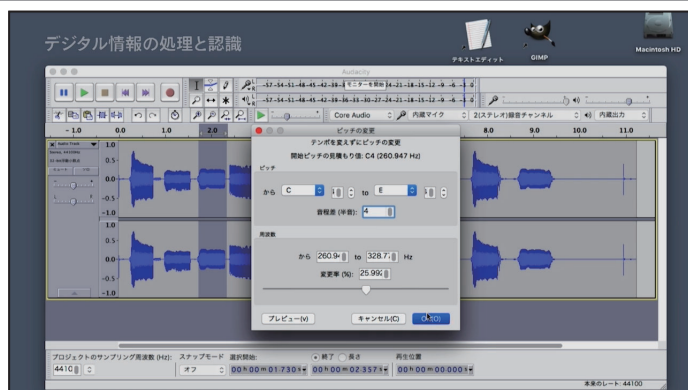
28



デジタル情報の処理と認識（'22）

第6回 音声の認識と合成

29



30

