統計的機械学習問題としての音声合成

徳田　恵一
（名古屋工業大学）

音学シンポジウム2013, 東京
2013年5月11日

はじめに

- ルールベース： フォルマント合成（~'90s）
  - 手動ルールによる各音素の素片を構築
- コーパスベース： 波形接続型音声合成（'90s~）
  - 音声データベースから音声素片を接続し合成
    - 単一インベントリ: ダイフォン音声合成
    - 複数インベントリ: 単位選択型音声合成
- コーパスベース： 統計的パラメトリック音声合成
  - ソース･フィルタモデル＋統計的音響モデル
    - 隠れマルコフモデルによる音声合成（HMM音声合成）

統計的な枠組みにおいて音声合成に含まれる処理はどのように定式化されるべきか？

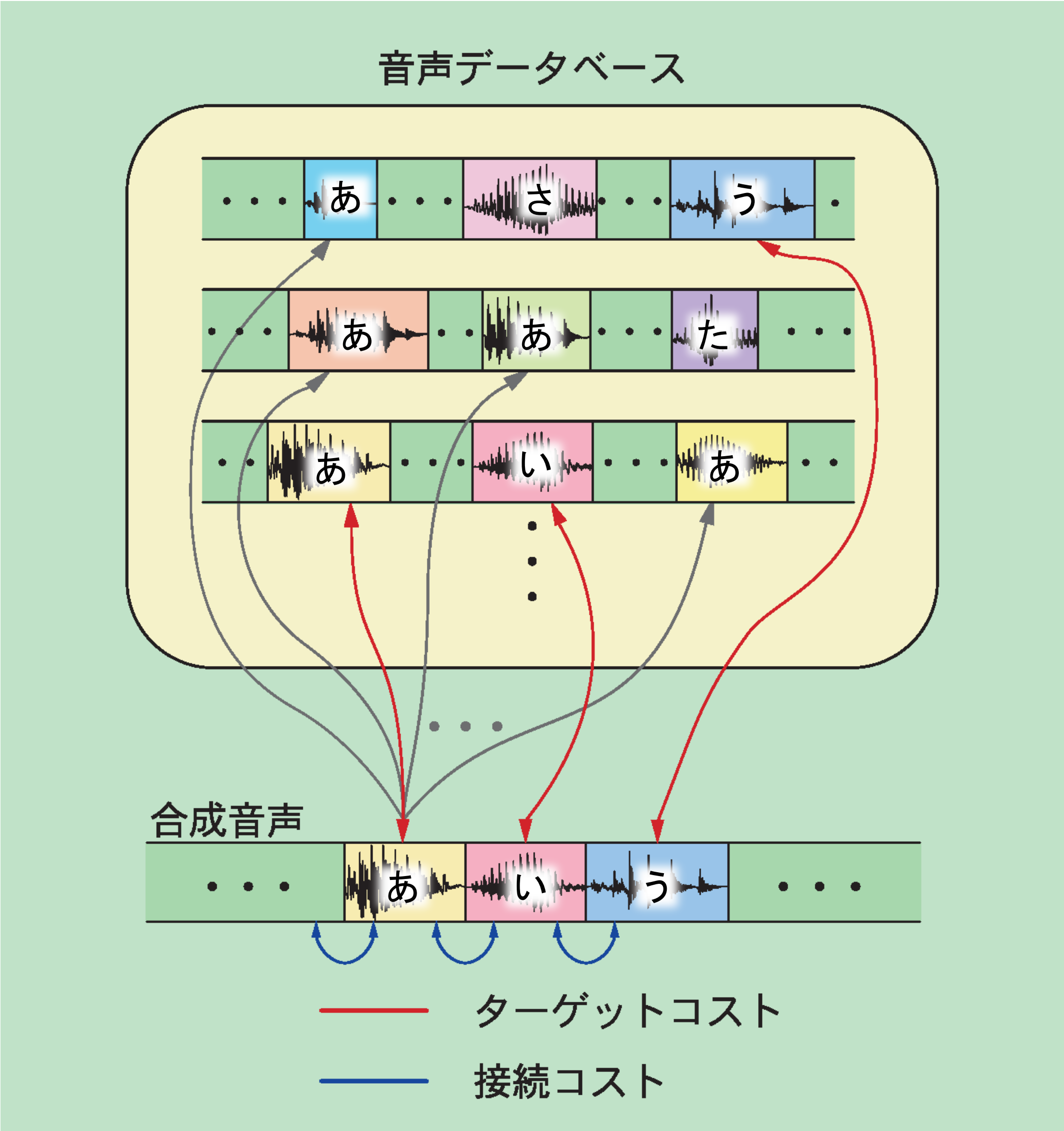
→
→
→

音声認識は<u>**技術**</u>

音声合成も<u>**技術**</u>

←

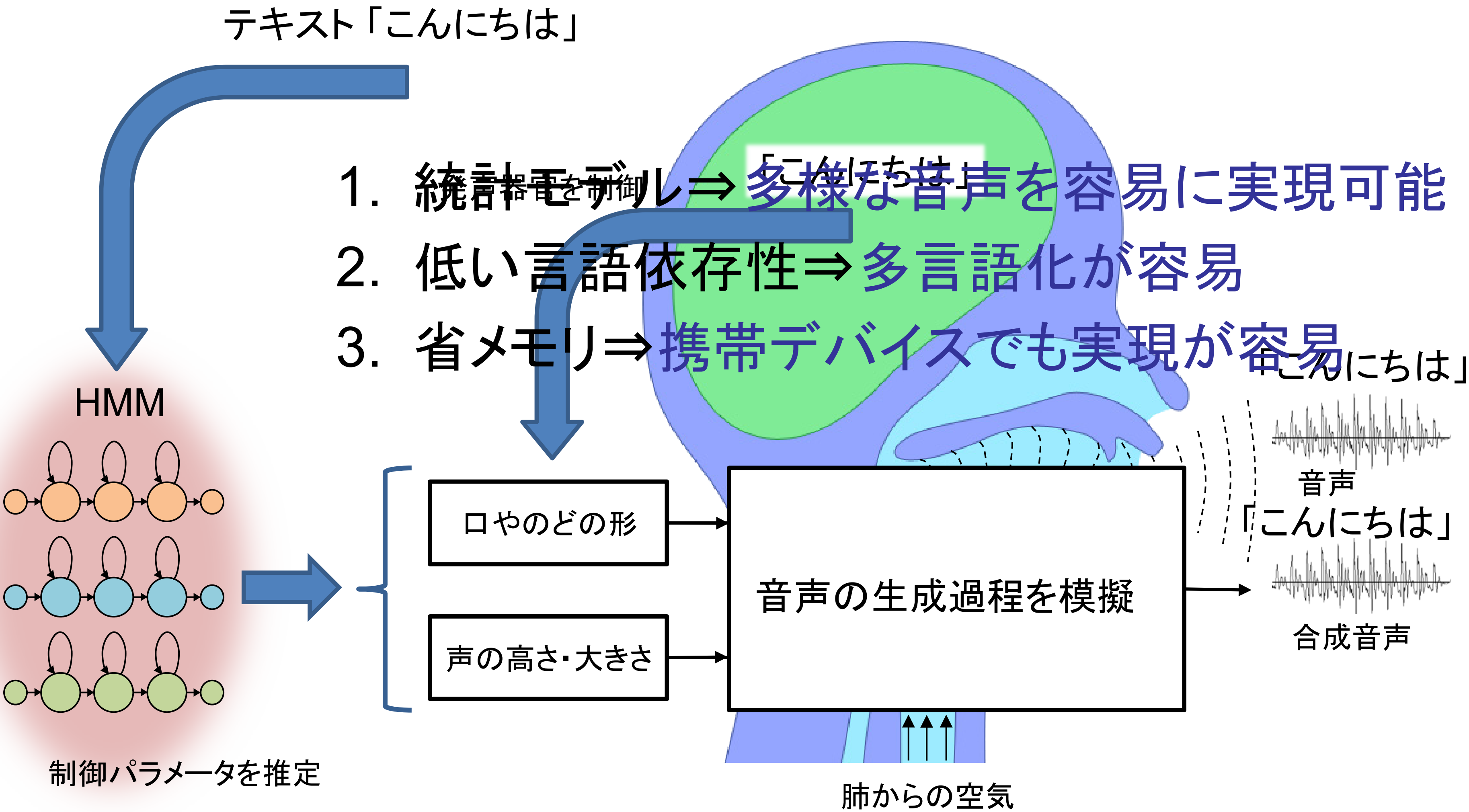
波形の接続による音声合成



- 多様な声質，発話スタイルを得るためには，大量の音声データベースが必要
- 数十MB～数百MB

HMM音声合成

1. 統計モデル⇒多様な音声を容易に実現可能
2. 低い言語依存性⇒多言語化が容易
3. 省メモリ⇒携帯デバイスでも実現が容易



音声合成の基本問題

音声データベース（音声波形と対応するテキスト）があるとき，任意に与えられたテキストに対応する音声波形を求めよ.

$\boldsymbol{W}$: テキスト
$\boldsymbol{X}$: 音声波形
（$\boldsymbol{W}$, $\boldsymbol{X}$）→ 音声データベース

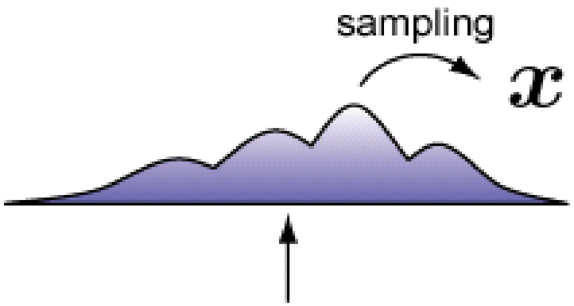
$\boldsymbol{w}$: 合成テキスト
（音声データベース, $\boldsymbol{w}$）→ 与えられる情報

$\boldsymbol{x}$: 合成音声波形 ← ？

音声合成の統計的な定式化(1/8)

ベイズの枠組みによる予測問題

$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$ から $\tilde{\boldsymbol{x}}$ を生成



$\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{w}$

$\boldsymbol{W}$: テキスト

$\boldsymbol{X}$: 音声波形

（$\boldsymbol{W}$, $\boldsymbol{X}$）音声データベース

$\boldsymbol{w}$: 合成テキスト

（$\boldsymbol{W}$, $\boldsymbol{X}$, $\boldsymbol{w}$）与えられる情報

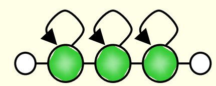
$\boldsymbol{x}$: 合成音声波形 ← ?

1. 与えられた情報から予測分布を推定
2. 予測分布からサンプルを生成

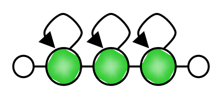
音声合成の統計的な定式化(2/8)

1. 予測分布の推定は困難☹
   → 音響モデルのパラメータを導入

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$$

⇓ 音響モデル λ の導入 

$$= \int p(\boldsymbol{x}, \lambda \mid \boldsymbol{w}, \boldsymbol{W}, \boldsymbol{X}) d\lambda = \int \underbrace{p(\boldsymbol{x} \mid \boldsymbol{w}, \lambda)}_{\text{生成}} \underbrace{p(\lambda \mid \boldsymbol{W}, \boldsymbol{X})}_{\text{学習}} d\lambda$$

λ : 音響モデル (e.g. HMM) 

音声合成の統計的な定式化(3/8)

2. 音声波形の直接的な利用は困難☹

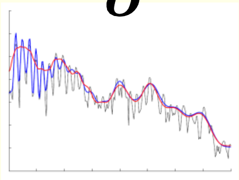
→ 音声波形のパラメトリックな表現を導入

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$$

$$= \int \underline{p(\boldsymbol{x} \mid \boldsymbol{w}, \lambda)} \, \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})} d\lambda$$

$\boldsymbol{x}$

$\boldsymbol{o}$



⇓ 音声波形のパラメトリック表現 $\boldsymbol{o}$ の導入

$$= \iint \underline{p(\boldsymbol{x} \mid \boldsymbol{o}) p(\boldsymbol{o} \mid \boldsymbol{w}, \lambda)} \, \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})} d\lambda d\boldsymbol{o}$$

$\boldsymbol{o}$: 音声波形 $\boldsymbol{x}$ のパラメトリック表現
(e.g., メルケプストラム, LPC, LSP, F0, 非周期性)

音声合成の統計的な定式化(4/8)

3. 同一のテキストであっても複数の発音や品詞情報を持ちうる☹
→ ラベルの導入

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$$

$$= \iint \underline{p(\boldsymbol{x} \mid \boldsymbol{o}) p(\boldsymbol{o} \mid \boldsymbol{w}, \lambda)} \, \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})} d\lambda d\boldsymbol{o}$$

⇓ テキストから得られるラベル $\boldsymbol{l}$ の導入

$$= \iint \sum_{\forall \boldsymbol{l}} \underline{p(\boldsymbol{x} \mid \boldsymbol{o}) p(\boldsymbol{o} \mid \boldsymbol{l}, \lambda) P(\boldsymbol{l} \mid \boldsymbol{w})} \, \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})} d\lambda d\boldsymbol{o}$$

$\boldsymbol{l}$: テキスト $\boldsymbol{w}$ から得られるラベル
（e.g., 読み，品詞，アクセント，ポーズ等）

音声合成の統計的な定式化(5/8)

4. 全変数の積分計算は困難☹
→ 全変数の同時最適化による近似

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$$

$$= \iint \sum_{\forall \boldsymbol{l}} \underline{p(\boldsymbol{x} \mid \boldsymbol{o}) p(\boldsymbol{o} \mid \boldsymbol{l}, \lambda) P(\boldsymbol{l} \mid \boldsymbol{w})} \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})} d\lambda d\boldsymbol{o}$$

⇓ 全変数の同時最適化による積分計算の近似を導入

$$\approx \underline{p(\boldsymbol{x} \mid \hat{\boldsymbol{o}}) p(\hat{\boldsymbol{o}} \mid \hat{\boldsymbol{l}}, \hat{\lambda}) P(\hat{\boldsymbol{l}} \mid \boldsymbol{w})} \underline{p(\hat{\lambda} \mid \boldsymbol{X}, \boldsymbol{W})}$$

ただし

$$\left\{ \hat{\boldsymbol{o}}, \hat{\boldsymbol{l}}, \hat{\lambda} \right\} = \arg\max_{\boldsymbol{o}, \boldsymbol{l}, \lambda} \underline{p(\boldsymbol{x} \mid \boldsymbol{o}) p(\boldsymbol{o} \mid \boldsymbol{l}, \lambda) P(\boldsymbol{l} \mid \boldsymbol{w})} \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})}$$

音声合成の統計的な定式化(6/8)

5. 同時最大化は困難☹

→ ステップ毎の最適化に分解

$$\left\{\hat{\boldsymbol{o}}, \hat{\boldsymbol{l}}, \hat{\lambda}\right\} = \arg\max_{\boldsymbol{o},\boldsymbol{l},\lambda} \underline{p(\boldsymbol{x} \mid \boldsymbol{o})p(\boldsymbol{o} \mid \boldsymbol{l}, \lambda)P(\boldsymbol{l} \mid \boldsymbol{w})}\,\underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})}$$

⇓ ステップ毎の最適化に分解

$$\hat{\lambda} = \arg\max_{\lambda} p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})$$ ⇐ 学習

$$\hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{l}} P(\boldsymbol{l} \mid \boldsymbol{w})$$ ⇐ テキスト解析

$$\hat{\boldsymbol{o}} = \arg\max_{\boldsymbol{o}} p(\boldsymbol{o} \mid \hat{\boldsymbol{l}}, \hat{\lambda})$$ ⇐ 音声パラメータ生成

音声合成の統計的な定式化(7/8)

6. 学習時も同様なパラメータ表現が必要☹
→ 同様の近似を導入

$$\hat{\lambda} = \arg\max_{\lambda} \underline{p(\lambda \mid \boldsymbol{X}, \boldsymbol{W})}$$

⇓ ステップ毎の最適化に分解

$$\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} P(\boldsymbol{L} \mid \boldsymbol{W})$$ ⇐ ラベリング

$$\hat{\boldsymbol{O}} = \arg\max_{\boldsymbol{O}} p(\boldsymbol{X} \mid \boldsymbol{O})$$ ⇐ 特徴抽出

$$\hat{\lambda} = \arg\max_{\lambda} p(\hat{\boldsymbol{O}} \mid \hat{\boldsymbol{L}}, \lambda) p(\lambda)$$ ⇐ 音響モデル学習

$\boldsymbol{O}$: 音声波形 $\boldsymbol{X}$ のパラメトリック表現

$\boldsymbol{L}$: テキスト $\boldsymbol{W}$ から抽出されたラベル

音声合成の統計的な定式化(8/8)

Draw $\tilde{\boldsymbol{x}}$ from $p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$

⇩

$\hat{\boldsymbol{O}} = \arg\max_{\boldsymbol{O}} p(\boldsymbol{X} \mid \boldsymbol{O})$ ⇐ 特徴抽出

$\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} P(\boldsymbol{L} \mid \boldsymbol{W})$ ⇐ ラベリング

$\hat{\lambda} = \arg\max_{\lambda} p(\hat{\boldsymbol{O}} \mid \hat{\boldsymbol{L}}, \lambda) p(\lambda)$ ⇐ 音響モデル学習

$\hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{l}} P(\boldsymbol{l} \mid \boldsymbol{w})$ ⇐ テキスト解析

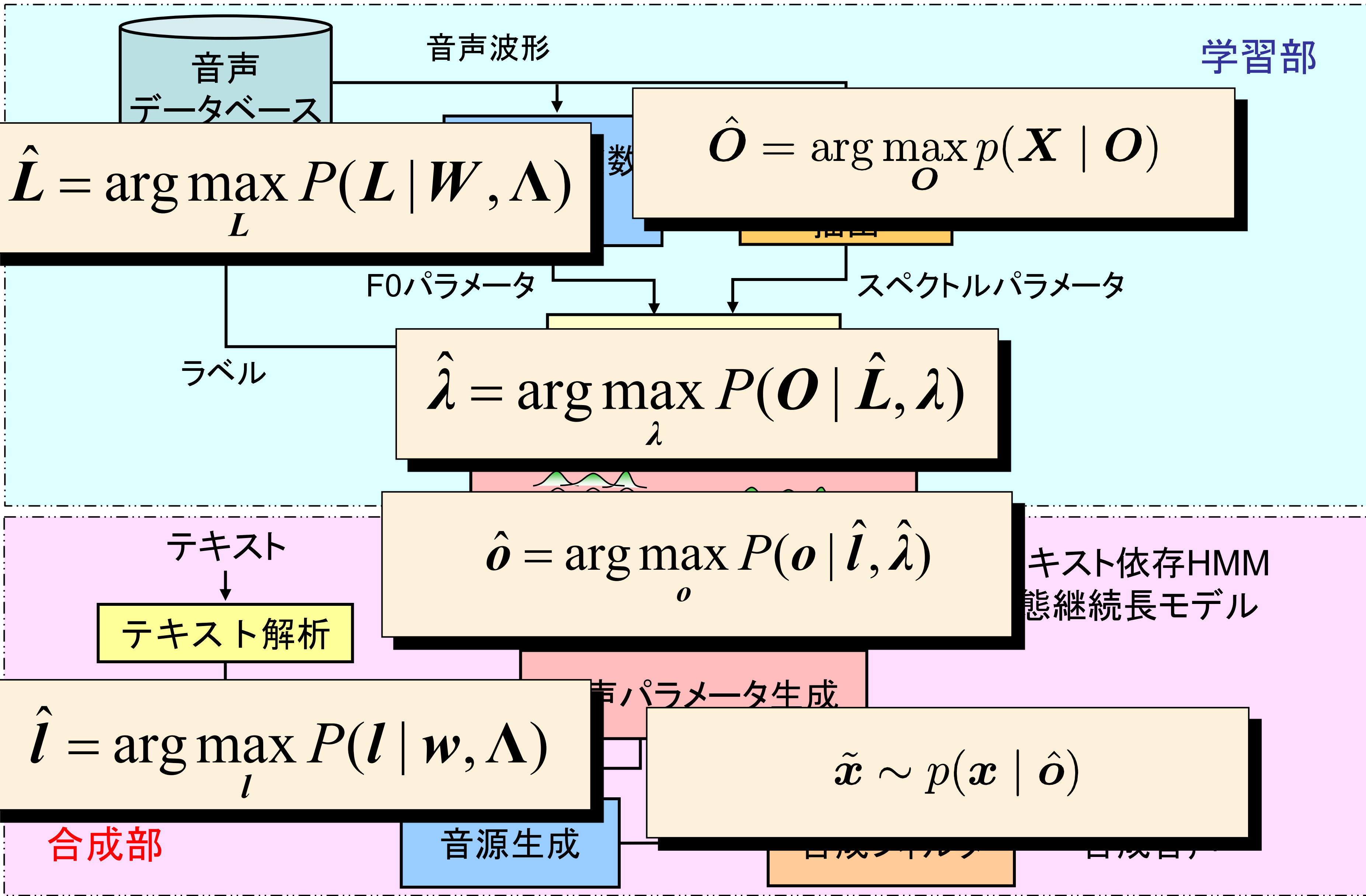
$\hat{\boldsymbol{o}} = \arg\max_{\boldsymbol{o}} p(\boldsymbol{o} \mid \hat{\boldsymbol{l}}, \hat{\lambda})$ ⇐ 音声パラメータ生成

$\tilde{\boldsymbol{x}}$ from $p(\boldsymbol{x} \mid \hat{\boldsymbol{o}})$ ⇐ 音声波形生成

概要

1. 定式化
2. 各要素の実現
3. 多様な音声合成
4. むすび

HMM音声合成の枠組み



$$\hat{L} = \arg\max_{L} P(L \mid W, \Lambda)$$

$$\hat{O} = \arg\max_{O} p(X \mid O)$$

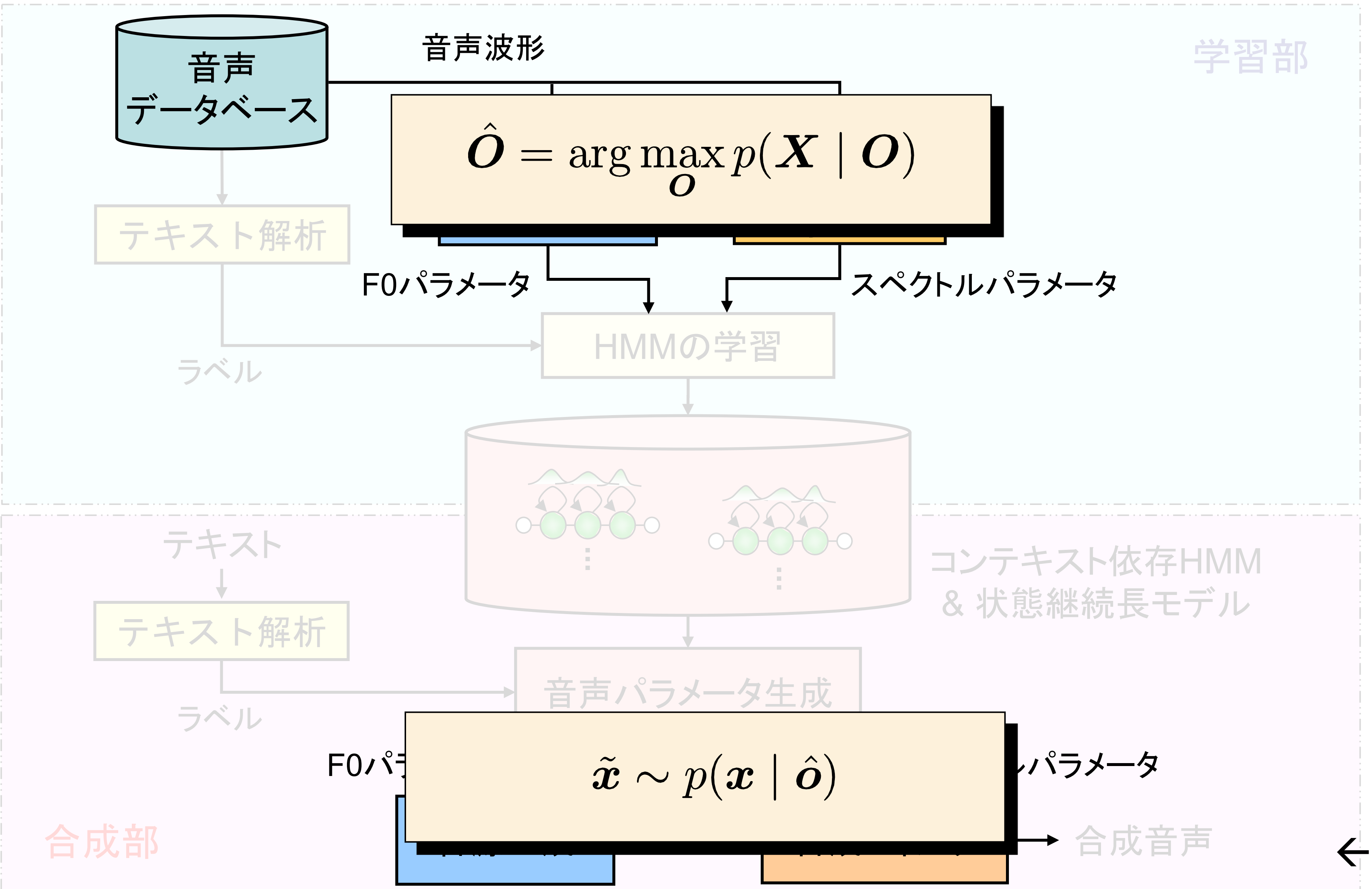
$$\hat{\lambda} = \arg\max_{\lambda} P(O \mid \hat{L}, \lambda)$$

$$\hat{o} = \arg\max_{o} P(o \mid \hat{l}, \hat{\lambda})$$

$$\hat{l} = \arg\max_{l} P(l \mid w, \Lambda)$$

$$\tilde{x} \sim p(x \mid \hat{o})$$

HMM音声合成の枠組み

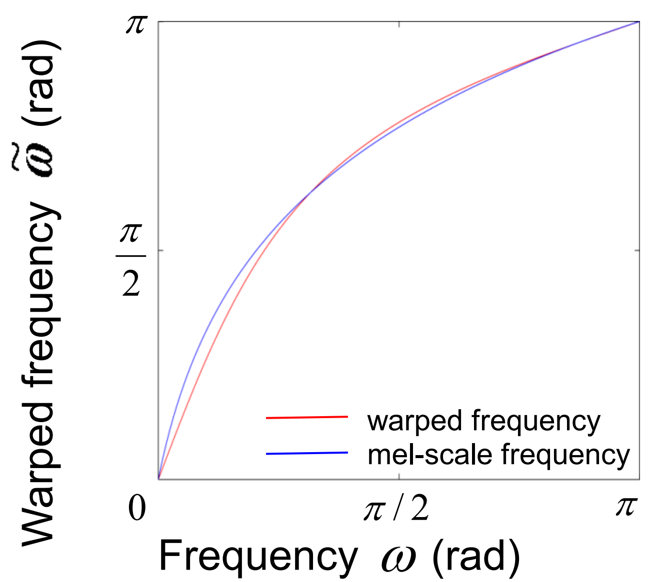


スペクトルパラメータの最尤推定

音声スペクトルのメルケプストラム表現

$$H(z) = \exp \sum_{m=0}^{M} c(m)\tilde{z}^{-m}$$

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} = e^{-j\tilde{\omega}}$$



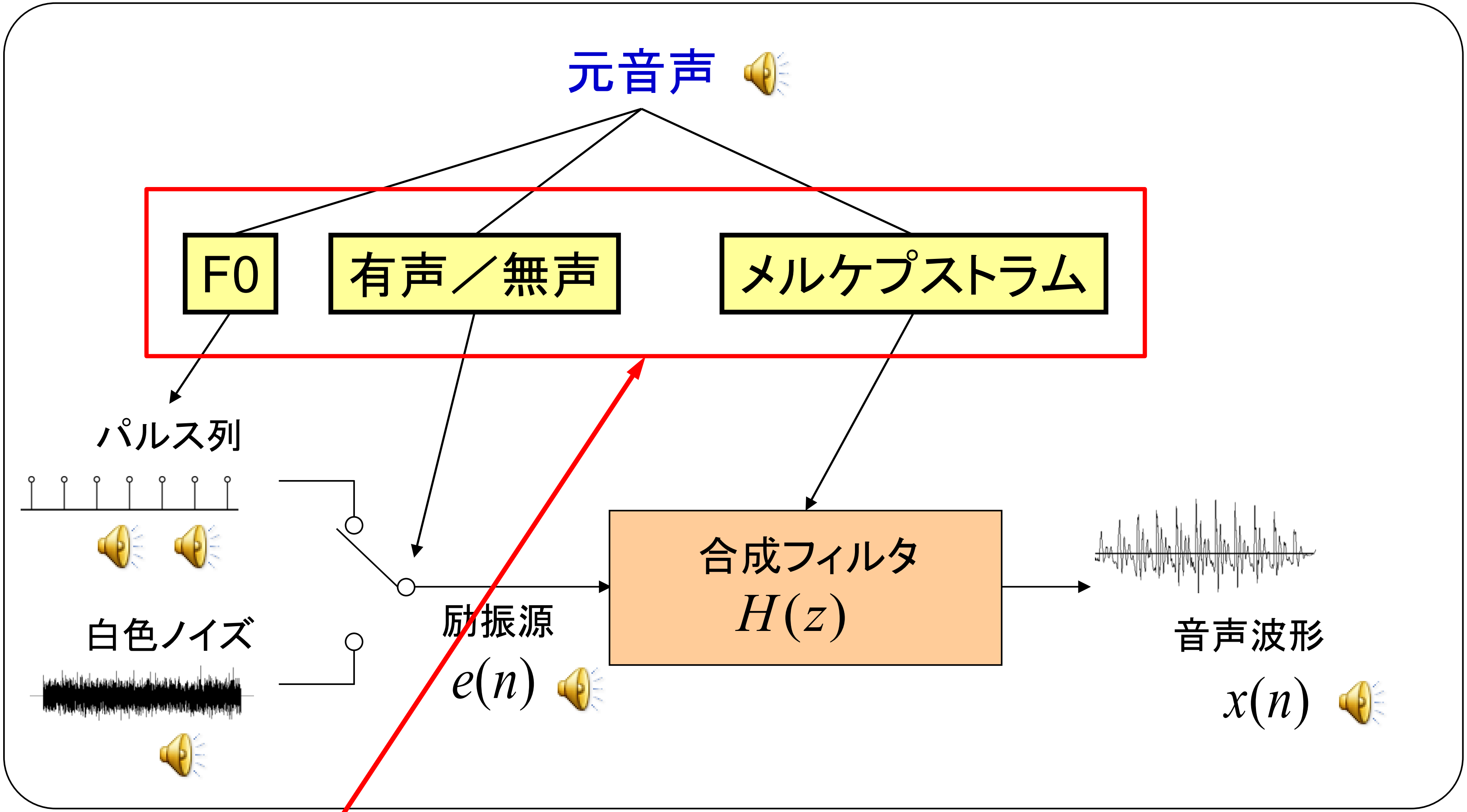
メルケプストラムの最尤推定

$$\boldsymbol{c} = \arg\max_{\boldsymbol{c}} p(\boldsymbol{x} \mid \boldsymbol{c})$$

$\boldsymbol{x}$: 音声波形（ガウス過程）

$\boldsymbol{c}$: メルケプストラム

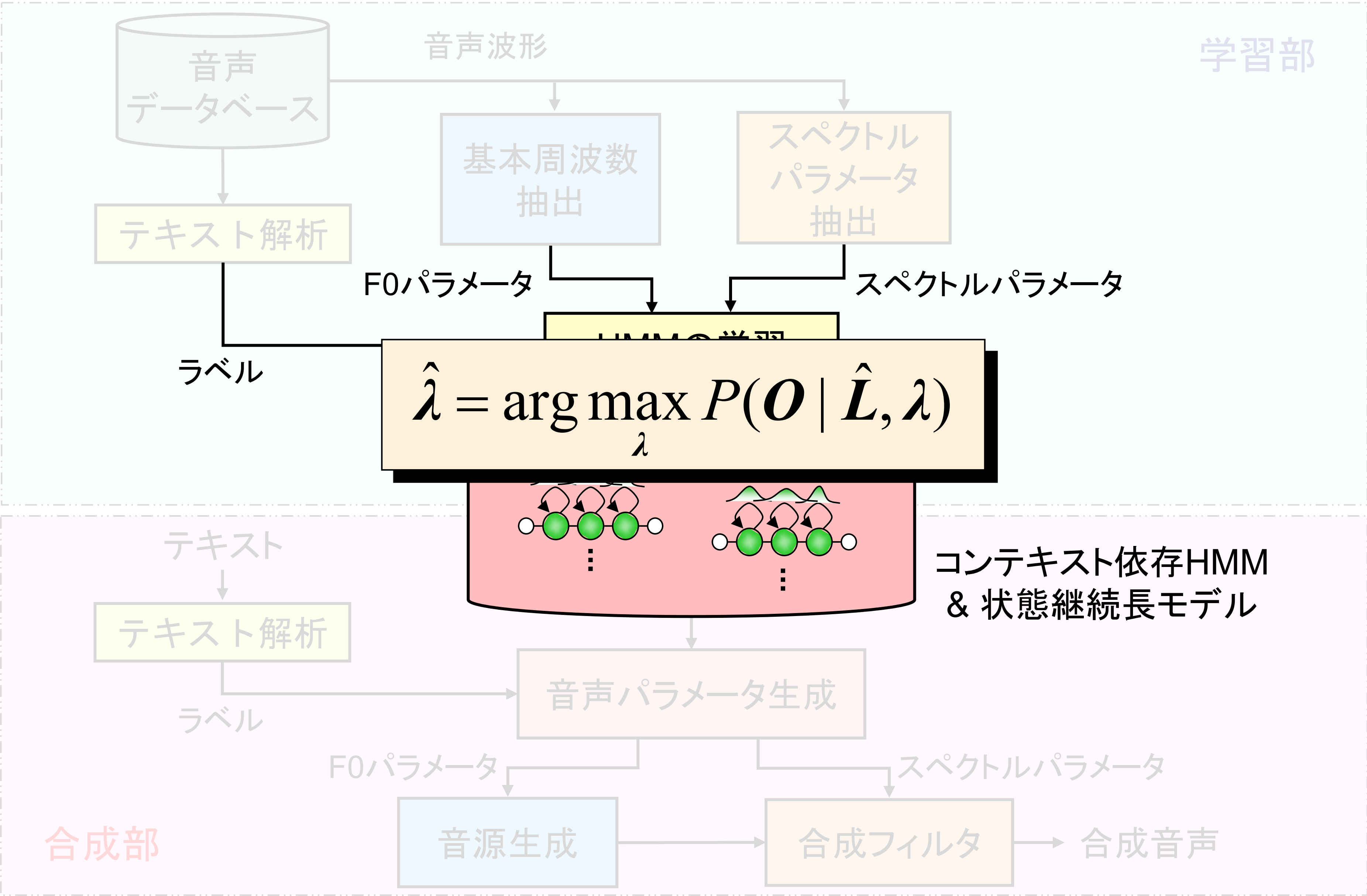
音声波形の分析再合成



これらのパラメータをHMMを用いてモデル化

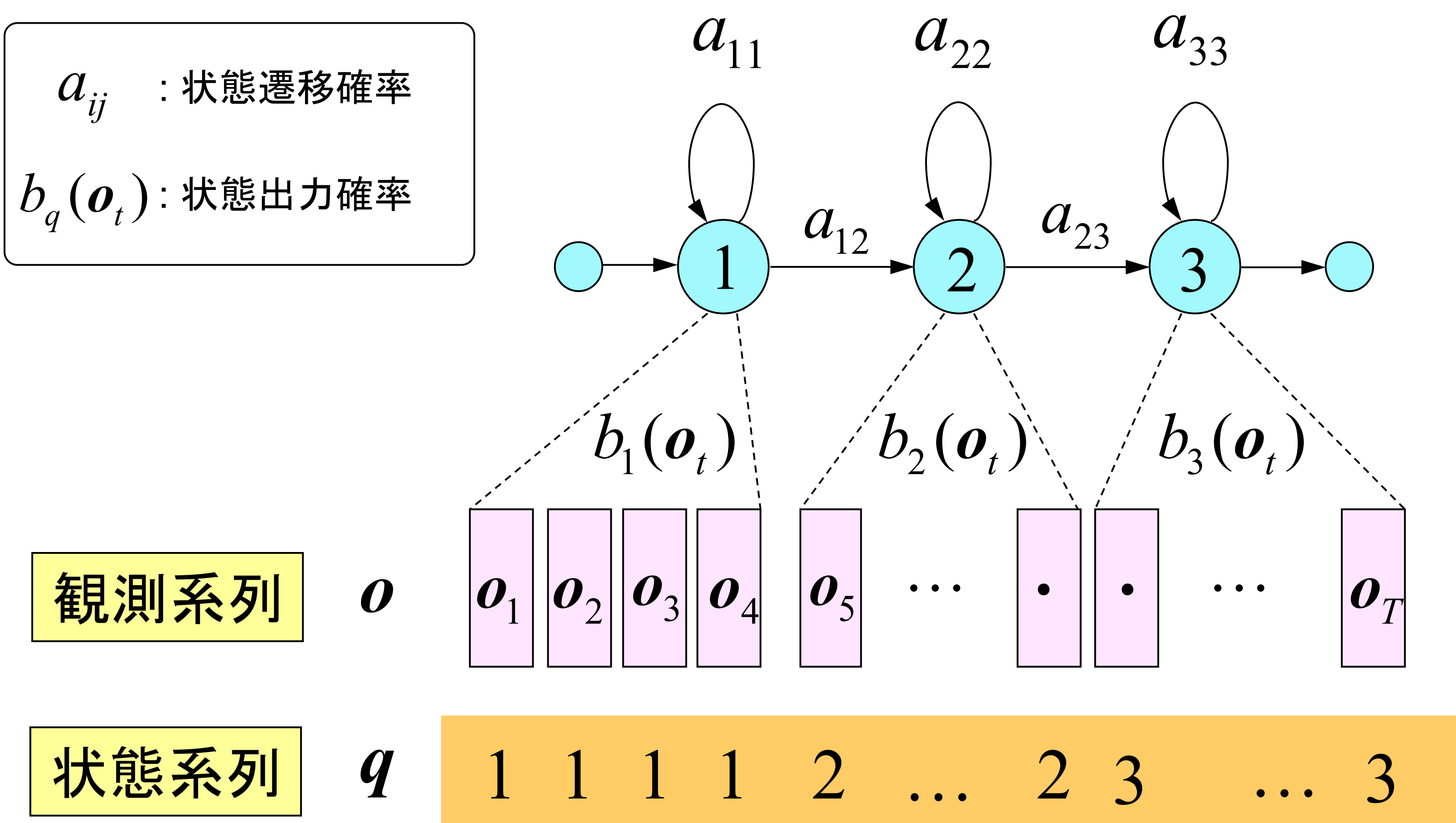
←

HMM音声合成の枠組み

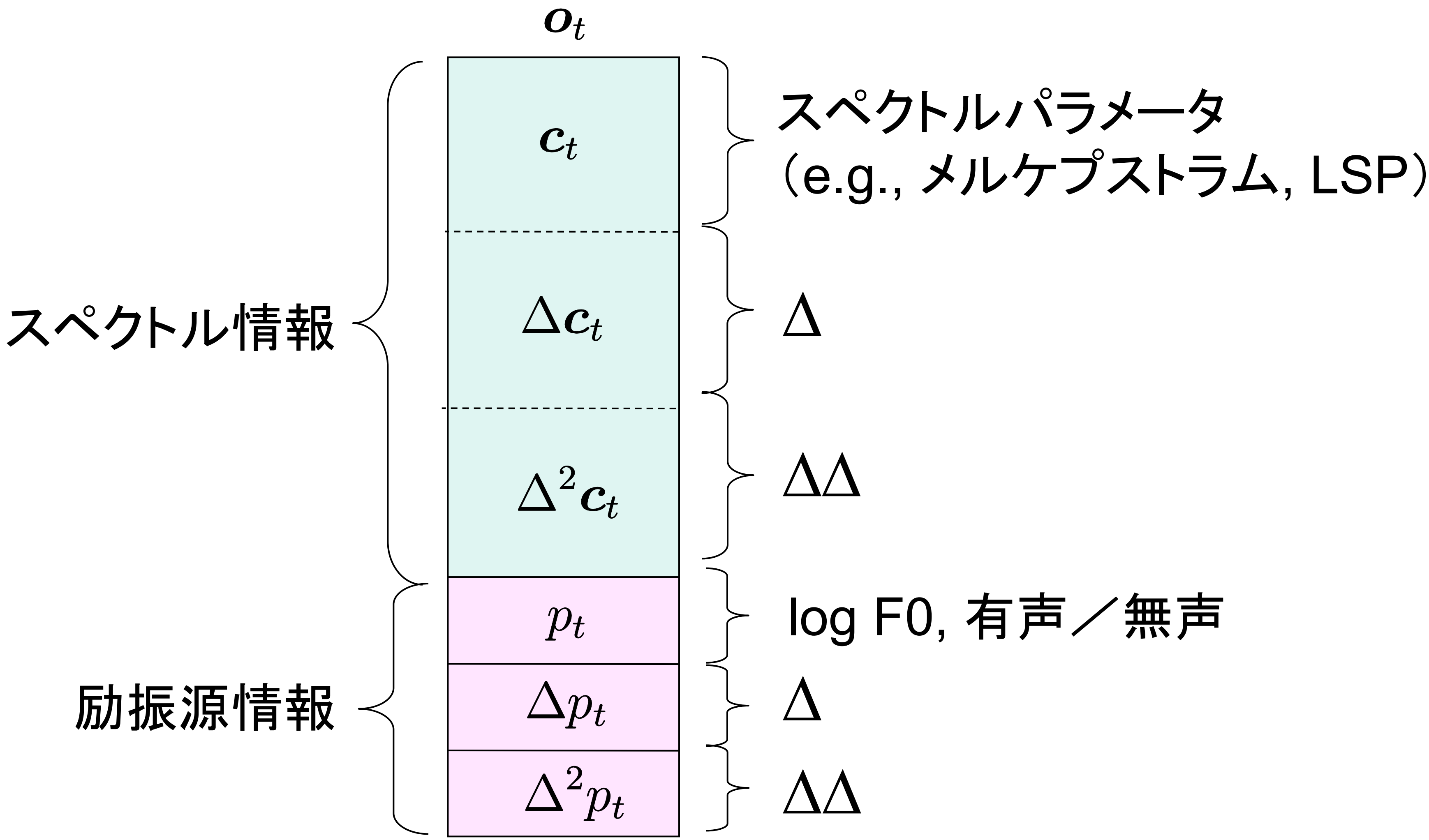


$$\hat{\lambda} = \arg\max_{\lambda} P(\boldsymbol{O} \mid \hat{\boldsymbol{L}}, \lambda)$$

隠れマルコフモデル（Hidden Markov model; HMM）



観測ベクトル（音響特徴量ベクトル）の構造



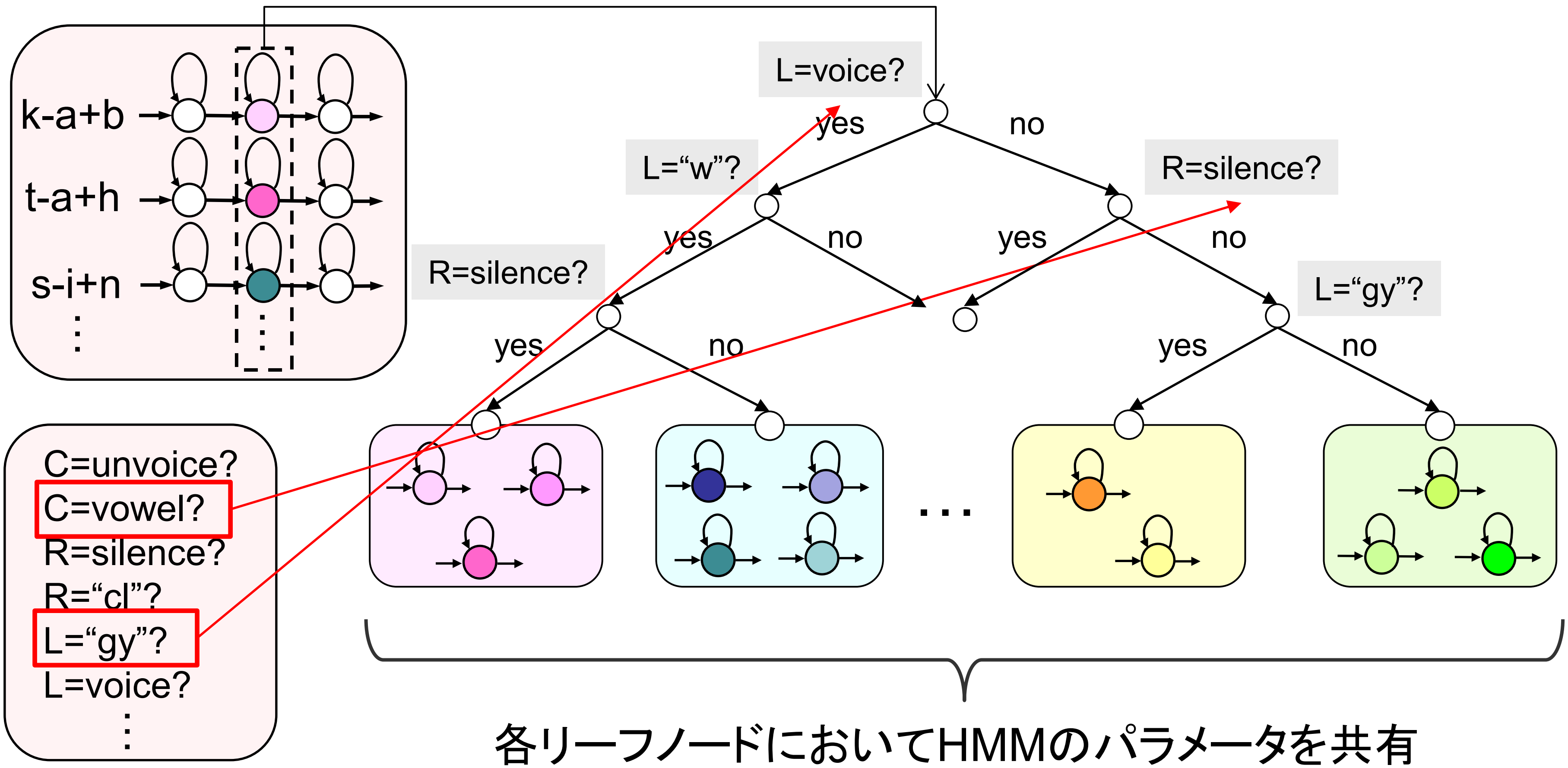
文脈要因（コンテキスト）

– 当該音素
– 先行・後続音素
– 当該音素のアクセント句でのモーラ位置
– {先行，当該，後続}の品詞，活用形，活用型
– {先行，当該，後続}のアクセント句の長さ，アクセント型
– 当該アクセント句の位置，前後のポーズの有無
– {先行，当該，後続}の呼気段落の長さ
– 当該呼気段落の位置
– 文の長さ

・・・・

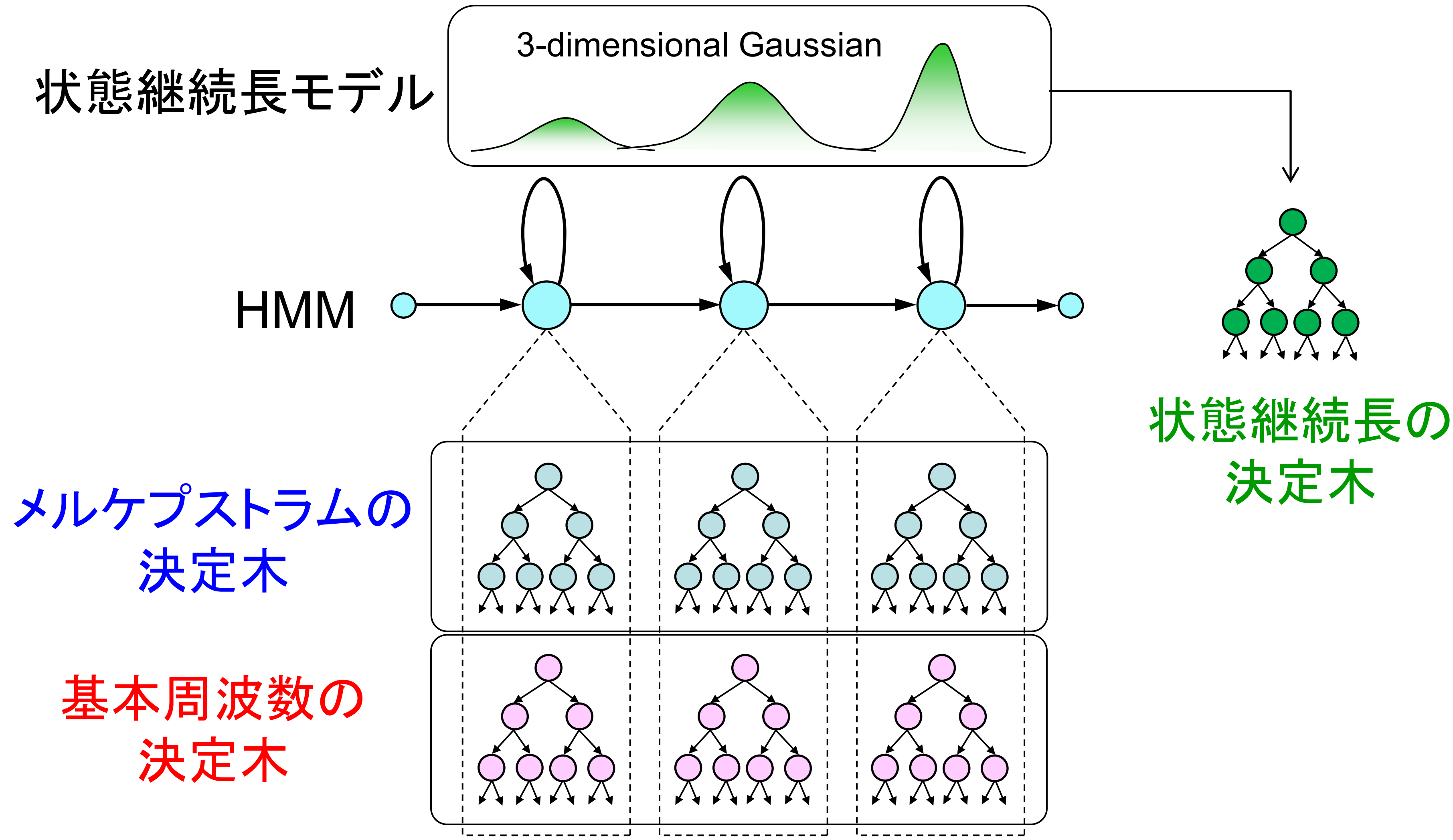
膨大な組み合わせ数 ⇒ 全コンテキストについて推定することは困難

決定木に基づくコンテキストクラスタリング [Odell; '95]

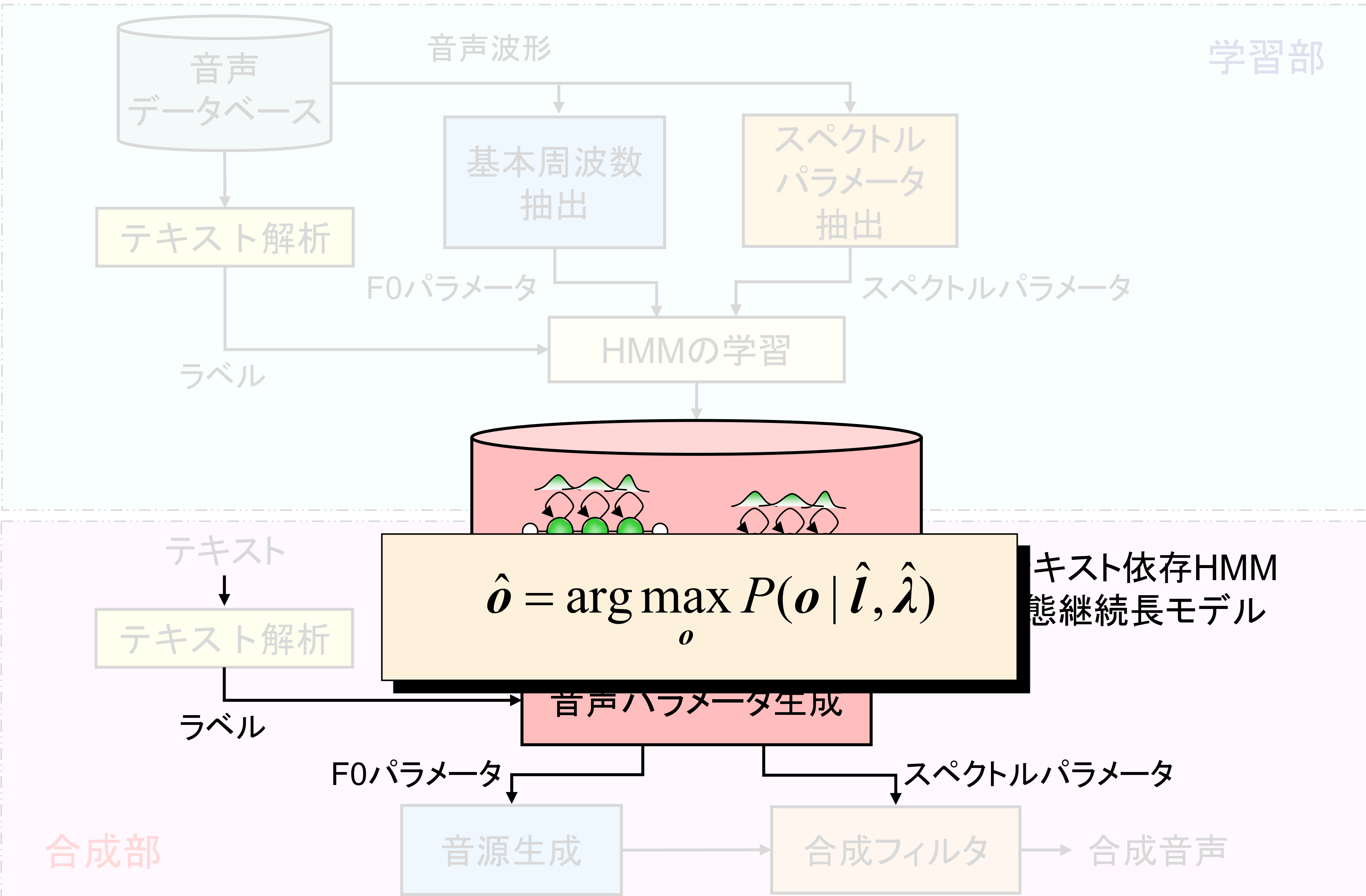


各リーフノードにおいてHMMのパラメータを共有

ストリーム依存決定木クラスタリング



HMM音声合成の枠組み



$$\hat{\boldsymbol{o}} = \arg\max_{\boldsymbol{o}} P(\boldsymbol{o} \mid \hat{\boldsymbol{l}}, \hat{\boldsymbol{\lambda}})$$

音声パラメータ生成アルゴリズム [Tokuda; ’00]

与えられた文HMMに対して出力確率が最大となる音声パラメータ系列 $\boldsymbol{o} = [\boldsymbol{o}_1^\top, \boldsymbol{o}_2^\top, \ldots, \boldsymbol{o}_T^\top]^\top$ を推定

$$P(\boldsymbol{o} \mid \hat{\boldsymbol{l}}, \hat{\lambda}) = \sum_{\boldsymbol{q}} P(\boldsymbol{o} \mid \boldsymbol{q}, \hat{\lambda}) P(\boldsymbol{q} \mid \hat{\boldsymbol{l}}, \hat{\lambda})$$

$$\approx \max_{\boldsymbol{q}} P(\boldsymbol{o} \mid \boldsymbol{q}, \hat{\lambda}) P(\boldsymbol{q} \mid \hat{\boldsymbol{l}}, \hat{\lambda})$$

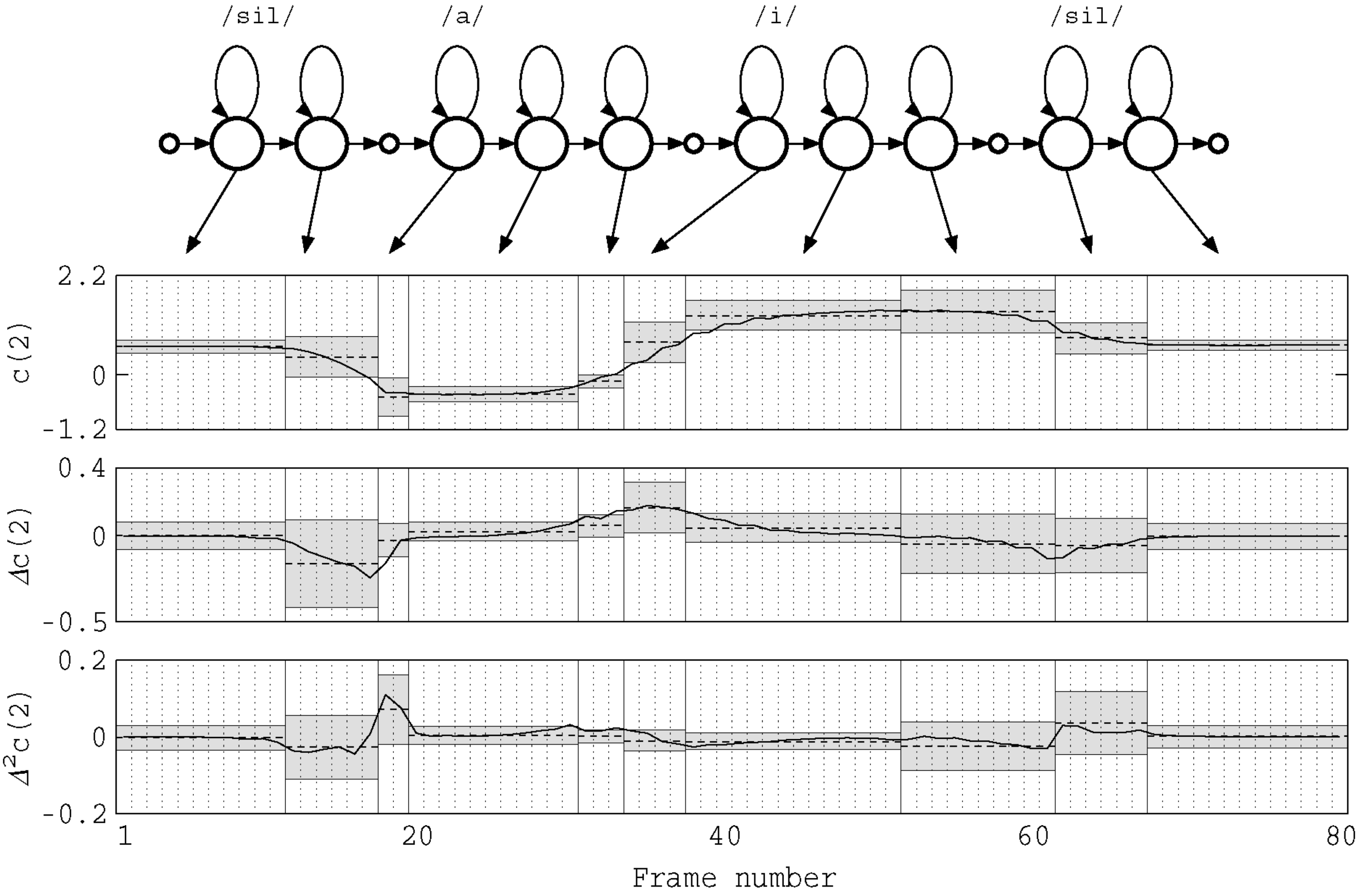
$$\Downarrow$$

$$\hat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} P(\boldsymbol{q} \mid \hat{\boldsymbol{l}}, \hat{\lambda})$$

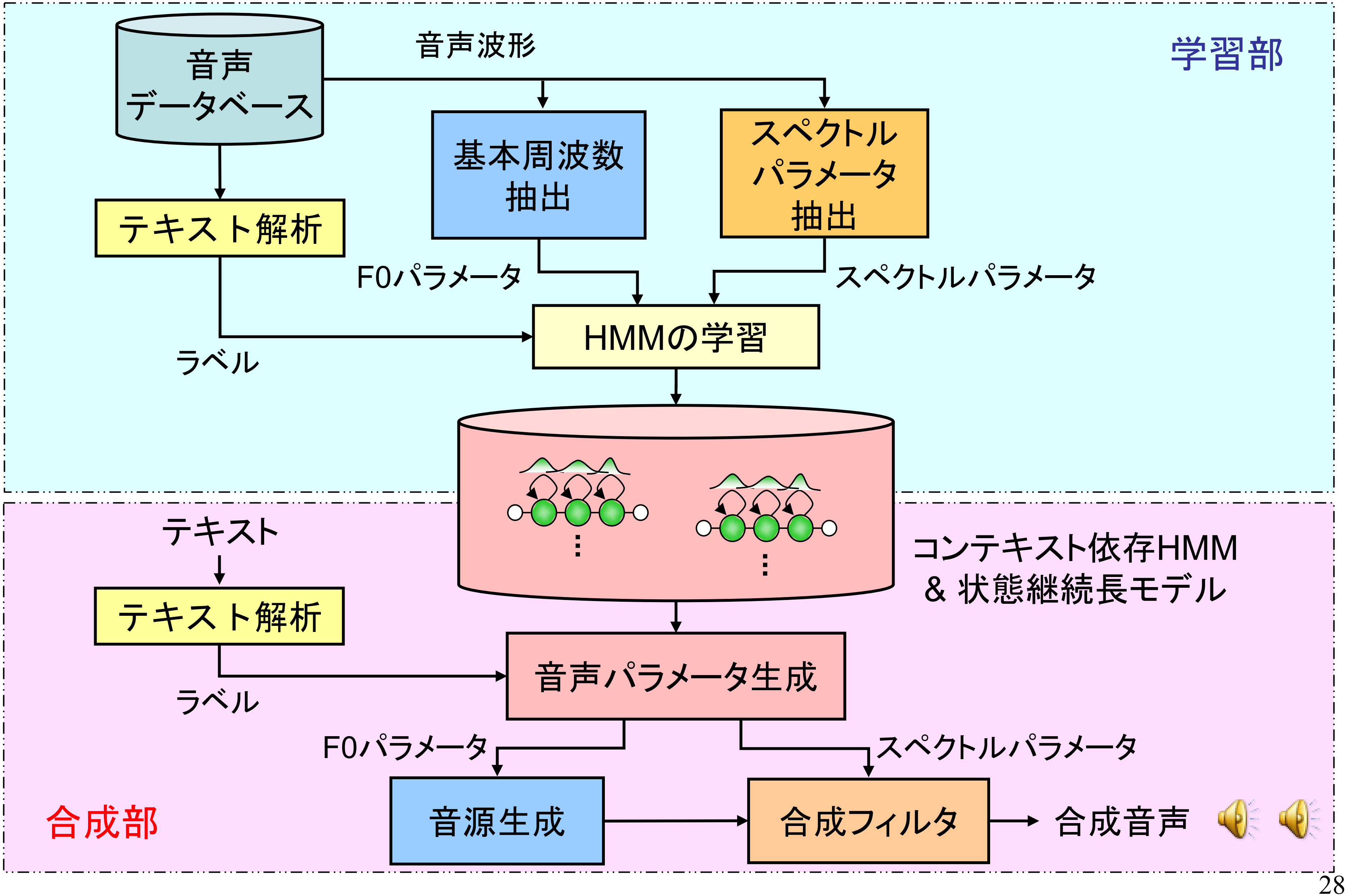
$$\hat{\boldsymbol{o}} = \arg\max_{\boldsymbol{o}} P(\boldsymbol{o} \mid \hat{\boldsymbol{q}}, \hat{\lambda})$$

→

生成される音声パラメータ系列の軌跡



HMM音声合成の枠組み



概要

1. 定式化
2. 各要素の実現
3. 多様な音声合成
4. むすび

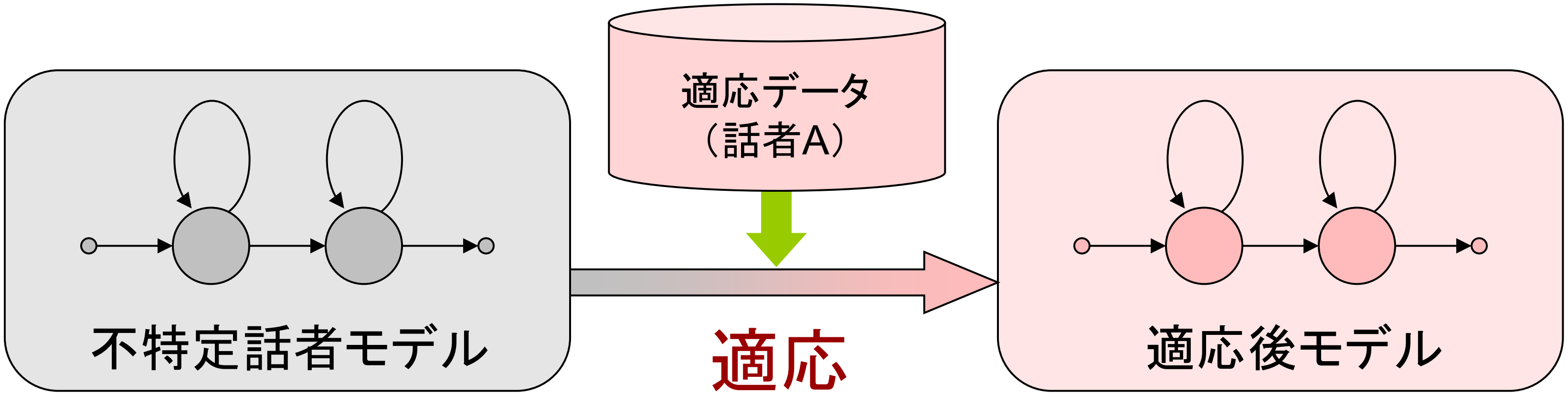
感情音声合成

	平静	怒り
「授業中に携帯いじってんじゃねえよ！電源切っとけ！」		
「ミーティングには毎週参加しなさい！」		

※学習データ200発話

話者適応(声を真似る)

線形回帰による話者適応



初期モデル

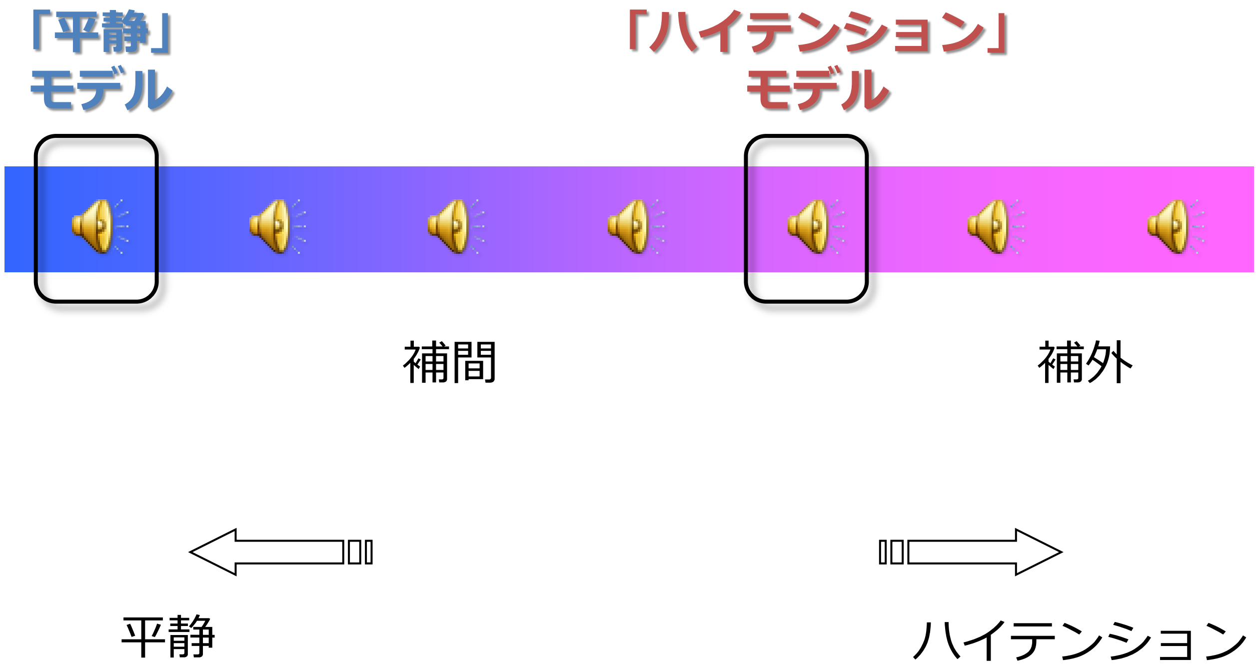
4発話による適応

50発話による適応

特定話者モデル(話者A)

?

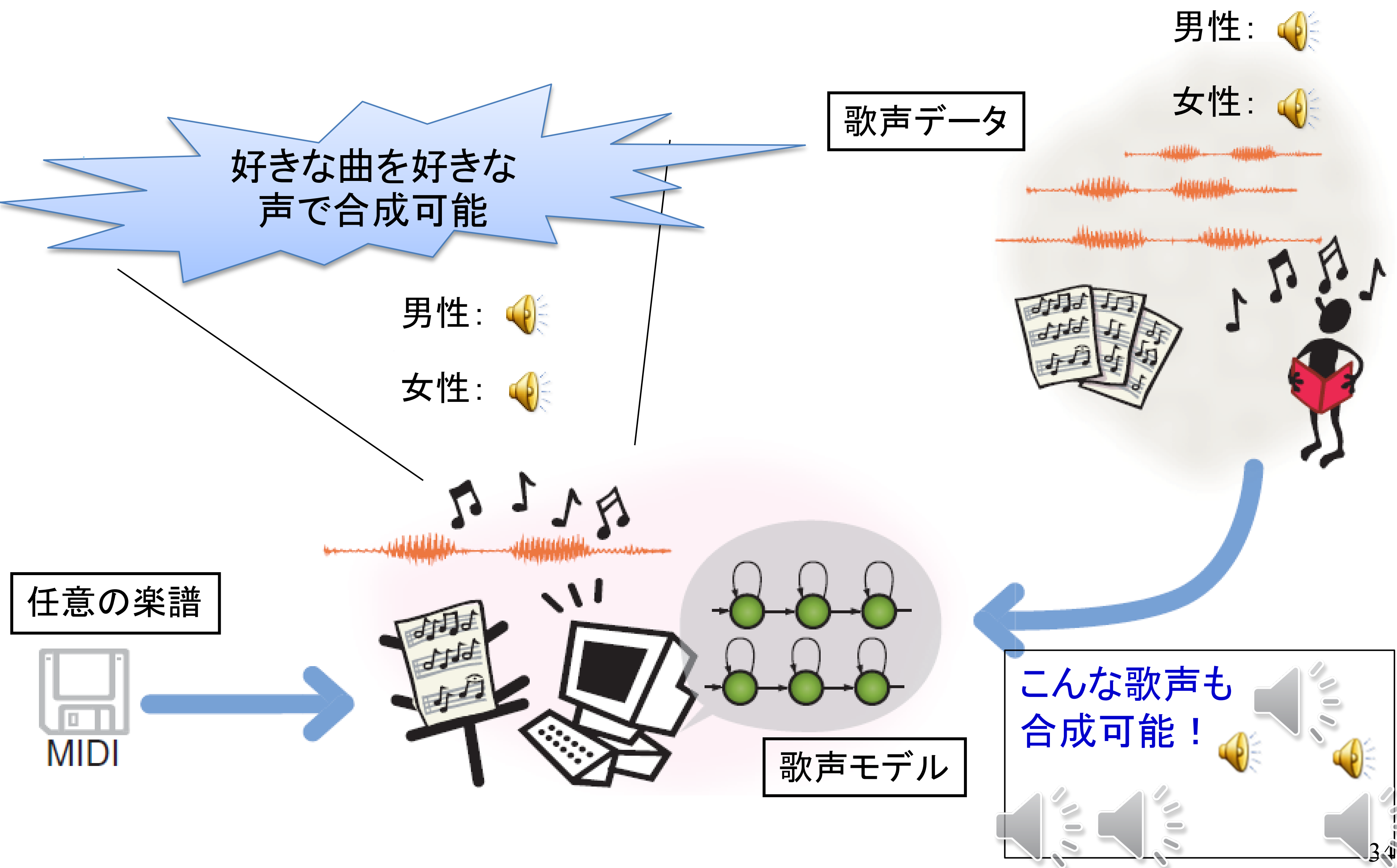
発話スタイルの補間（声を混ぜる）



多言語音声合成

- 日本語
- American English
- Chinese (Mandarin) (by ATR)
- Brazilian Portuguese (by Nitech, and UFRJ)
- European Portuguese (by Nitech, Univ of Porto, and UFRJ)
- Slovenian
  (by Bostjan Vesnicer, University of Ljubljana, Slovenia)
- Swedish (by Anders Lundgren, KTH, Sweden)
- German (by University of Bonn, and Nitech)
- Korean (by Sang-Jin Kim, ETRI, Korea)
- Finish (by TKK, Finland)
- Baby English (by Univ of Edinburgh, UK)
- Polish, Slovak, Arabic, Farsi, Croatian, Polyglot, etc.

歌声合成

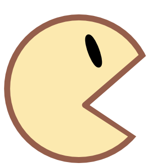


好きな曲を好きな
声で合成可能
男性：
女性：
歌声データ
男性：
女性：
任意の楽譜
MIDI
歌声モデル
こんな歌声も
合成可能！

デモ（HMM歌声合成）

才野作詞作曲のオリジナル曲
（HMMは収録した13曲すべてから学習）

Sinsy
HMM-based Singing Voice Synthesis System



リズムに乗っかってけば
いくつになったって get down
ひとつ残らず今日は
飲み込んでこう ほら
いつもと同じノリで
はしゃごうぜ皆で all the day
このビートのもっと向こうが 俺らの本場

概要

1. 定式化
2. 各要素の実現
3. 多様な音声合成
4. まとめ

全要素の統計的統合

統計的パラメトリック音声合成の問題設定

$P(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W})$ から $\hat{\boldsymbol{x}}$ を生成

$$P(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{X}, \boldsymbol{W}) = \sum_{\boldsymbol{l},\boldsymbol{L}} \iint \underbrace{P(\boldsymbol{x} \mid \boldsymbol{c})}_{\text{波形生成}} \underbrace{P(\boldsymbol{c} \mid \boldsymbol{l}, \boldsymbol{\lambda})}_{\text{音声パラメータ生成}} \underbrace{P(\boldsymbol{l} \mid \boldsymbol{w}, \boldsymbol{\Lambda})}_{\text{テキスト解析}}$$

$$\times \underbrace{P(\boldsymbol{\lambda} \mid \boldsymbol{C}, \boldsymbol{L})}_{\text{音響モデルパラメータ}}$$

$$\times \underbrace{P(\boldsymbol{L} \mid \boldsymbol{W}, \boldsymbol{\Lambda})}_{\text{ラベリング}} \underbrace{P(\boldsymbol{C} \mid \boldsymbol{X})}_{\text{特徴抽出}} \underbrace{P(\boldsymbol{\Lambda})}_{\text{事前分布}} d\boldsymbol{\lambda}\, d\boldsymbol{\Lambda}\, d\boldsymbol{c}\, d\boldsymbol{C}$$

むすび

音声合成における統計的アプローチ

- 音声合成の全要素を統計的な枠組みによって記述
- 統計的な問題として統一された視点
- データベースの重要性

今後の課題

- まだまだ解くべき問題はたくさんある:
  - 音声波形モデリング
  - テキスト解析部との統合
  - etc.

←

最後のメッセージ

音声合成は泥臭い問題か?

No!

皆様も音声合成研究に是非ご参加を！

ご清聴ありがとうございました

Acknowledgement

Keiichi Tokuda would like to thank HTS working group members, including Heiga Zen, Keiichiro Oura, Junichi Yamagichi, Tomoki Toda, Yoshihiko Nankaku, Kei Hahimoto, and Sayaka Shiota for their help.

References (1/4)

Sagisaka;'92 - "ATR nu-TALK speech synthesis system," ICSLP, '92.
Black;'96 - "Automatically clustering similar units...," Euro speech, '97.
Beutnagel;'99 - "The AT&T Next-Gen TTS system," Joint ASA, EAA, & DAEA meeting, '99.
Yoshimura;'99 - "Simultaneous modeling of spectrum ...," Eurospeech, '99.
Itakura;'70 - "A statistical method for estimation of speech spectral density...," Trans. IEICE, J53-A, '70.
Imai;'88 - "Unbiased estimator of log spectrum and its application to speech signal...," EURASIP, '88.
Kobayashi;'84 - "Spectral analysis using generalized cepstrum," IEEE Trans. ASSP, 32, '84.
Tokuda;'94 - "Mel-generalized cepstral analysis -- A unified approach to speech spectral...," ICSLP, '94.
Imai;'83 - "Cepstral analysis synthesis on the mel frequency scale," ICASSP, '83.
Fukada;'92 - "An adaptive algorithm for mel-cepstral analysis of speech," ICASSP, '92.
Itakura;'75 - "Line spectrum representation of linear predictive coefficients of speech...," J. ASA (57), '75.
Tokuda;'02 - "Multi-space probability distribution HMM," IEICE Trans. E85-D(3), '02.
Odell;'95 - "The use of context in large vocaburary...," PhD thesis, University of Cambridge, '95.
Shinoda;'00 - "MDL-based context-dependent subword modeling...," Journal of ASJ(E) 21(2), '00.
Yoshimura;'98 - "Duration modeling for HMM-based speech synthesis," ICSLP, '98.
Tokuda;'00 - "Speech parameter generation algorithms for HMM-based speech synthesis," ICASSP, '00.
Kobayashi;'85 - "Mel generalized-log spectrum approximation...," IEICE Trans. J68-A (6), '85.
Hunt;'96 - "Unit selection in a concatenative speech synthesis system using...," ICASSP, '96.
Donovan;'95 - "Improvements in an HMM-based speech synthesiser," Eurospeech, '95.
Kawai;'04 - "XIMERA: A new TTS from ATR based on corpus-based technologies," ISCA SSW5, '04.
Hirai;'04 - "Using 5 ms segments in concatenative speech synthesis," Proc. ISCA SSW5, '04.

References (2/4)

Rouibia;'05 - "Unit selection for speech synthesis based on a new acoustic target cost," Interspeech, '05.
Huang;'96 - "Whistler: A trainable text-to-speech system," ICSLP, '96.
Mizutani;'02 - "Concatenative speech synthesis based on HMM," ASJ autumn meeting, '02.
Ling;'07 - "The USTC and iFlytek speech synthesis systems...," Blizzard Challenge workshop, 07.
Ling;'08 - "Minimum unit selection error training for HMM-based unit selection...," ICASSP, 08.
Plumpe;'98 - "HMM-based smoothing for concatenative speech synthesis," ICSLP, '98.
Wouters;'00 - "Unit fusion for concatenative speech synthesis," ICSLP, '00.
Okubo;'06 - "Hybrid voice conversion of unit selection and generation...," IEICE Trans. E89-D(11), '06.
Aylett;'08 - "Combining statistical parametric speech synthesis and unit selection..." LangTech, '08.
Pollet;'08 - "Synthesis by generation and concatenation of multiform segments," Interspeech, '08.
Yamagishi;'06 - "Average-voice-based speech synthesis," PhD thesis, Tokyo Inst. of Tech., '06.
Yoshimura;'97 - "Speaker interpolation in HMM-based speech synthesis system," Eurospeech, '97.
Tachibana;'05 - "Speech synthesis with various emotional expressions...," IEICE Trans. E88-D(11), '05.
Kuhn;'00 - "Rapid speaker adaptation in eigenvoice space," IEEE Trans. SAP 8(6), '00.
Shichiri;'02 - "Eigenvoices for HMM-based speech synthesis," ICSLP, '02.
Fujinaga;'01 - "Multiple-regression hidden Markov model," ICASSP, '01.
Nose;'07 - "A style control technique for HMM-based expressive speech...," IEICE Trans. E90-D(9), '07.
Yoshimura;'01 - "Mixed excitation for HMM-based speech synthesis," Eurospeech, '01.
Kawahara;'97 - "Restructuring speech representations using a ...", Speech Communication, 27(3), '97.
Zen;'07 - "Details of the Nitech HMM-based speech synthesis system...", IEICE Trans. E90-D(1), '07.
Abdl-Hamid;'06 - "Improving Arabic HMM-based speech synthesis quality," Interspeech, '06.

References (3/4)

Hemptinne;'06 - "Integration of the harmonic plus noise model into the...," Master thesis, IDIAP, '06.
Banos;'08 - "Flexible harmonic/stochastic modeling...," V. Jornadas en Tecnologias del Habla, '08.
Cabral;'07 - "Towards an improved modeling of the glottal source in...," ISCA SSW6, '07.
Maia;'07 - "An excitation model for HMM-based speech synthesis based on ...," ISCA SSW6, '07.
Ratio;'08 - "HMM-based Finnish text-to-speech system utilizing glottal inverse filtering," Interspeech, '08.
Drugman;'09 - "Using a pitch-synchronous residual codebook for hybrid HMM/frame...", ICASSP, '09.
Dines;'01 - "Trainable speech synthesis with trended hidden Markov models," ICASSP, '01.
Sun;'09 - "Polynomial segment model based statistical parametric speech synthesis...," ICASSP, '09.
Bulyko;'02 - "Robust splicing costs and efficient search with BMM models for...," ICASSP, '02.
Shannon;'09 - "Autoregressive HMMs for speech synthesis," Interspeech, '09.
Zen;'06 - "Reformulating the HMM as a trajectory model...", Computer Speech & Language, 21(1), '06.
Wu;'06 - "Minimum generation error training for HMM-based speech synthesis," ICASSP, '06.
Hashimoto;'09 - "A Bayesian approach to HMM-based speech synthesis," ICASSP, '09.
Wu;'08 - "Minimum generation error training with log spectral distortion for...," Interspeech, '08.
Toda;'08 - "Statistical approach to vocal tract transfer function estimation based on...," ICASSP, '08.
Oura;'08 - "Simultaneous acoustic, prosodic, and phrasing model training for TTS...," ISCSLP, '08.
Ferguson;'80 - "Variable duration models...," Symposium on the application of HMM to text speech, '80.
Levinson;'86 - "Continuously variable duration hidden...," Computer Speech & Language, 1(1), '86.
Beal;'03 - "Variational algorithms for approximate Bayesian inference," PhD thesis, Univ. of London, '03.
Masuko;'03 - "A study on conditional parameter generation from HMM...," Autumn meeting of ASJ, '03.
Yu;'07 - "A novel HMM-based TTS system using both continuous HMMs and discrete...," ICASSP, '07.

References (4/4)

Qian;'08 - "Generating natural F0 trajectory with additive trees," Interspeech, '08.
Latorre;'08 - "Multilevel parametric-base F0 model for speech synthesis," Interspeech, '08.
Tiomkin;'08 - "Statistical text-to-speech synthesis with improved dynamics," Interspeech, '08.
Toda;'07 - "A speech parameter generation algorithm considering global...," IEICE Trans. E90-D(5), '07.
Wu;'08 - "Minimum generation error criterion considering global/local variance...," ICASSP, '08.
Toda;'09 - "Trajectory training considering global variance for HMM-based speech...," ICASSP, '09.
Saino;'06 - "An HMM-based singing voice synthesis system," Interspeech, '06.
Tsuzuki;'04 - "Constructing emotional speech synthesizers with limited speech...," Interspeech, '04.
Sako;'00 - "HMM-based text-to-audio-visual speech synthesis," ICSLP, '00.
Tamura;'98 - "Text-to-audio-visual speech synthesis based on parameter generation...," ICASSP, '98.
Haoka;'02 - "HMM-based synthesis of hand-gesture animation," IEICE Technical report,102(517), '02.
Niwase;'05 - "Human walking motion synthesis with desired pace and...," IEICE Trans. E88-D(11), '05.
Hofer;'07 - "Speech driven head motion synthesis based on a trajectory model," SIGGRAPH, '07.
Ma;'07 - "A MSD-HMM approach to pen trajectory modeling for online handwriting...," ICDAR, '07.
Morioka;'04 - "Miniaturization of HMM-based speech synthesis," Autumn meeting of ASJ, '04.
Kim;'06 - "HMM-Based Korean speech synthesis system for...," IEEE Trans. Consumer Elec., 52(4), '06.
Klatt;'82 - "The Klatt-Talk text-to-speech system," ICASSP, '82.

Copyright for some samples

Some samples (icon is 🔊) are released under following license.