

音声から情報を得る —大量の音声データから見えてくる ものは？—

板橋 秀一

国立情報学研究所（NII）
産業技術総合研究所（AIST）

目次

1. 音と音声
2. 音声研究：合成・認識
3. 音声コーパスとは
4. 大規模コーパスから見えること
5. 音声コーパスの利用
6. 音声コーパス類似性の可視化
7. 音響データベース
8. まとめ

音と音声

音は媒質の振動

単純な音→正弦波(サイン波)→音叉

普通の音は複合波



正弦波の重ね合わせ

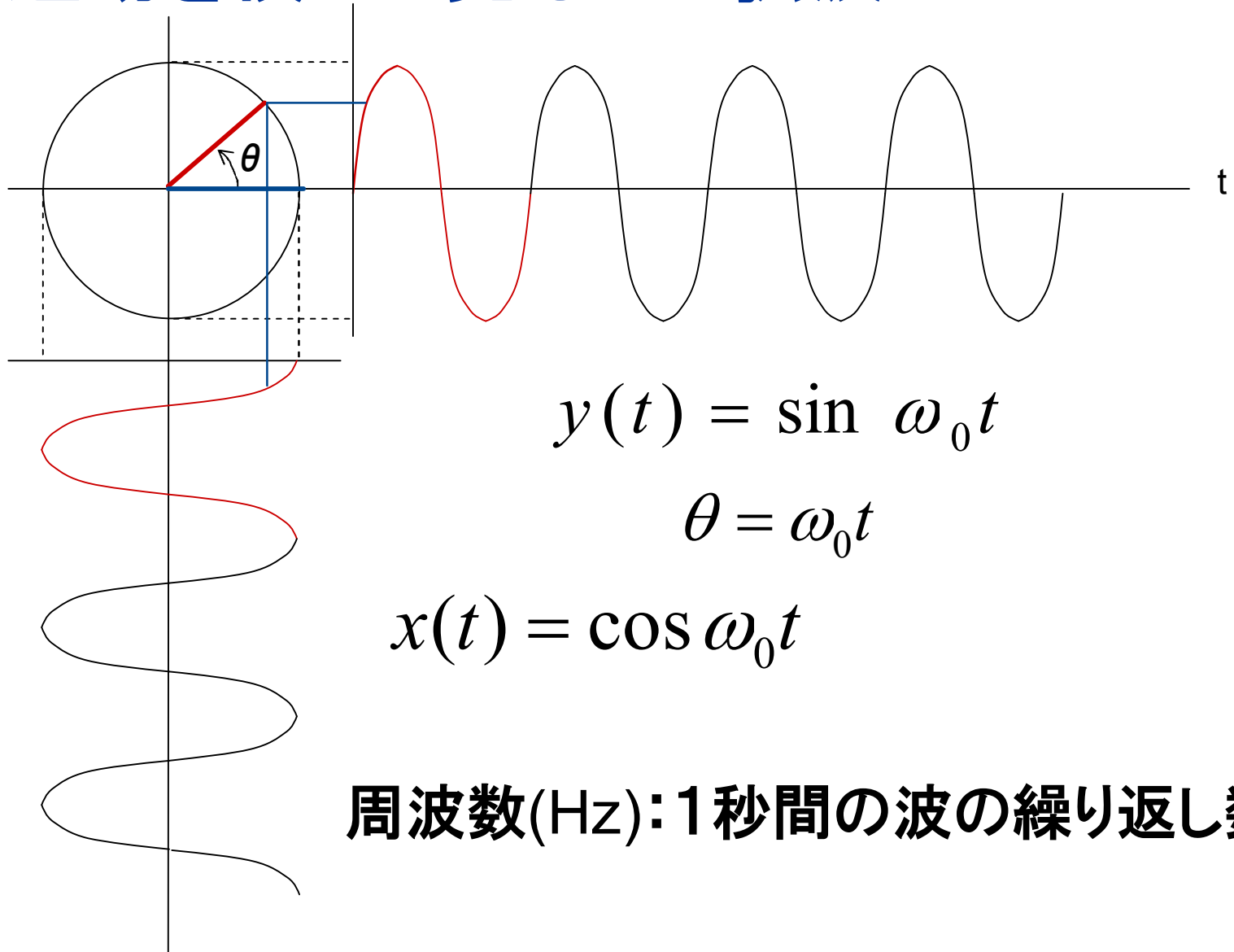
音声の源→声帯



ブザーのような音

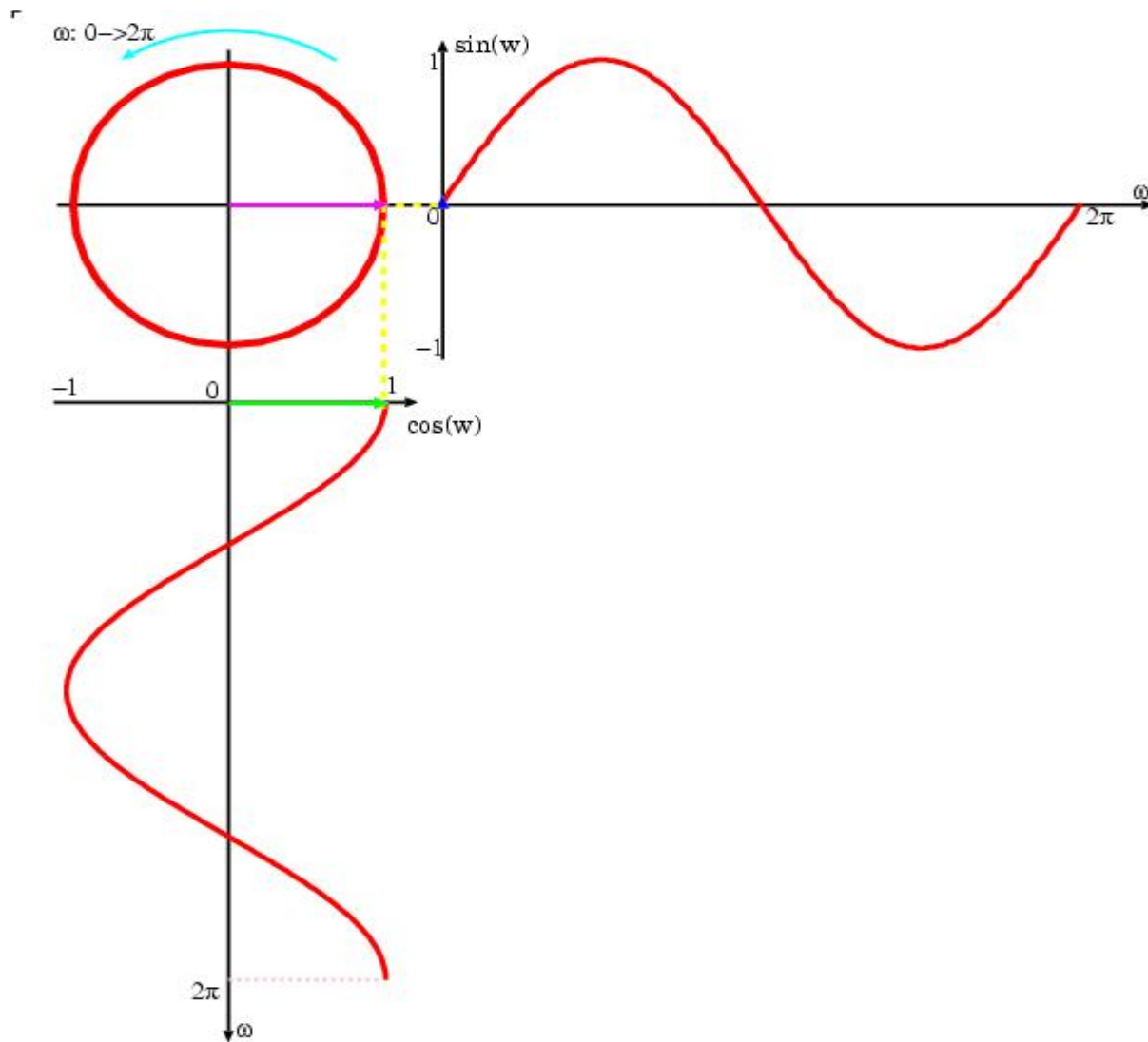


円運動を横から見ると正弦波



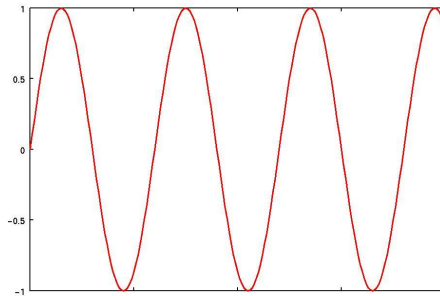
周波数(Hz): 1秒間の波の繰り返し数

正弦波(sine)と余弦波(cosine)



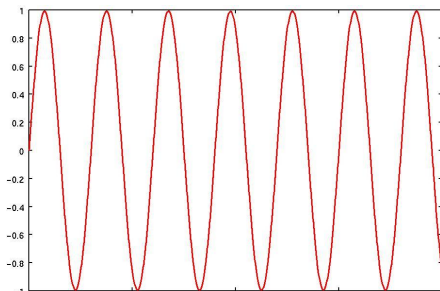
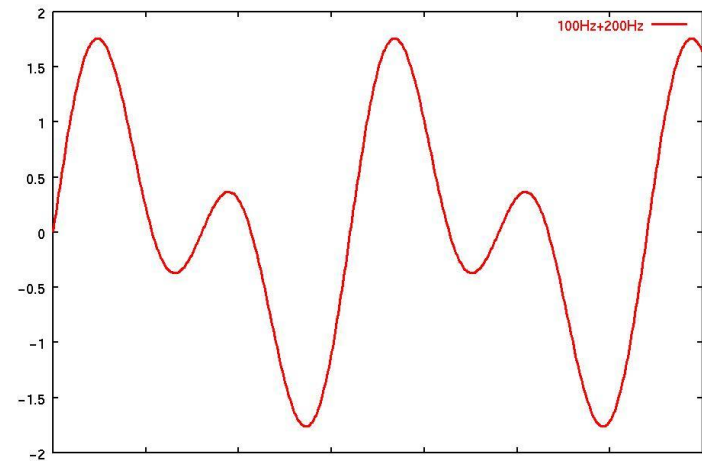
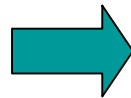
正弦波(サイン波)

Sampling : 6000Hz Bit: 16-bit



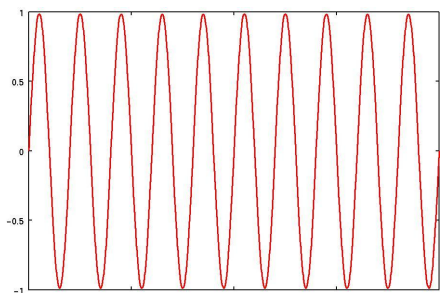
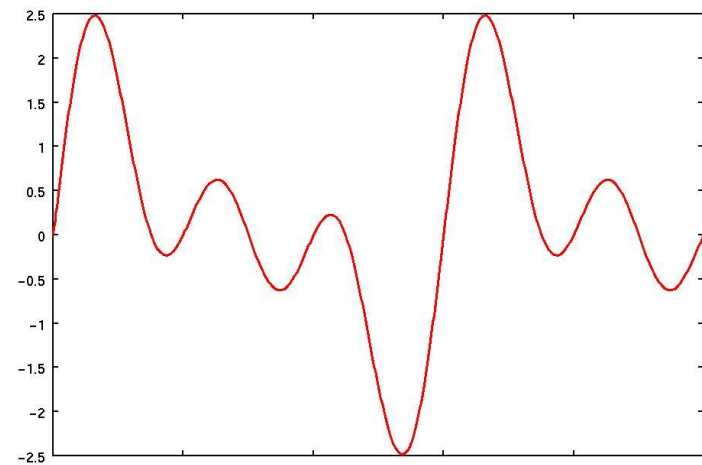
100Hz

100Hz+200Hz



200Hz

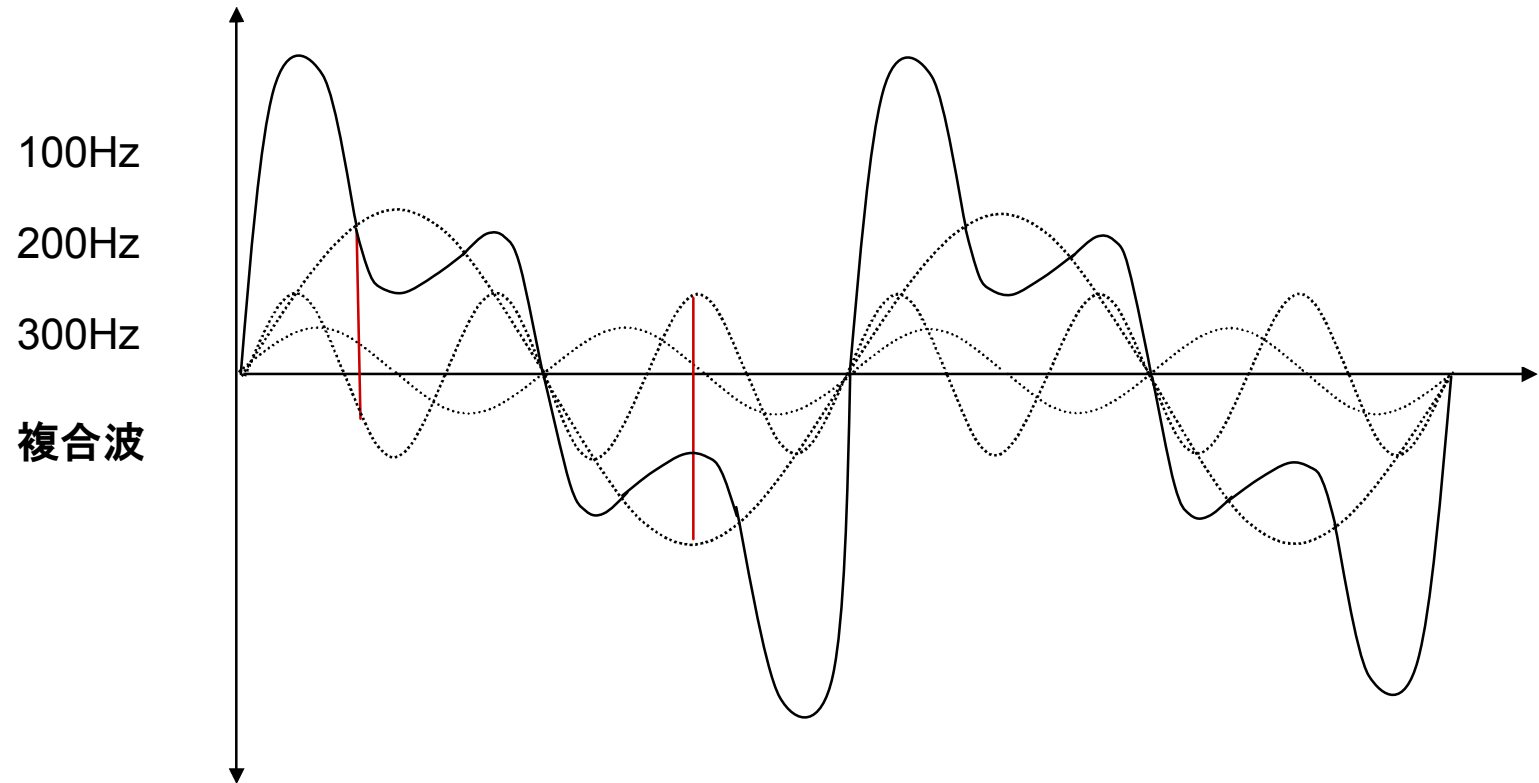
100Hz+200Hz+300Hz



300Hz

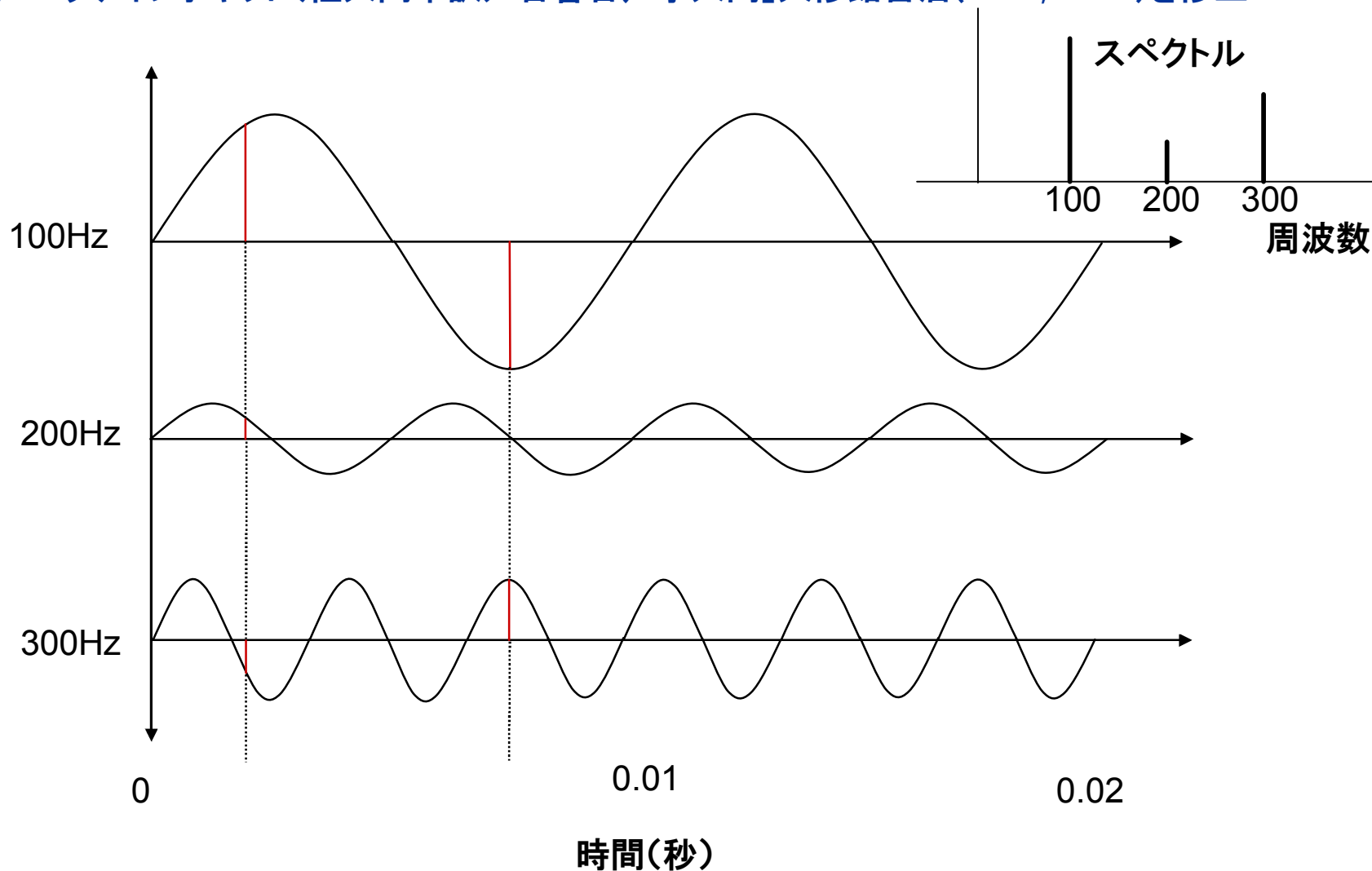
100, 200, 300Hzの正弦波の合成

出典:P. ラディフォギッド(佐久間章訳)「音響音声学入門」大修館書店(1976, 1982)を修正

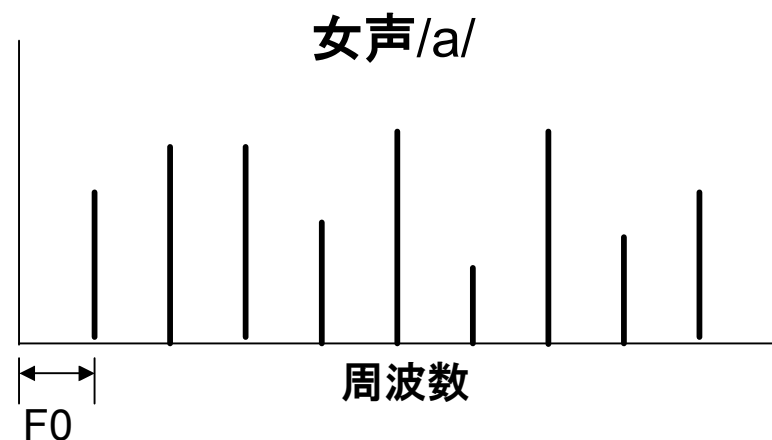
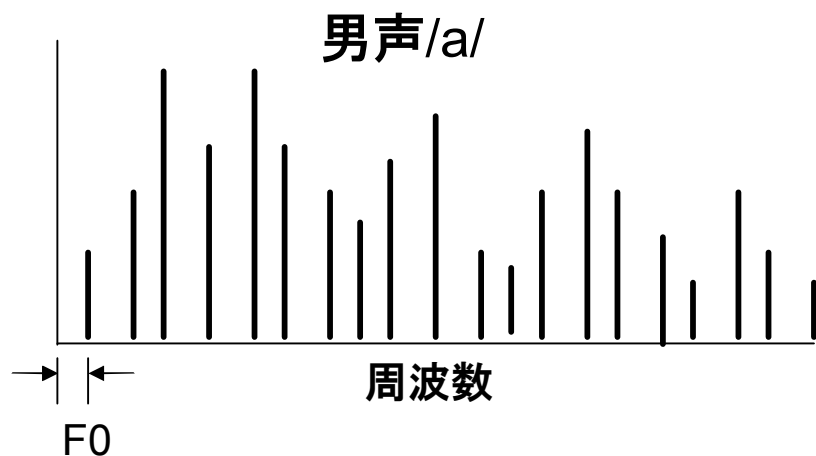


100, 200, 300Hzの複合波の分析

出典:P. ラディフォギッド(佐久間章訳)「音響音声学入門」大修館書店(1976, 1982)を修正



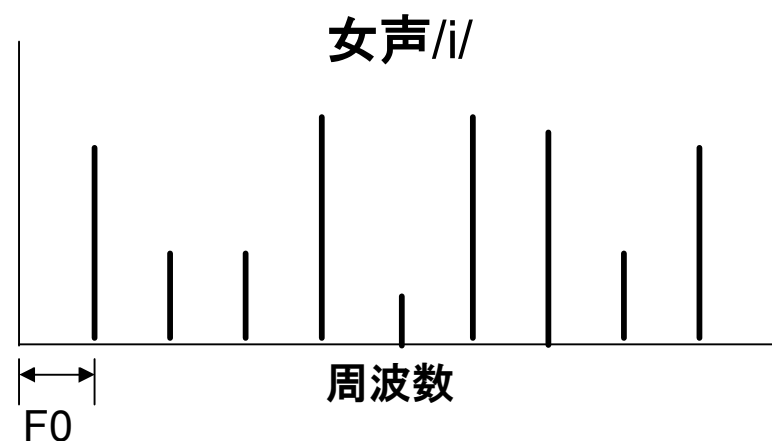
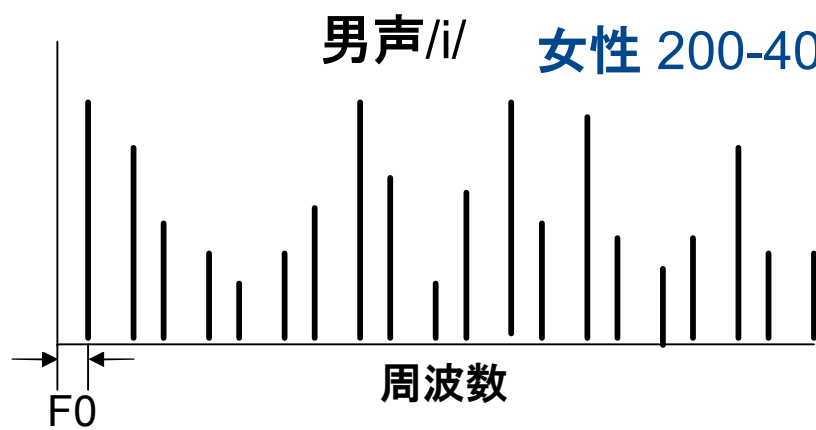
音声のスペクトルと基本周波数(F0)



基本周波数(声の高さ)

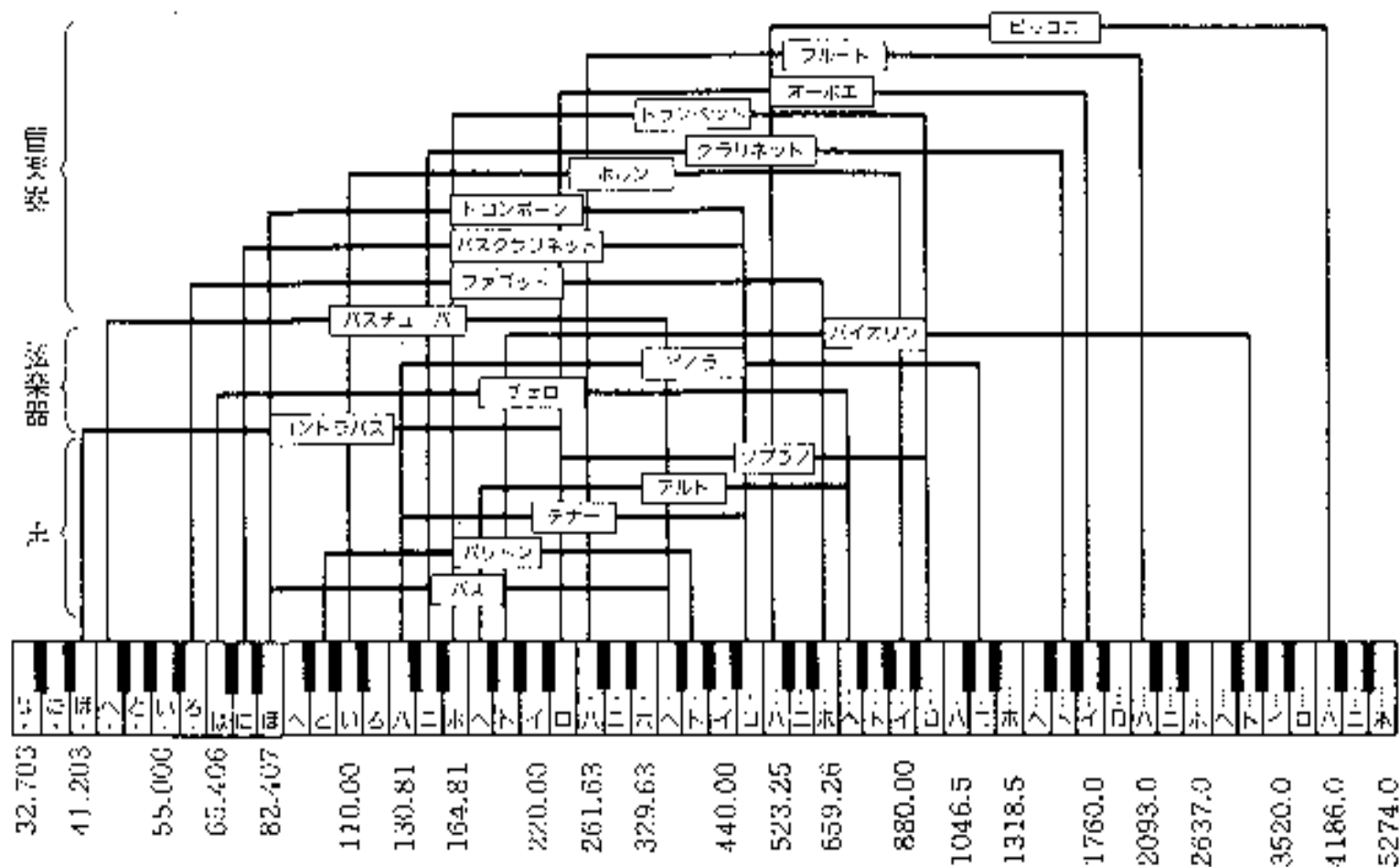
男声 100-200 Hz

女性 200-400 Hz



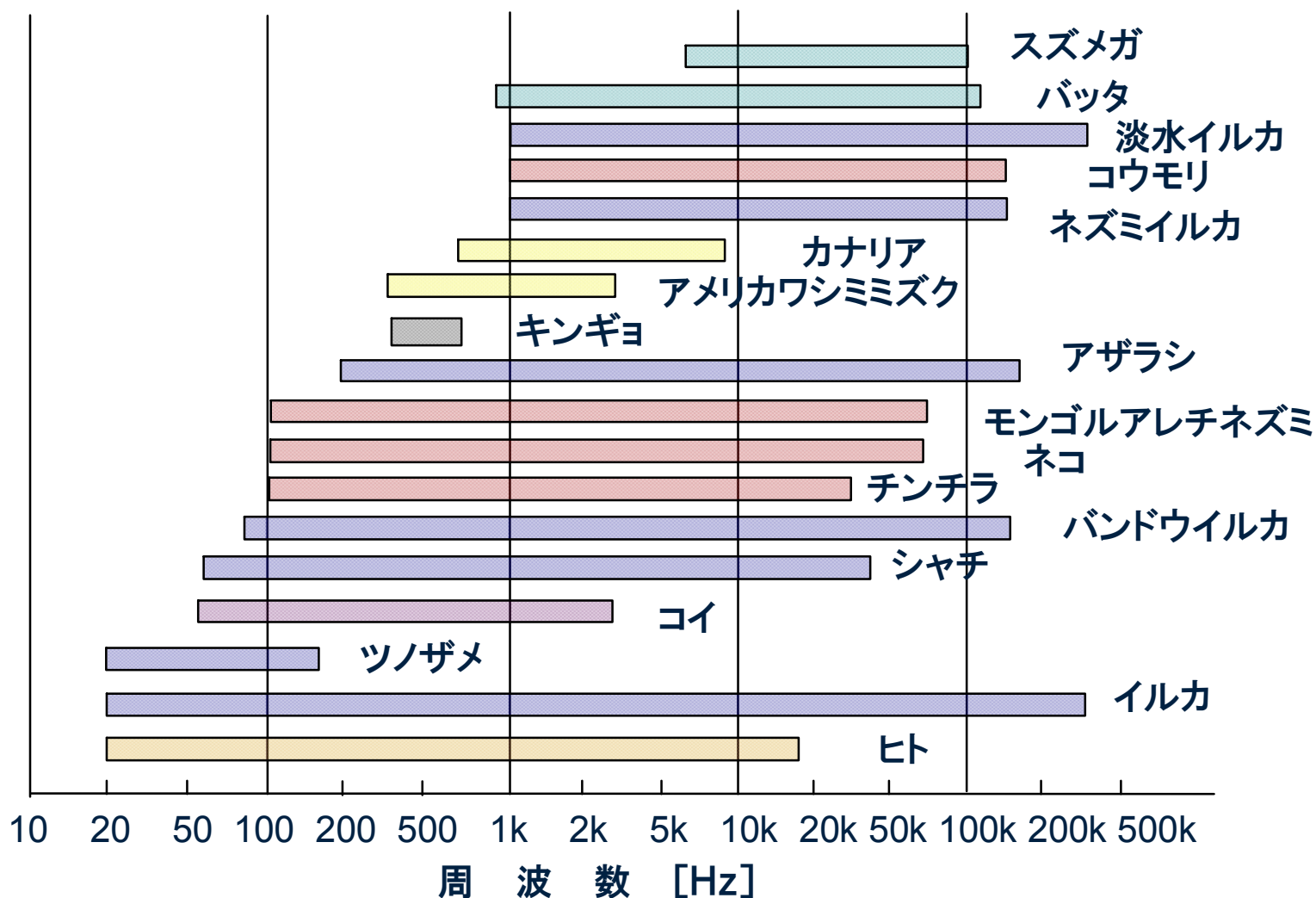
声と楽器

(出典: 曾根敏夫「くらしと音」裳華房(1991))



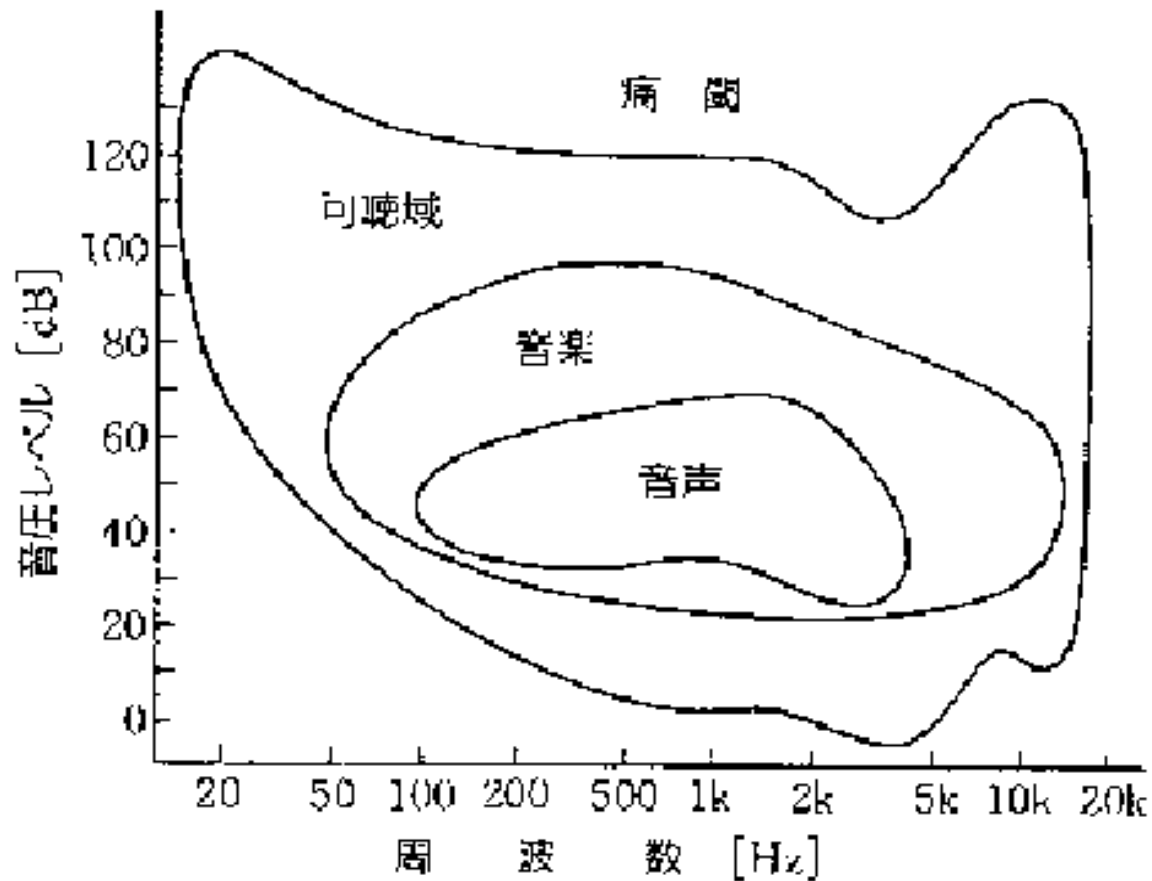
動物の可聴域

(出典: 曾根敏夫「くらしと音」裳華房(1991)に着色)

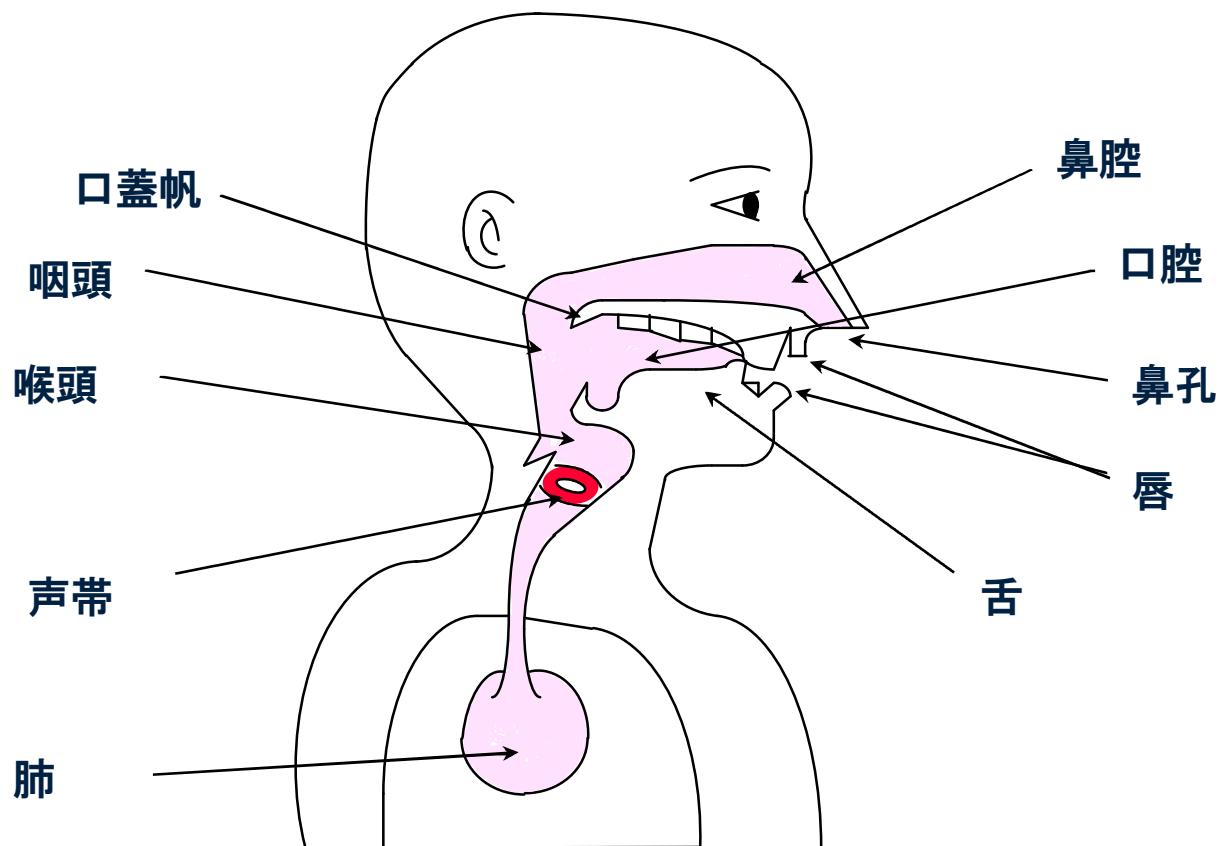


人の可聴域

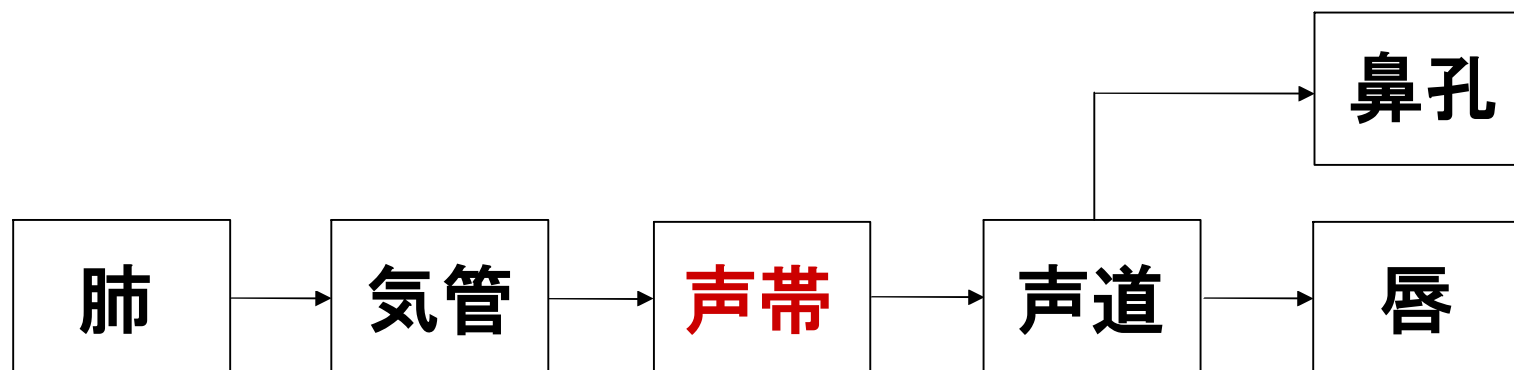
(出典: 曾根敏夫「くらしと音」裳華房(1991))



音声器官の模式図



音声の生成過程

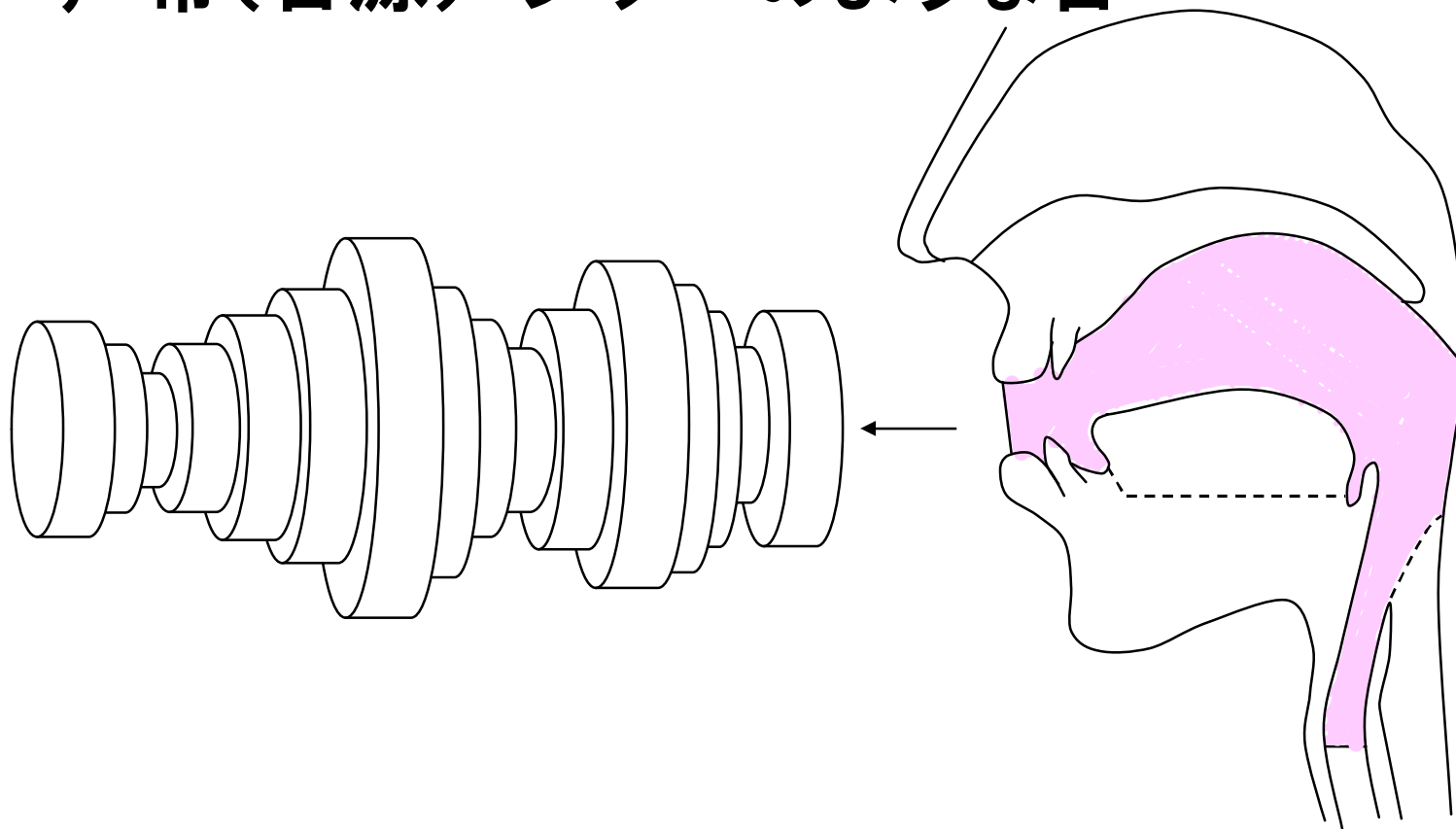


音素:ローマ字1文字 /a, k, s, t, n, h, m, y, r, w/

音節:かな1文字 あ/a/, け/ke/, さ/sa/, と/to/,

声道模型

- ・声道：声帯から唇までの音の通路
- ・声帯(音源)：ブザーのような音

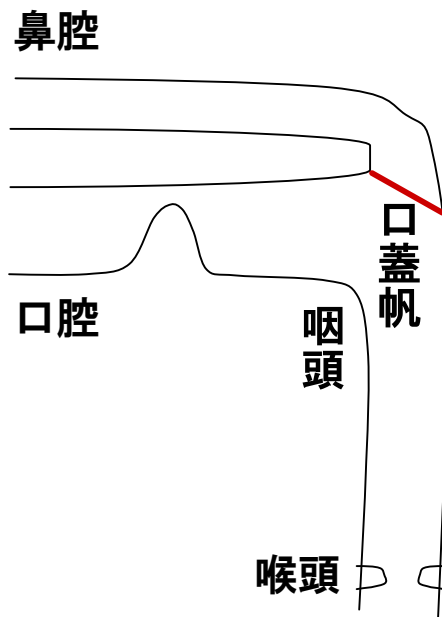


母音・子音・鼻音の生成と声道



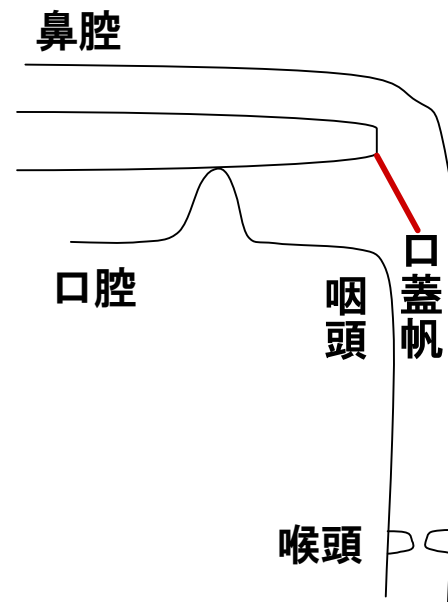
母音

/a, i, u, e, o/



子音

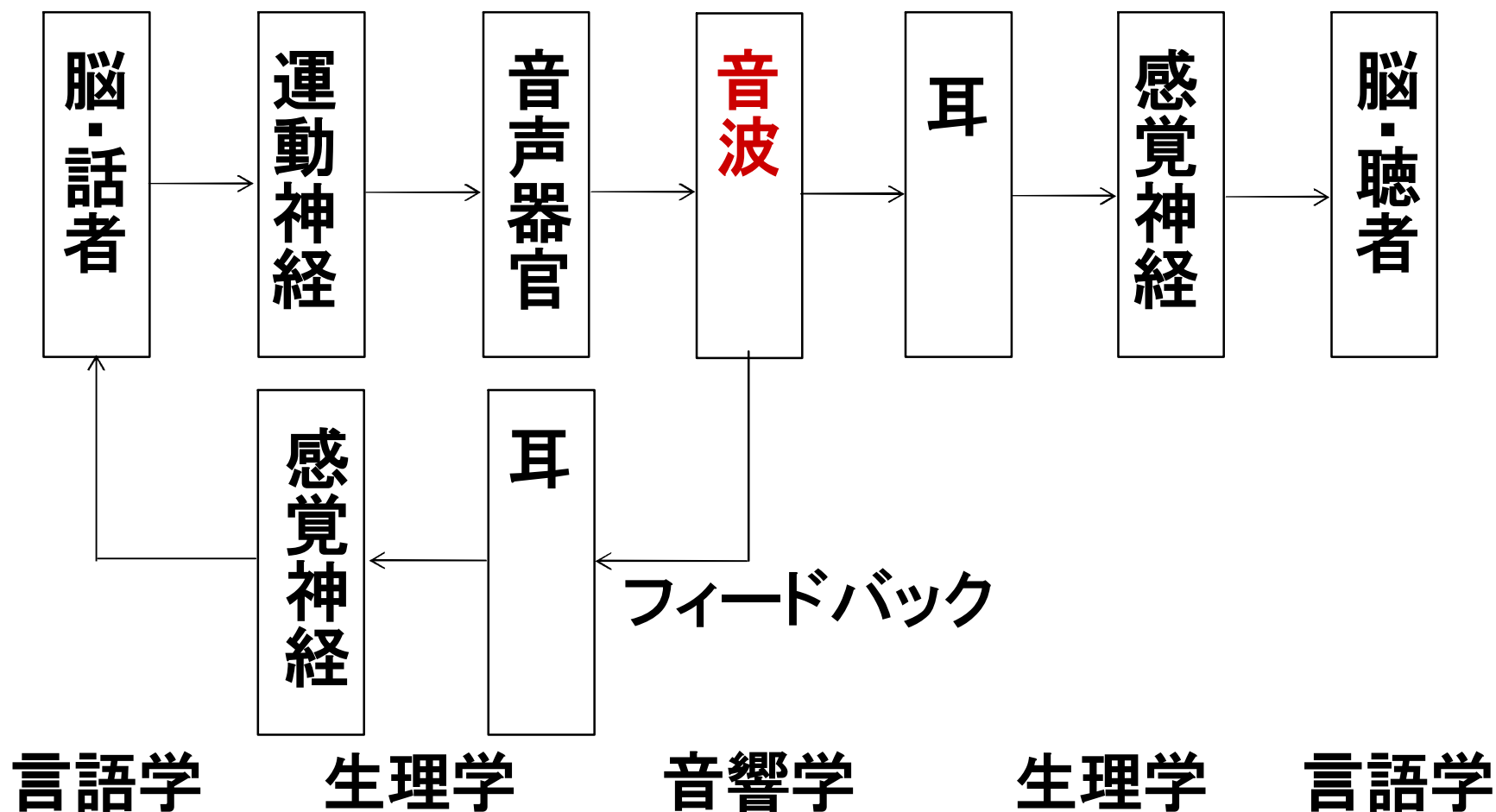
/p, t, k, b, d, g, s, z, r, h/



鼻音

/m, n, η , N/

音声の生成と知覚（ことばの鎖）



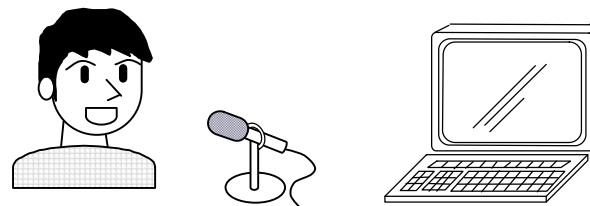
音声研究

音声情報処理

音声自動認識

音声分析・合成

言語・話者認識

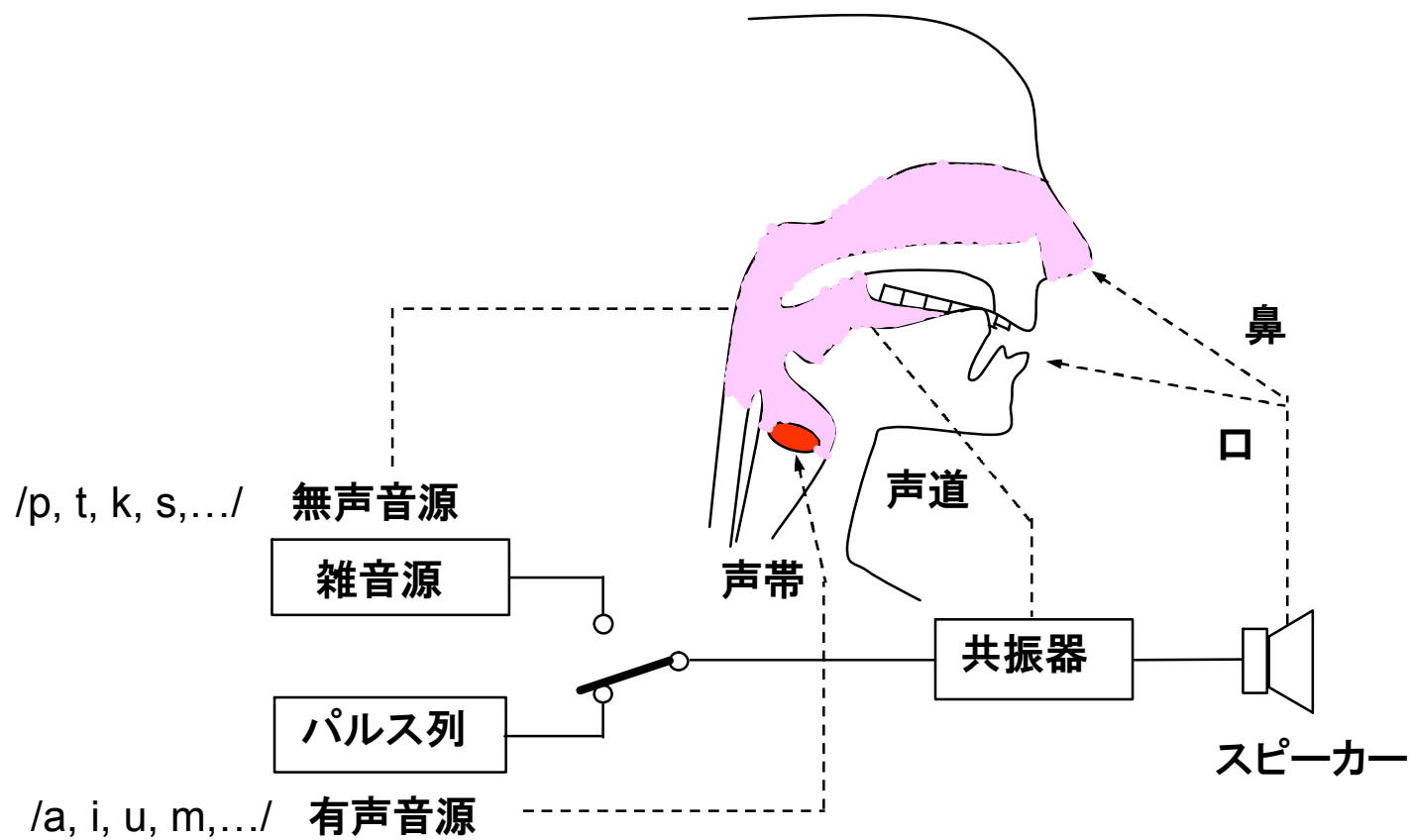


言語学・音声学

音声分析/特徴抽出

音声言語データの保存: 危機言語

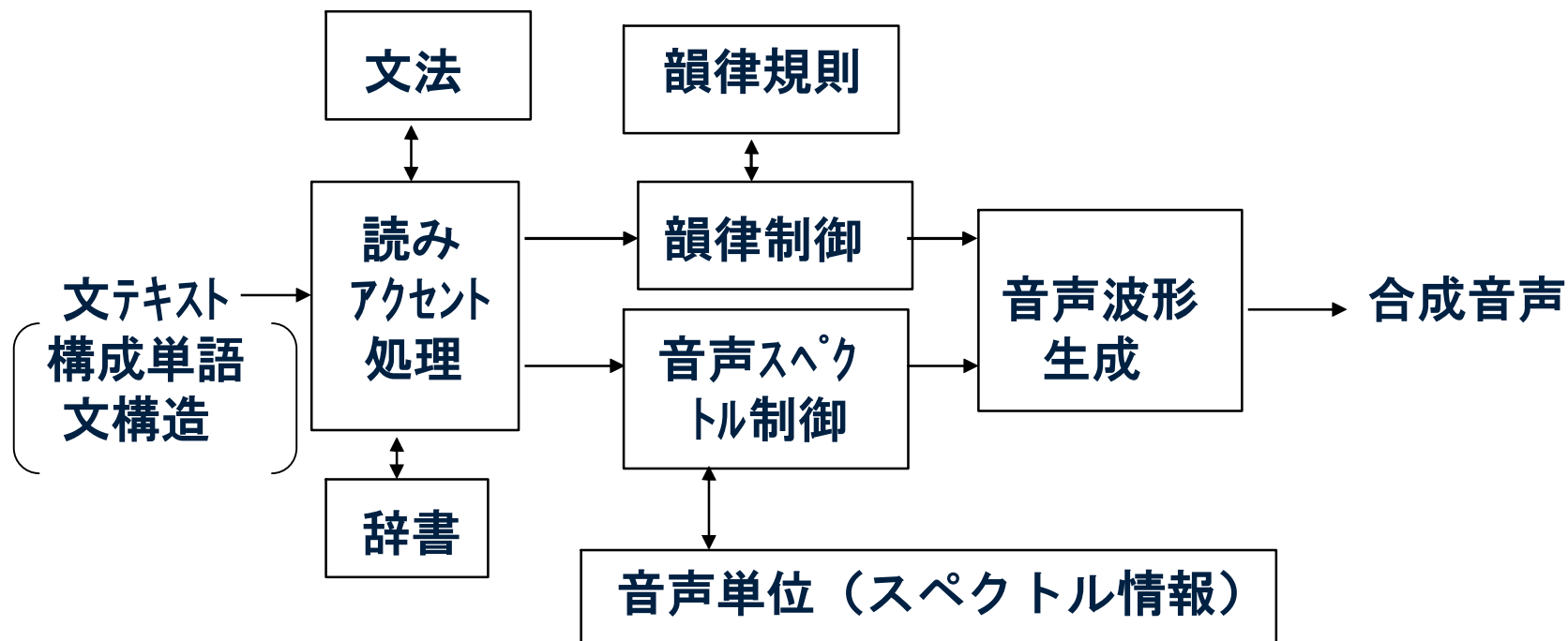
音声合成の原理



音声合成の難しさ

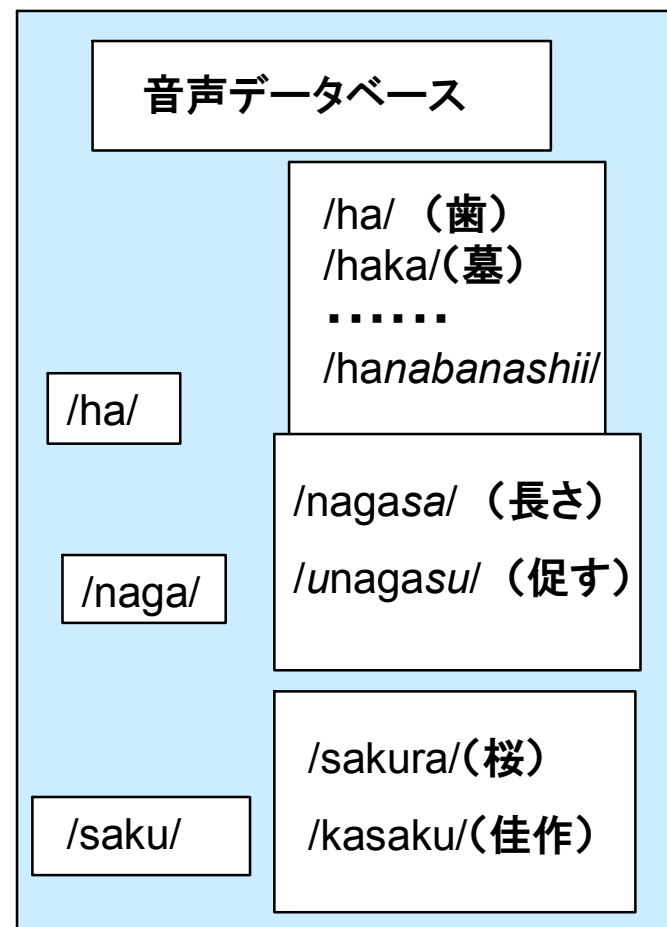
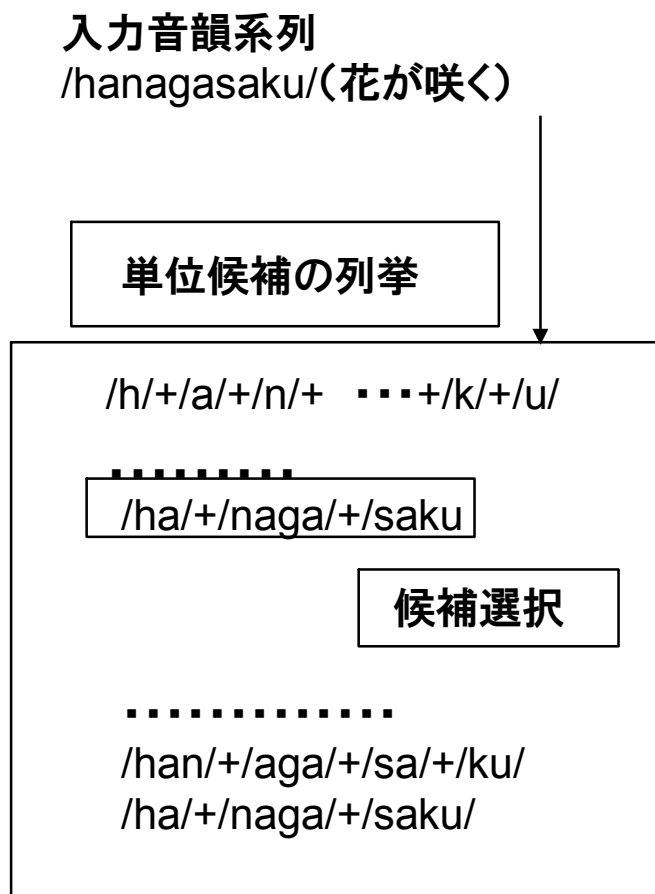
- ・任意の人の声
- ・調音結合(音声のつながり)
/a/: /**a**oi/, /**a**ida/
- ・多言語・方言
- ・快適な音声

音声合成システム



出典: 匂坂芳典「コーパスベース音声合成」J. Signal Processing 信号処理, 2 (6) 407-414 (1998)

コーパスベース音声合成



音声スペクトル

パラメータ時系列

出典: 匂坂芳典「コーパスベース音声合成」J. Signal Processing 信号処理, 2 (6) 407-414 (1998)

合成音声の例

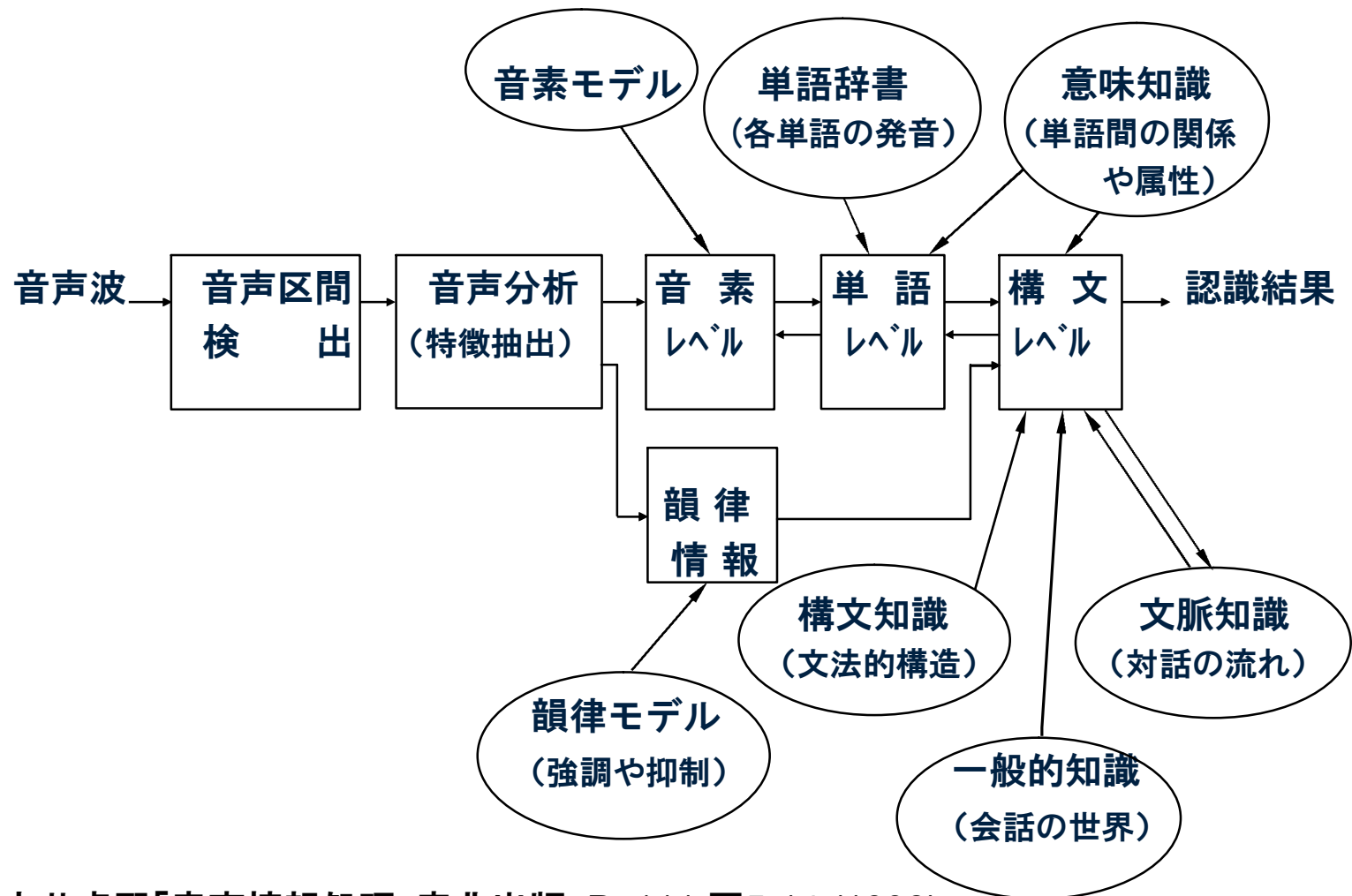
ストックホルム王立工科大学 (1977)
日経エレクトロニクス (1984)
JEIDA 最近の規則合成音声 (1996)

JEIDA: (社)日本電子工業振興協会(現在 電子情報技術産業協会)

音声認識の難しさ

- ・話者変動
- ・時間変動
- ・調音結合(音声のつながり)
- ・背景雑音
- ・言語・方言

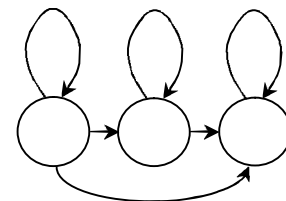
文音声認識システムの基本的構成



出典: 古井貞熙「音声情報処理」森北出版, P. 111 図5-14 (1998)

音声情報処理における統計的手法

認識: 隠れマルコフモデル (HMM)
言語モデル: 2字組, 3字組



合成: コーパスベース音声合成

多量の音声データが必要

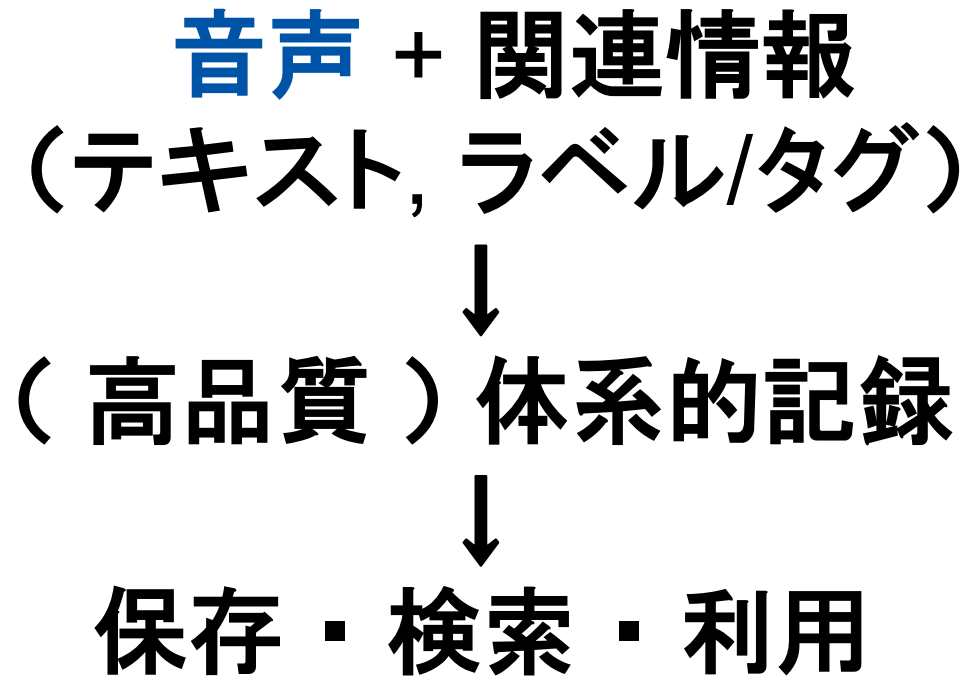
今、なぜ、音声コーパスなのか？

- 音声情報処理: 認識・分析・合成等
- 世界の諸言語の音声研究・方言研究
- 計量的な言語研究

各種・大量の音声データが必要

- 単語音声・連続音声・多言語・方言など
各種コーパス(データベース)の作成
- 危機的言語の保存

音声コーパスとは



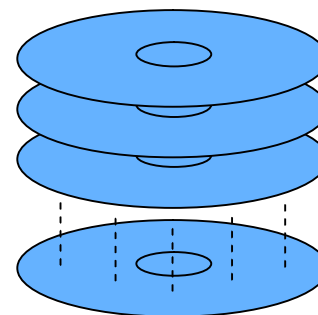
音声コーパス

各種大量の音声データの体系的蓄積
共通利用・公開を前提とする.

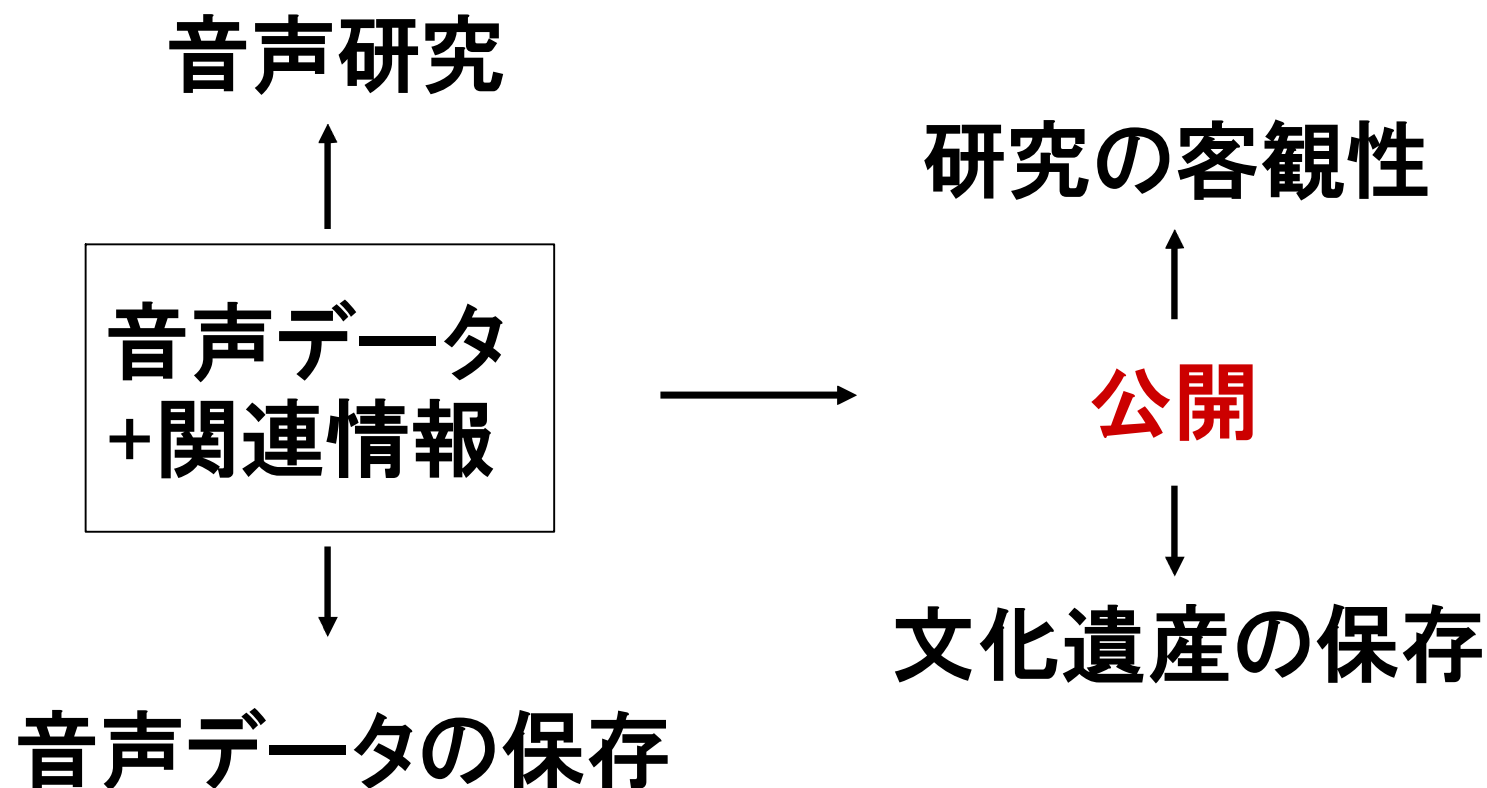
データベース: データを検索するデータ
ベースシステムを指すことが多く, デー
タの集積そのものを指すために「**コーパ
ス**」が使われるようになった.

音声コーパスの要求条件

1. 多様性
 2. 不偏性
 3. 多量性
 4. 使い易さ
- 統一性



音声コーパスの必要性



世界の言語の現状

世界中で約7,000の言語

世界の人口の94%の人が世界の言語の5%の言語を用いている。

残りの95%の言語は6%の人が用いている。

音声コーパス

原則: 各種大量のデータ

公開・共通利用

目的:

- ・研究開発: より良い手法の探索
- ・ユーザー: 認識装置の性能評価

構成:

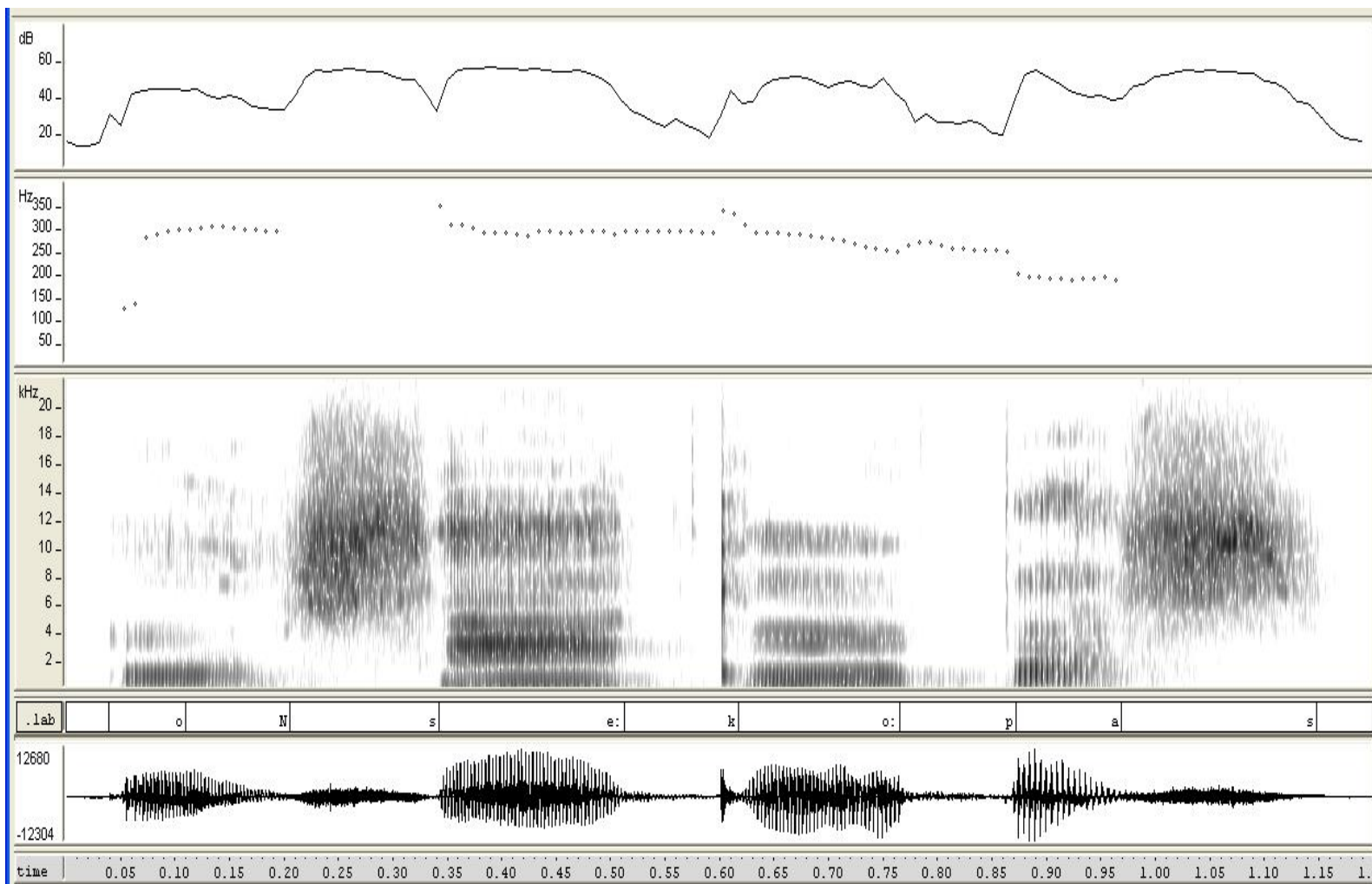
- ・音声データ (分析パラメータ)
- ・関連情報 (ラベル・タグ)

対象:

- ・音声分析・合成・認識
- ・話者・言語認識

音声波形, ラベル, スペクトログラム, Fo, パワー

女声「音声コーパス」/onse:ko:pas(u)/



日本語話し言葉コーパス(CSJ) (1999-2003)



大規模コーパス(CSJ)から見えること

言う: ゆう 99%; **いう** (13万項目)

Main: めいん 98%; **めーん**

日本: にほん 98%; **につぽん** (8千項目)

感ずる: かんじる 97%; **かんずる**

Simulation: しゅみれーしょん 97%; **しみゅれーしょん**

出典: K. Maekawa, "Quantitative analysis of word-form variation using a spontaneous speech corpus," Proc. Corpus Linguistics 2005, Birmingham (Jul. 2005).

大規模コーパス(CSJ)から見えること

NHK: えぬえちけー 96%; えぬえいちけー

体育: たいく 96%; たいいく

生ずる: しょうじる 94%; しょうずる

ぽい: っぽい 94%; ぽい

良い: いー 91%; よい (4千項目)

出典: K. Maekawa, "Quantitative analysis of word-form variation using a spontaneous speech corpus," Proc. Corpus Linguistics 2005, Birmingham (Jul. 2005).

音声コーパスの例

「桃太郎の話」
NHKアナウンサー
青森方言
鹿児島方言
沖縄方言



出典: 文部省科研費重点領域研究「日本語音声」代表 杉藤
美代子 大阪樟蔭女子大学教授

ヘリウム音声(10気圧)

提供: 郵政省電波研究所(現在 情報通信研究機構)

音声コーパスにおける問題点

- プロジェクト用・実験用に作成
- 非公開
- 作成の人的・経済的負担

音声データの作成・収集・蓄積・配布・共有のための
共通の枠組みが必要

各種音声データの収集・管理・配布を行う機関

音声資源コンソーシアム(NII-SRC)

世界の動き

- 1992年 LDC(言語データコンソーシアム)
- 1995年 ELRA(ヨーロッパ言語資源協会)
- 2001年 SITEC(韓国音声情報技術産業振興センター)
- 2002年 CCC(中国コーパスコンソーシアム)
- 2004年 Chinese LDC(中国LDC)

日本の動き

- 1999年 言語資源協会(GSK):テキストに重点
NPOとして組織変更(2003)
- 2006年 NII-SRC(音声資源コンソーシアム)

音声資源コンソーシアム

国立情報学研究所(NII)は日本の情報学のセンターとして、情報メディア特に音声メディアの未来価値創成に向けて**音声資源コンソーシアム(SRC)**を設置。

音声資源コンソーシアム(NII-SRC)

音声コーパスの収集・配付・研究事業

情報メディア特に音声メディアの未来価値創成に向けて
国立情報学研究所内に設置

2006年6月サービス開始

<http://research.nii.ac.jp/src/>

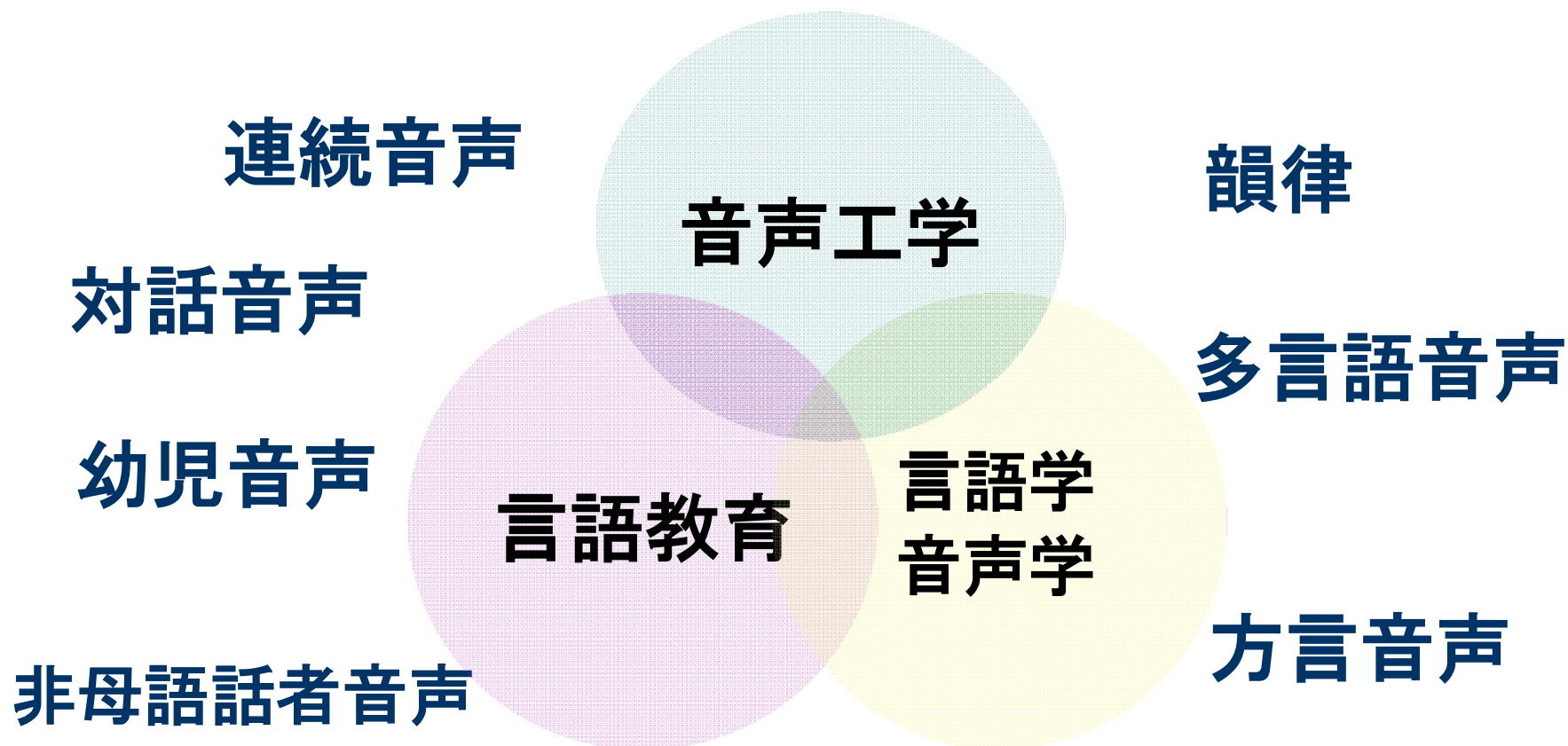
音声資源コンソーシアム

■ 音声資源の構築・配布・普及の促進

- 音声コーパスの内容・所在・利用方法に関する情報の収集・提供
- 未公開音声コーパスの提供依頼
- 既存の音声コーパスの配布・普及・広報活動
- 標準的契約書の作成(提供者—SRC—利用者)
- 音声コーパスに関する調査・研究

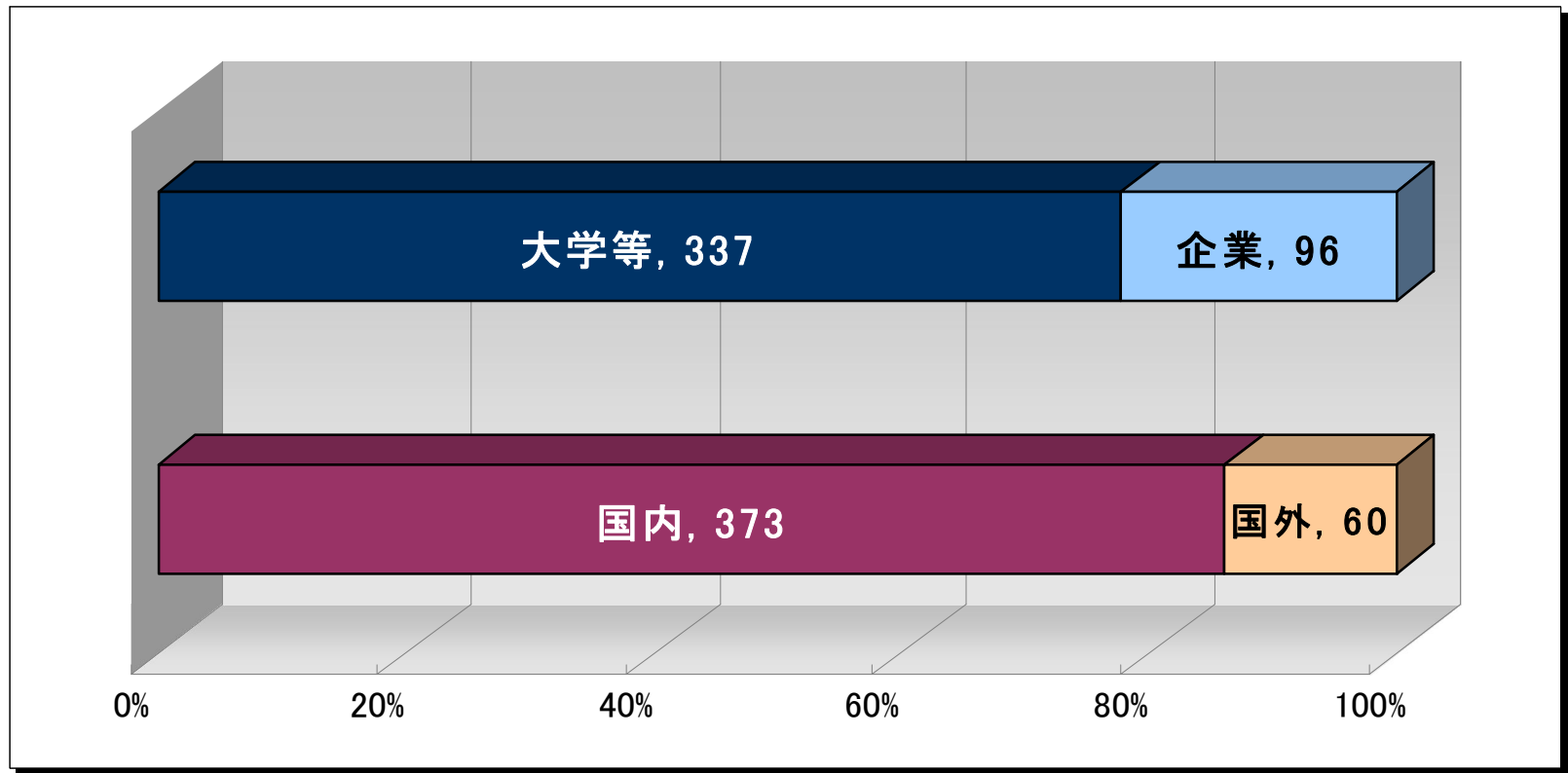
取り扱いコーパス

- 現在 31種類の音声コーパスをSRCより配布



コーパス配布状況 (2006.9~2009.3)

申込み件数：433件 ／ 配布コーパス数：1046



複数音声コーパス可視化研究の目的

現行のコーパスリスト

- テキストによる取り扱いコーパス一覧
- コーパスの基本情報(機関によって異なる)

膨大な量のコーパスリストから目的に合ったコーパスを選択する事は困難

コーパス検索の簡便化

- ・ コーパス特徴を表す属性項目による分類
- ・ 音声コーパス間の関連性, 類似性の可視化

コーパス特徴分類

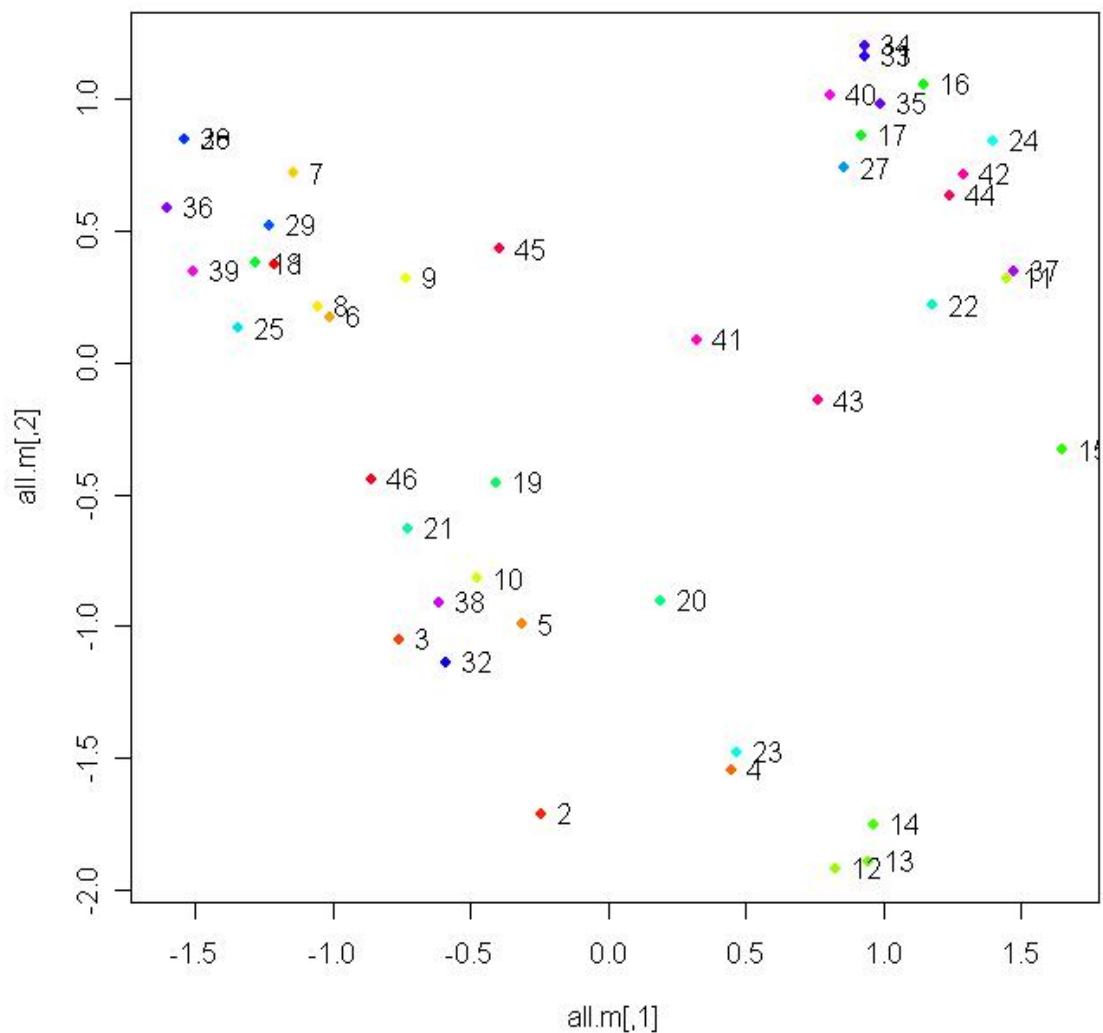
8属性58項目による分類

属性名	項目数	内容
入力デバイス	7項目	入力デバイスの種類
入力環境	5項目	収録環境
話者数	10項目	全体, 男女別
発話スタイル	4項目	連続／孤立音声
発話モード	5項目	対話／読み上げなど
データモード	9項目	標本化周波数など
言語	4項目	単一／多言語など
目的	14項目	構築目的, 利用目的

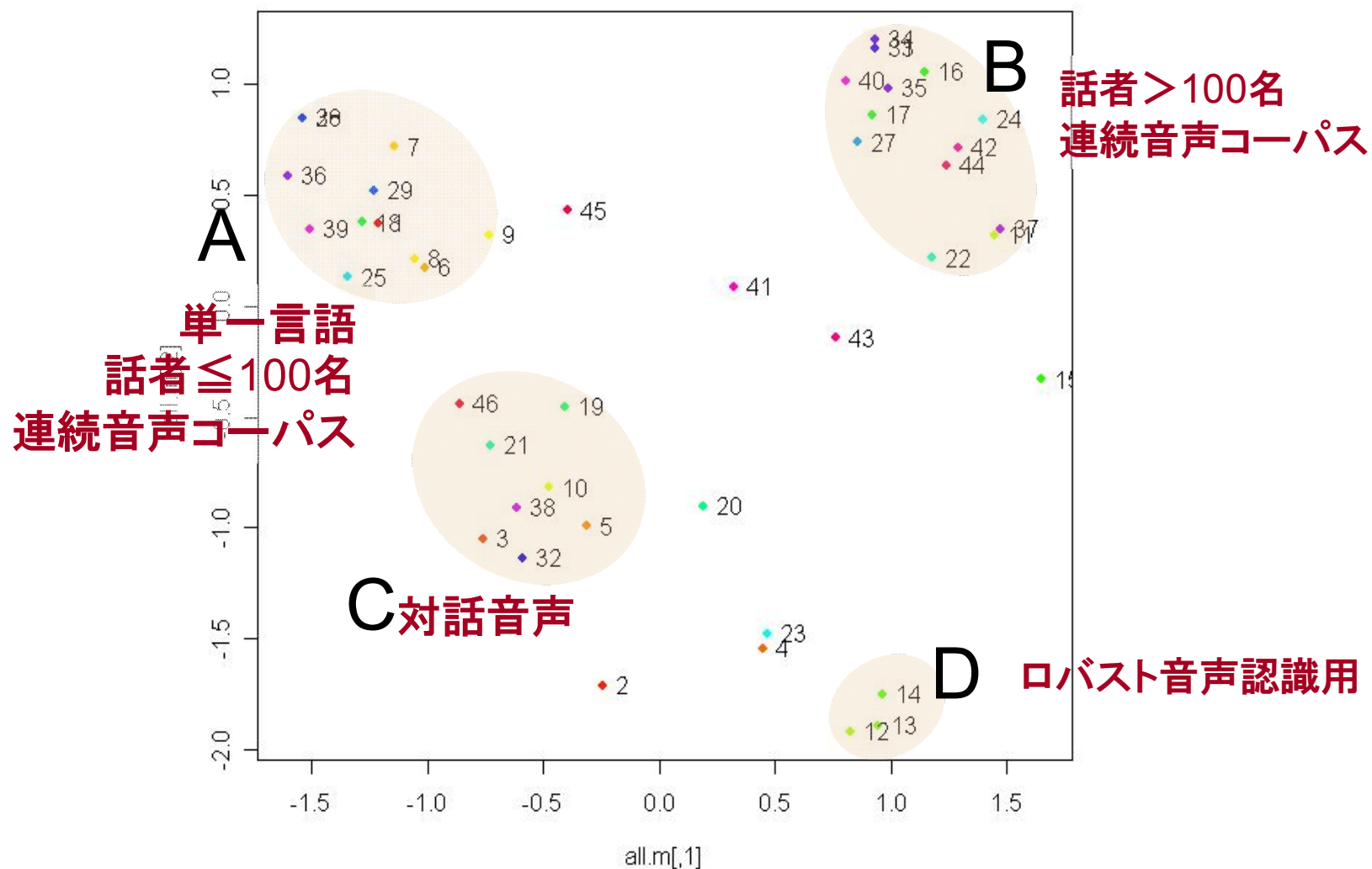
分析方法

- コーパス特徴属性に多次元尺度構成法(MDS)を適用して, コーパスの空間配置を導出
 - ・ コーパス特徴属性を58次元のベクトルで表現
 - ・ 各コーパスの特徴ベクトルを入力としたMDSによる解析
- 分析対象コーパス
音声コーパス46種類
 - ・ SRCで配布している音声コーパス 23種類
 - ・ 国内で配布されている音声コーパス 16種類
 - ・ 国外で配布されている音声コーパス 7種類

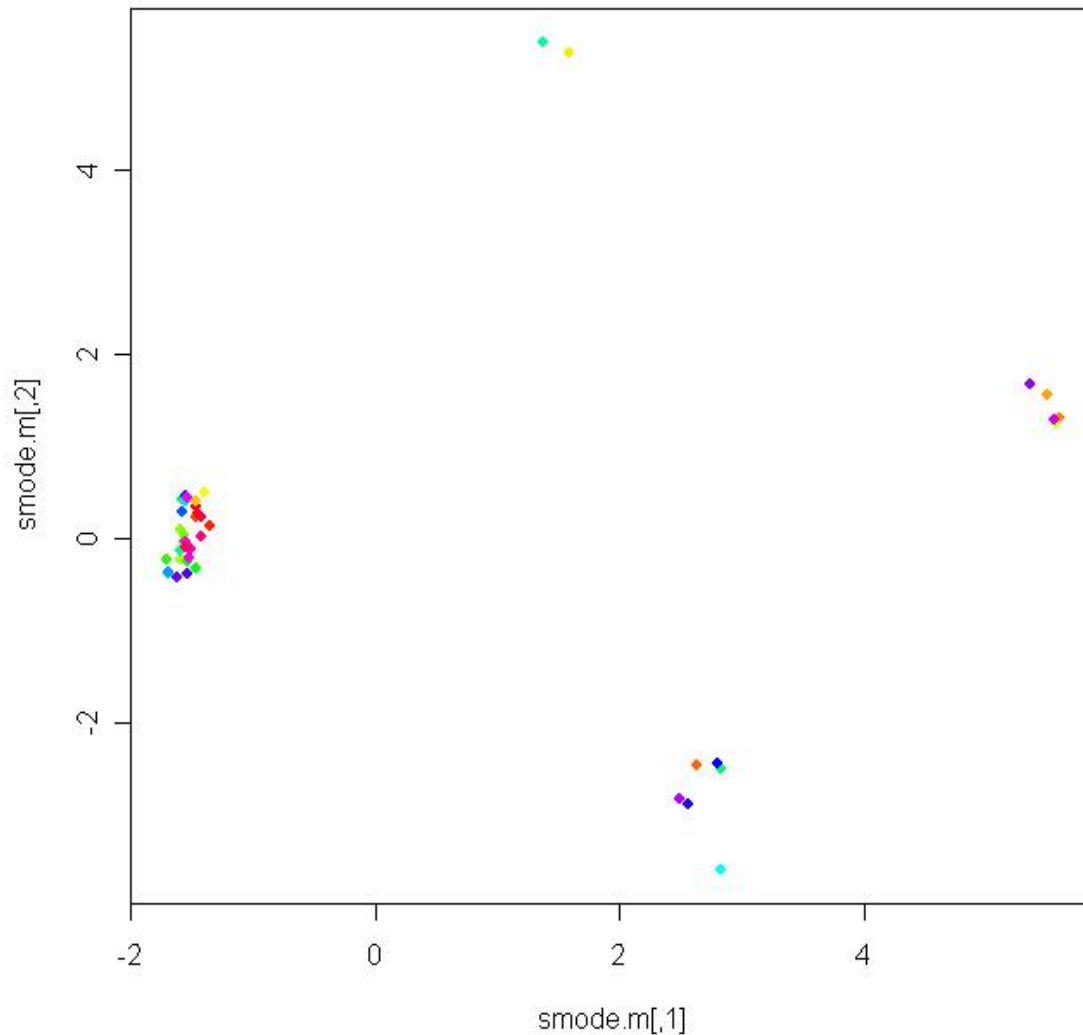
MDSによる空間配置



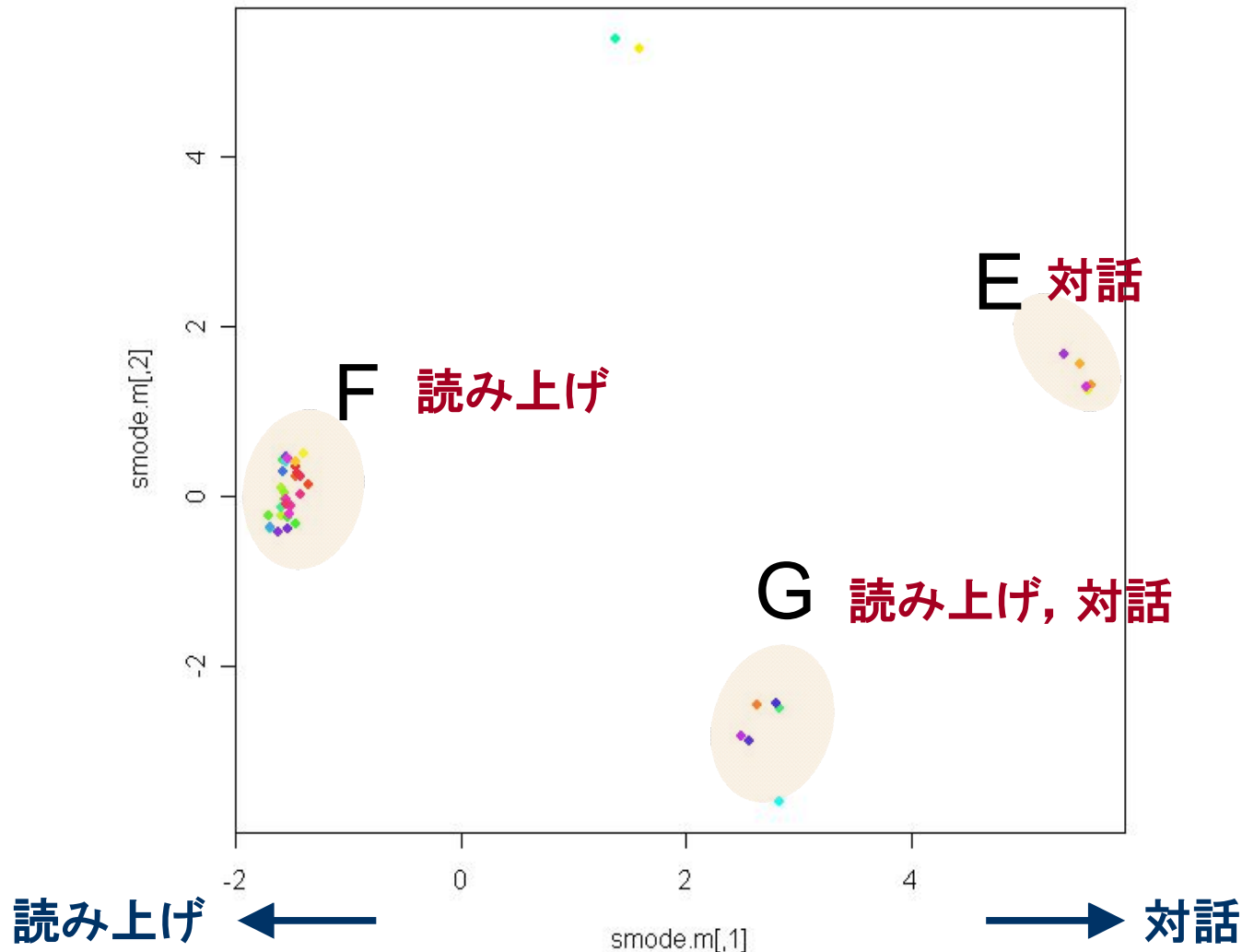
MDSによる空間配置



「発話モード」に重みをつけた配置



「発話モード」に重みをつけた配置



複数音声コーパス可視化法のまとめ

- コーパス特徴によるコーパス間の類似性の可視化
類似度の高いコーパス・特徴的な要素を持つコーパスの判断が容易
- 属性の値への重みづけ
利用者の意図に応じた空間配置の可視化の可能性を示唆
- 今後の課題
 - ・ 音響的特徴を表す項目の追加
 - ・ 可視化手法の客観・主観評価
 - ・ 複数コーパス検索のアプリケーション構築

音響データベース

- 音楽データベース
- 環境音データベース
- 騒音データベース

音楽データベース

RWCプロジェクト(1992-2001)

- ・ポピュラー音楽(100曲)
- ・著作権フリー音楽(15曲)
- ・クラシック(50曲)
- ・ジャズ(50曲)

オリジナル演奏・録音
CD33枚に収録

環境音データベース

RWCプロジェクト

音源定位, 音検索, 音認識等の研究用

1. 木, プラスチック, セラミックの衝突音等
2. 落下, ガス噴出, 物をこする音等
3. 金属, 紙の音, 楽器音等
4. 無響室, 残響室, 事務室, 動く音源の
インパルス応答 → 室の特性

DVD-ROM 3枚

騒音データベース

・環境騒音17種類

自動車内，展示会場，駅，工場，道路，交差点，人ごみ，列車，計算機室，空調機、エレベータホール等

DVD-R 2枚

(社)日本電子工業振興協会(JEIDA)作成

まとめ

- 音, 振動, 正弦波, 音声
- 音声合成・認識の概要
- 音声コーパス: 目的, 利用
- 大規模コーパスから見えること
- 音声コーパス類似性の可視化
- 音響データベースの紹介

参考図書紹介

板橋秀一編著「音声工学」
森北出版(2005, 2008) ¥3,400

板橋秀一編著「知識・知能と情報」
近代科学社(1992, 2000) ¥3,000

2009言語資源シンポジウム

「言語・音声データの学術利用に向けて」

日時:2009年10月15日(木)10:30ー16:30

会場:一橋記念講堂(東京都千代田区一ツ橋2-1-2)

【プログラム】

開会挨拶 東倉洋一 国立情報学研究所副所長

基調講演「言語資源の意義と重要性」 長尾真 国立国会図書館長

基調講演「データ中心科学の意義と重要性」 坂内正夫 国立情報学研究所長

招待講演「言語資源の利用と著作権」 岡村久道 弁護士, 国立情報学研究所客員教授

招待講演「高度言語情報融合フォーラムについて」 辻井潤一 東京大学教授

パネル討論「多様なニーズに応える言語資源」司会:板橋秀一 国立情報学研究所特任教授

「言語資源協会(GSK)の現状と課題」 橋田浩一 産総研サービス工学センター次長

「NTCIRの活動と言語資源の扱い」 神門典子 国立情報学研究所教授

「NICTにおける音声・言語資源」 鳥澤健太郎 情報通信研究機構グループリーダー

「国立国語研究所における言語資源」 前川喜久雄 国立国語研究所言語資源グループ長

「企業から見た言語資源」 松井くにお 富士通研究所主席研究員

閉会挨拶 板橋秀一 国立情報学研究所特任教授

講演題目等は仮題です。また、プログラムは都合により変更になる場合があります。

【問合せ先】

国立情報学研究所 音声資源コンソーシアム

URL <http://research.nii.ac.jp/src/>