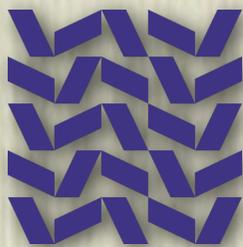


音声の構造的表象を通して 考察する幼児の音声模倣

～幼児は親の声の何を真似ようとしているのか？～

峯松 信明

東京大学大学院新領域創成科学研究科



本発表の流れ

幼児の言語獲得の持つ不思議さ

- Another POS (Poverty of Stimulus)
- 幼児は親の声の何を真似ようとしているのか？

話者不変の音声の構造的表象 ～音響的普遍構造～

- 峯松の声から峯松であることを除去する秘伝
- 階名としての「あいうえお」と音名としての「あいうえお」

普遍なる不変構造に基づく音声認識

- 一人の声で全ての人の声を認識する!?

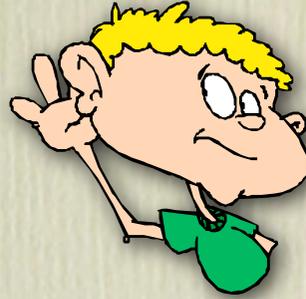
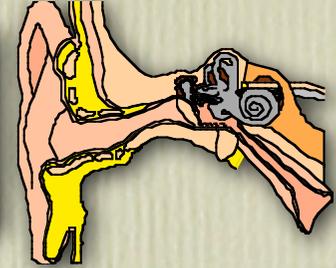
言語学／音楽学／脳科学／障害学／心理学／人類学・・・

- 失読症と自閉症 ～音声言語は勘違いの上に成立する～
- ヒトが進化のプロセスを通して獲得したコモンセンス

まとめ

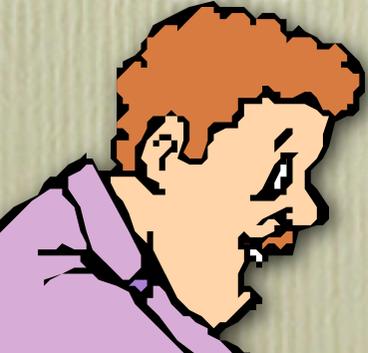
音声の多様性と容易性

音声の音響的側面 = 多様性のるつぼ

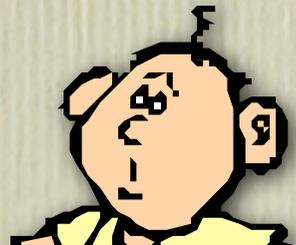


幼児の音声音響的環境 ~Another POS~

- 大部分は母親と父親の音声
- 極端に偏った音声コーパスに基づく超頑健な音声情報処理



おはよう



おはよう



音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

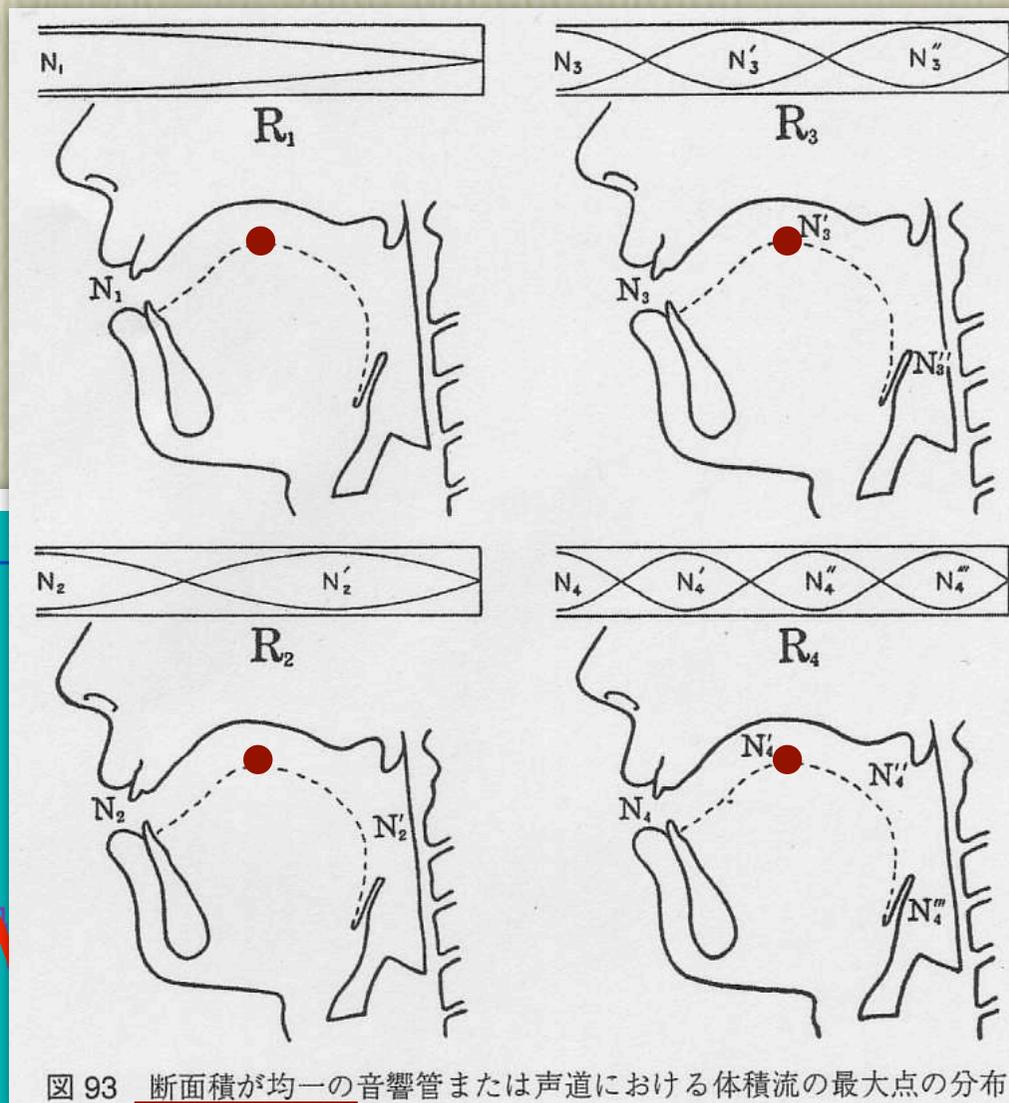
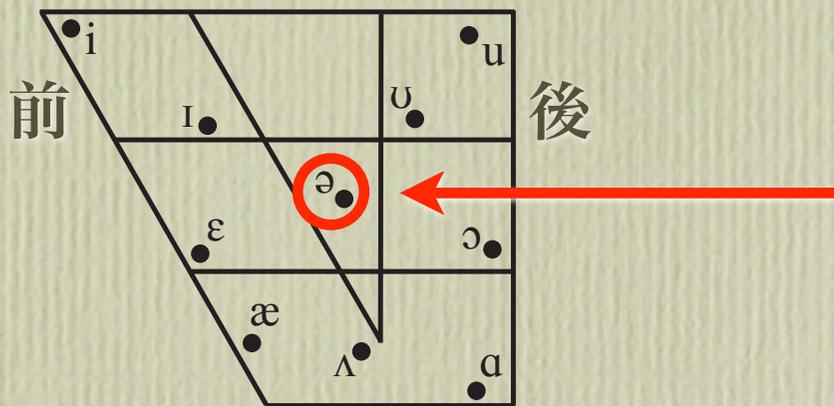
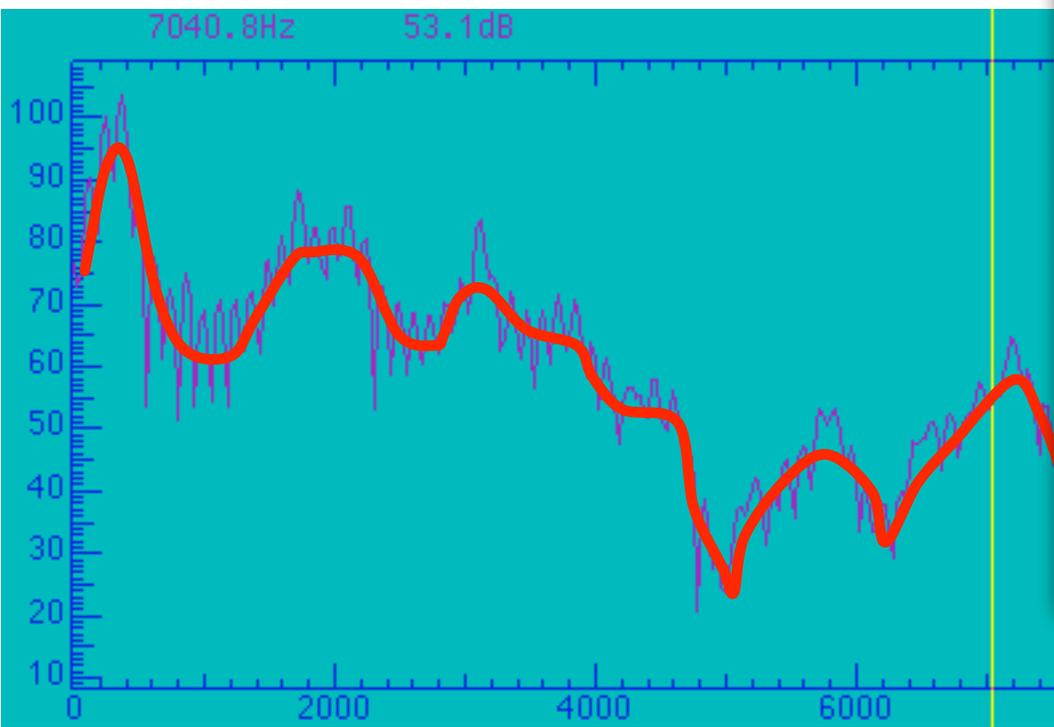


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

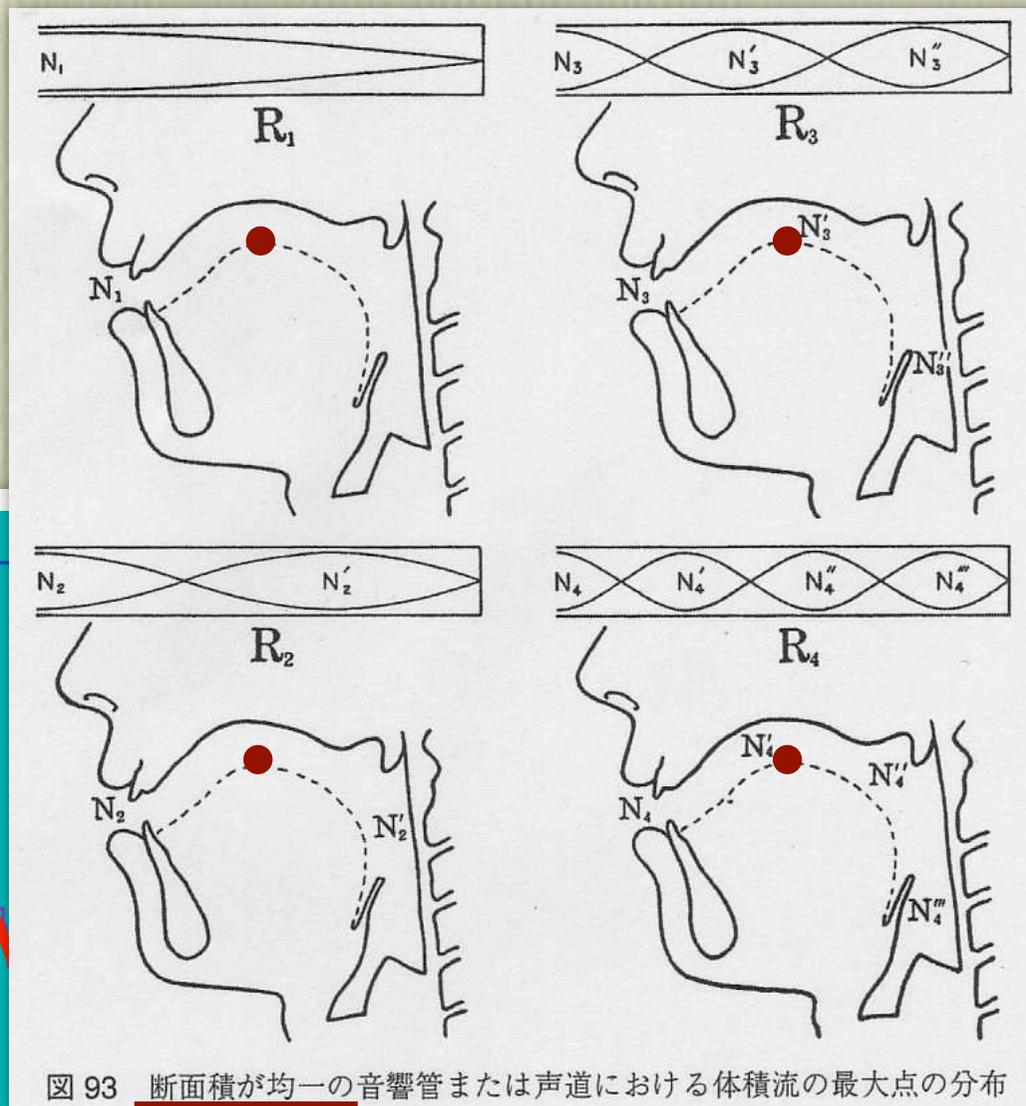
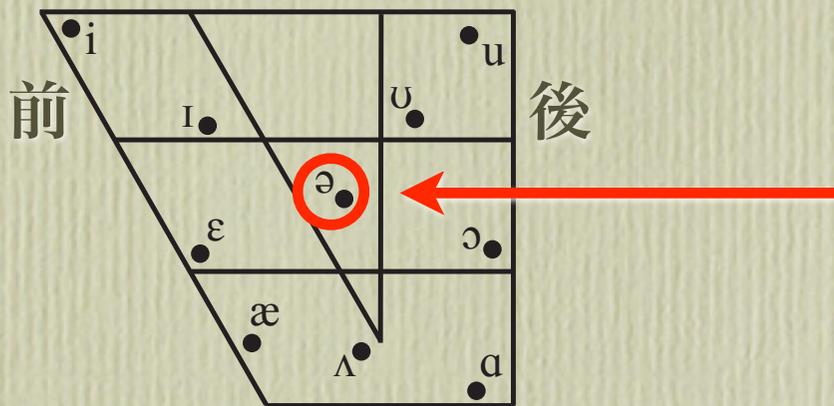
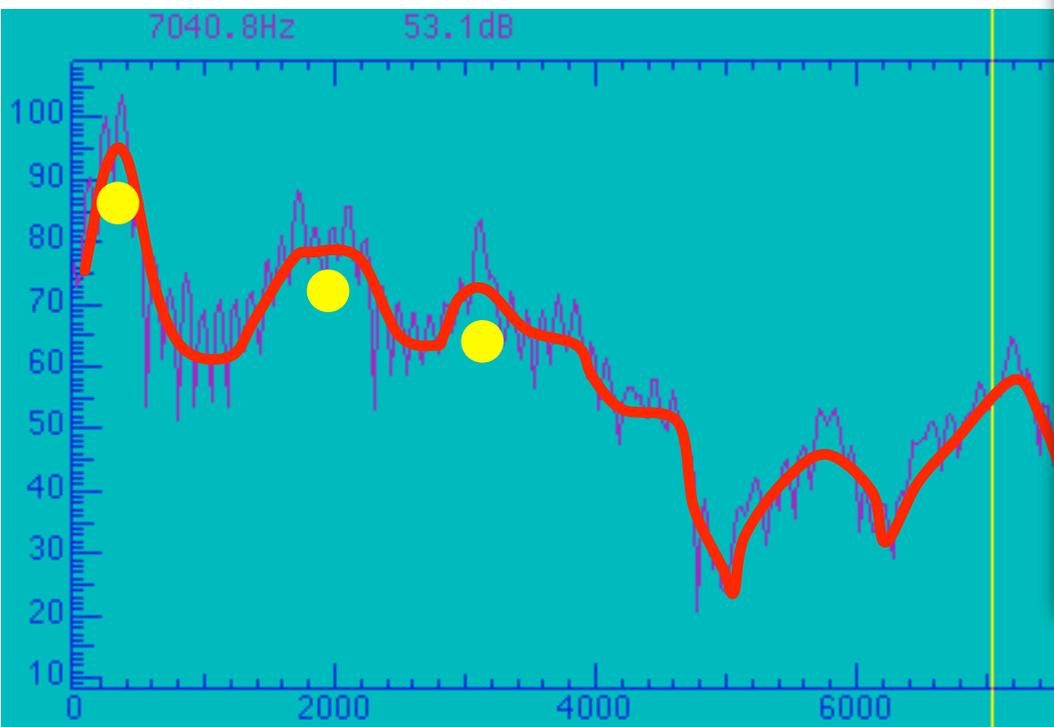


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

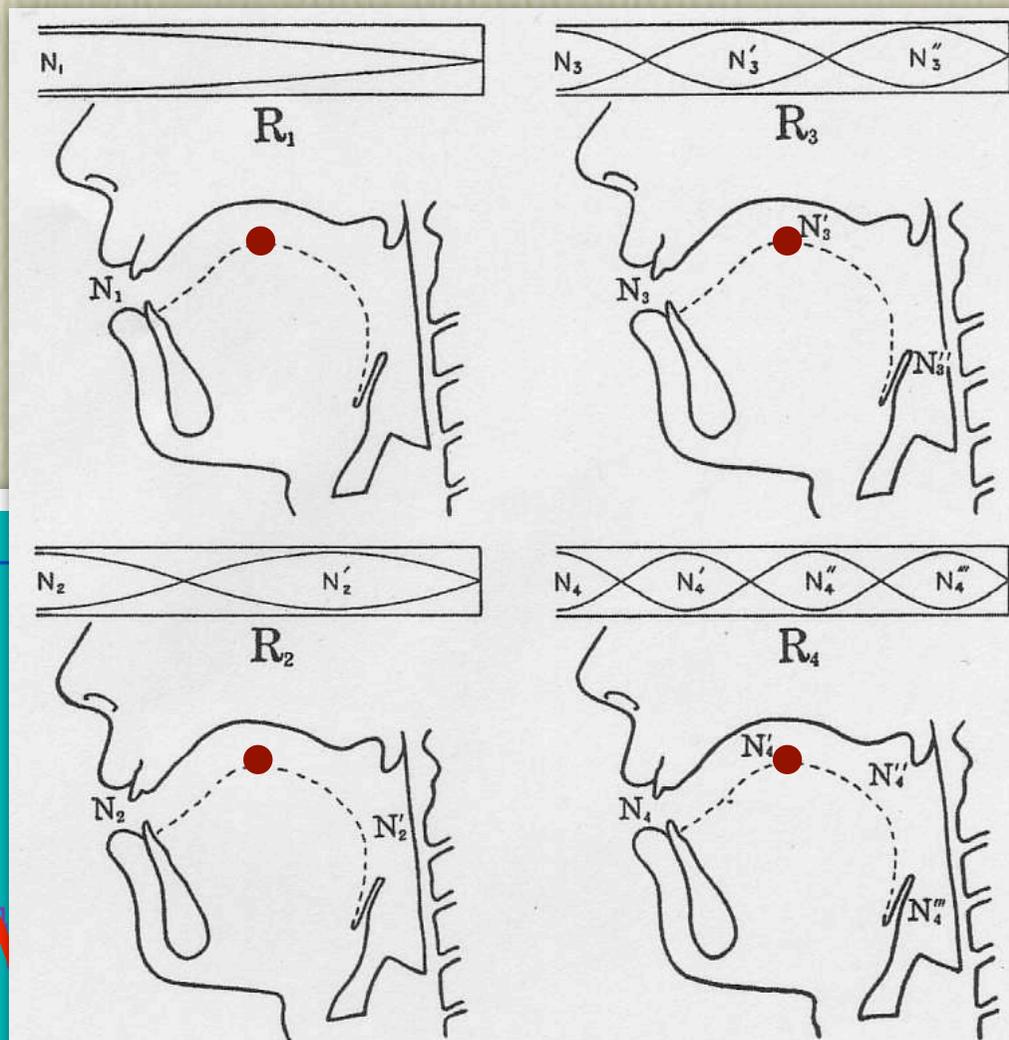
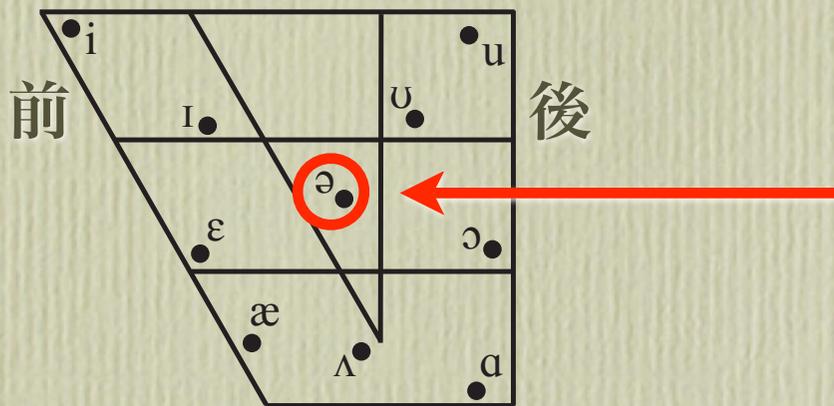
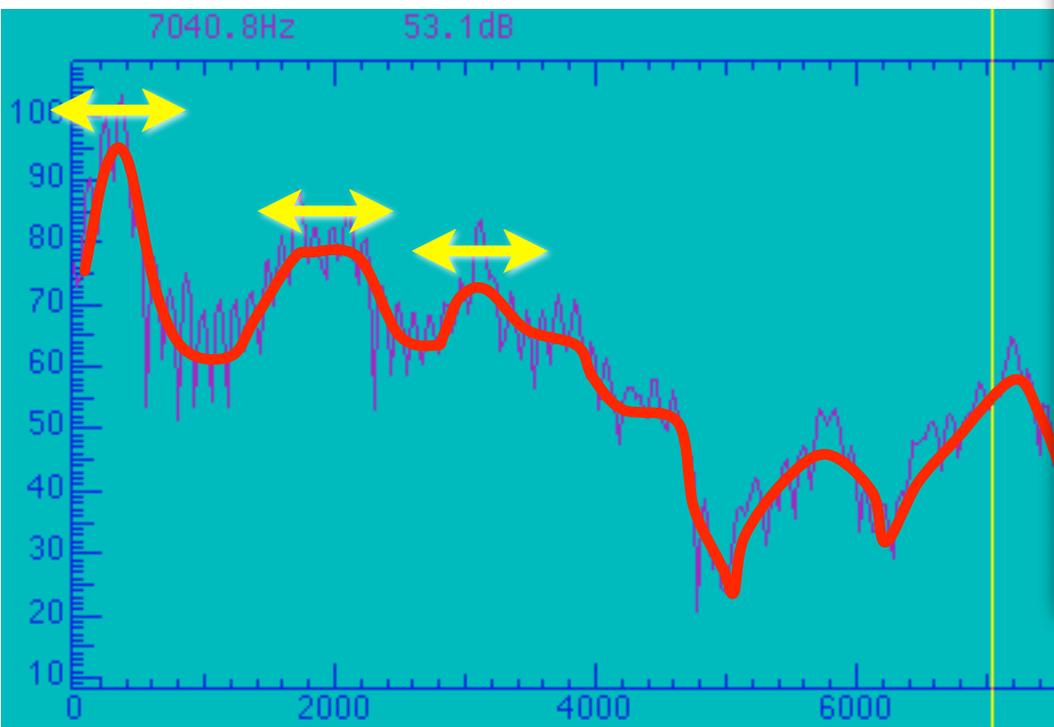


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

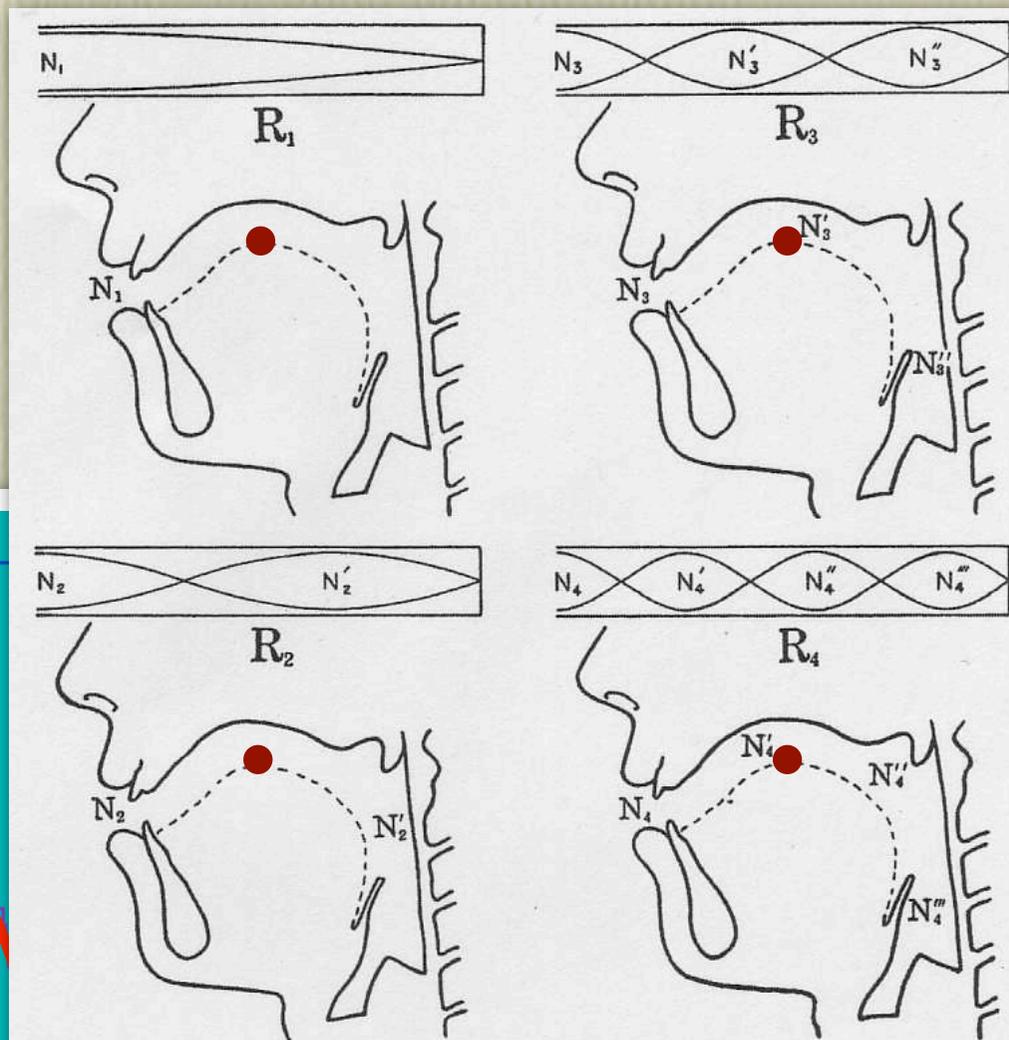
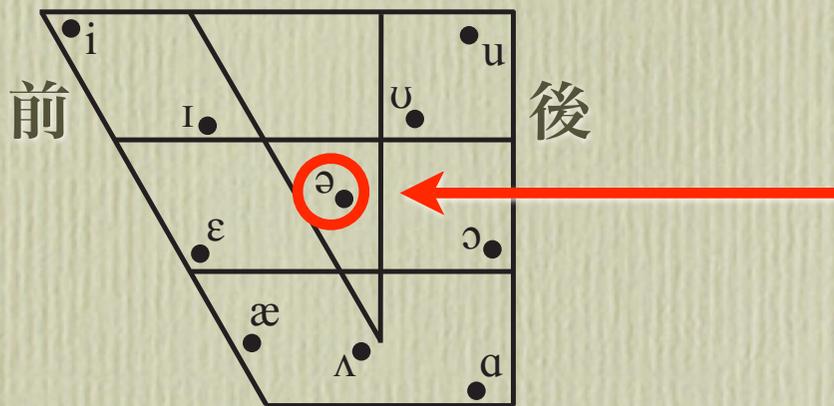
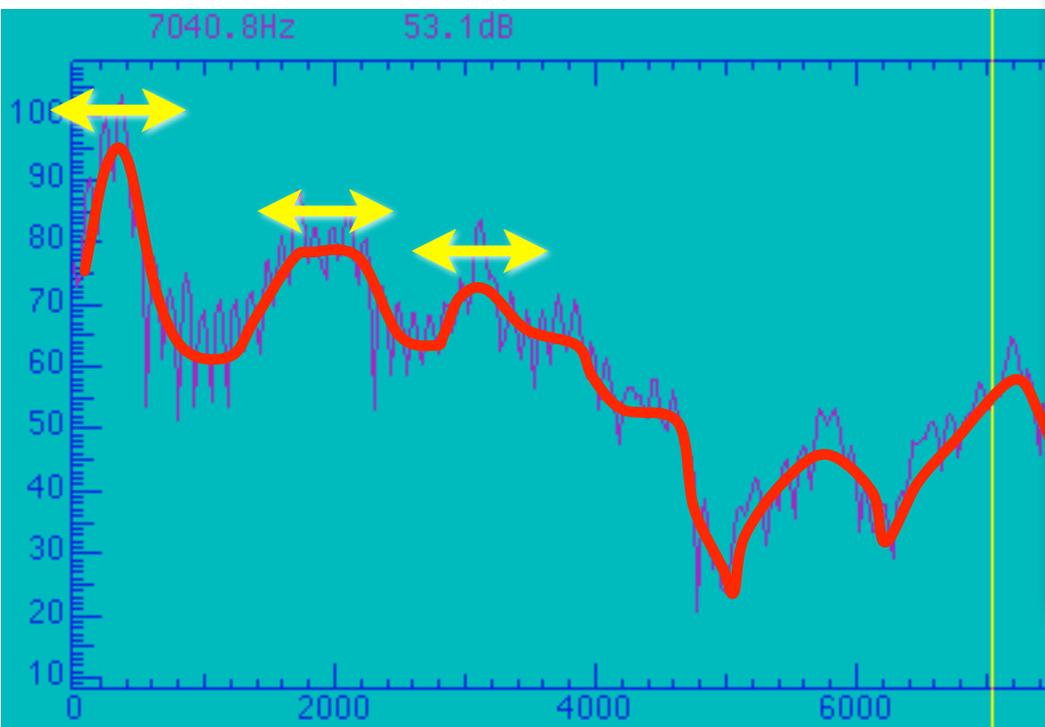


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

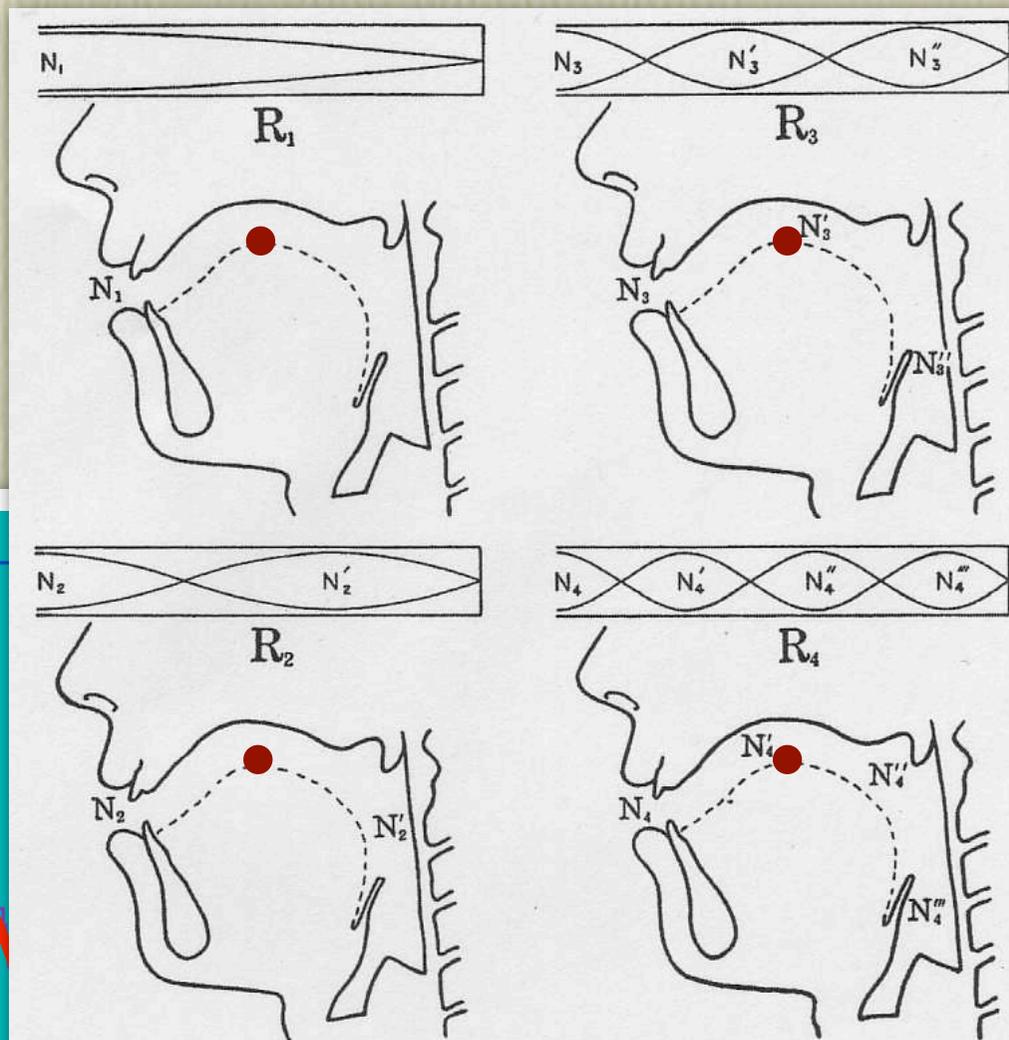
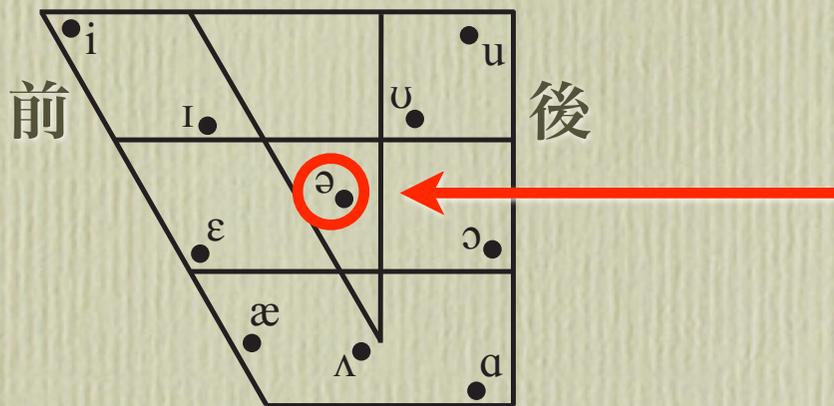
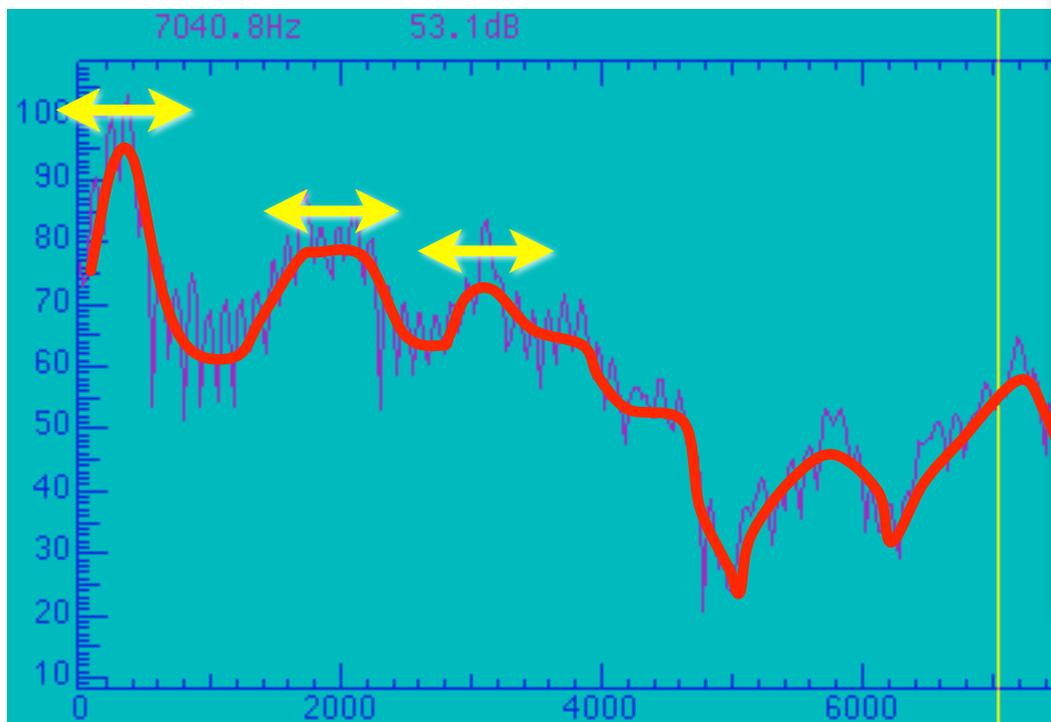


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

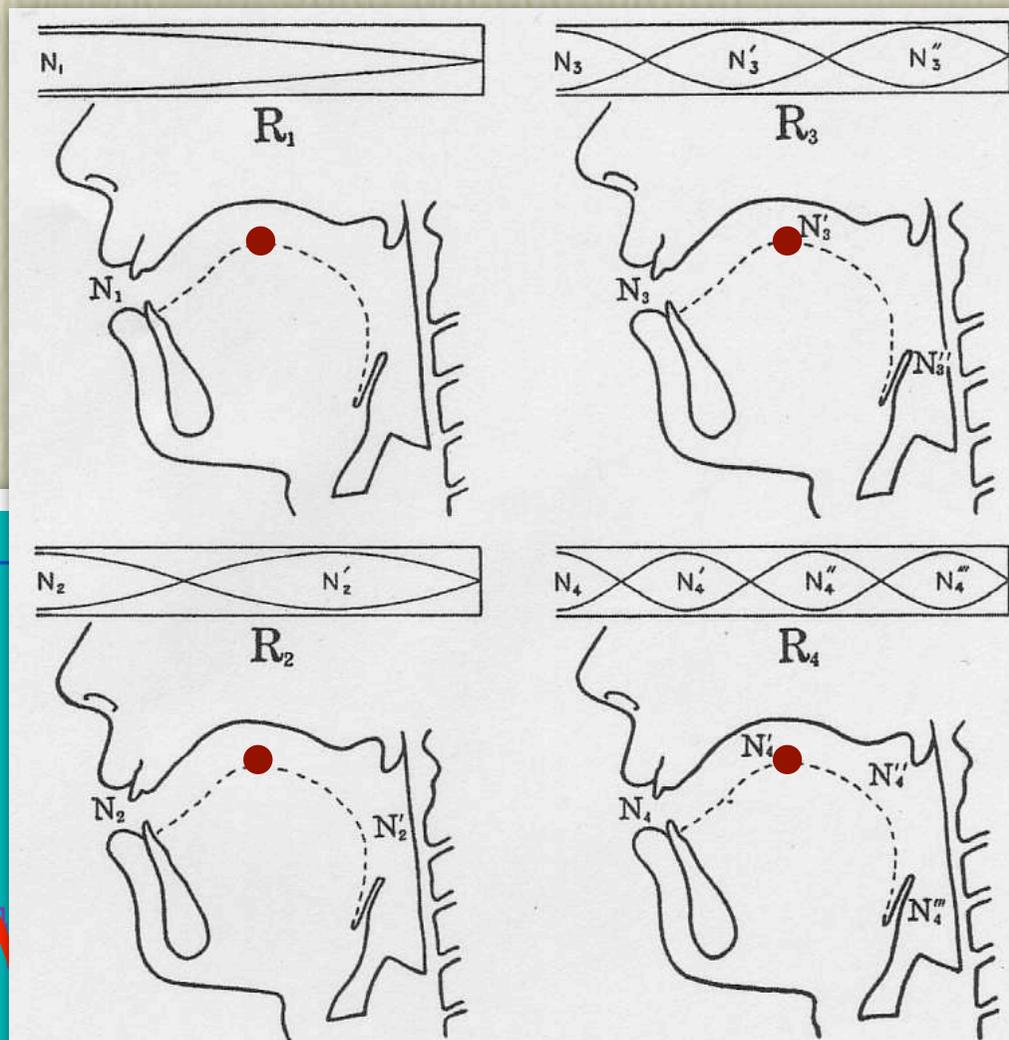
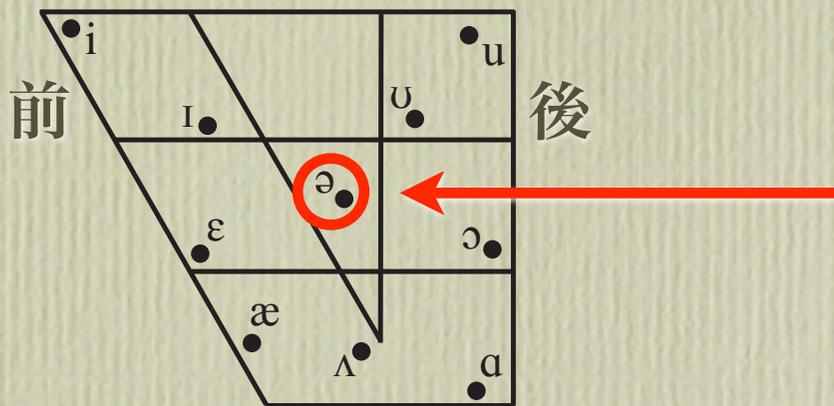
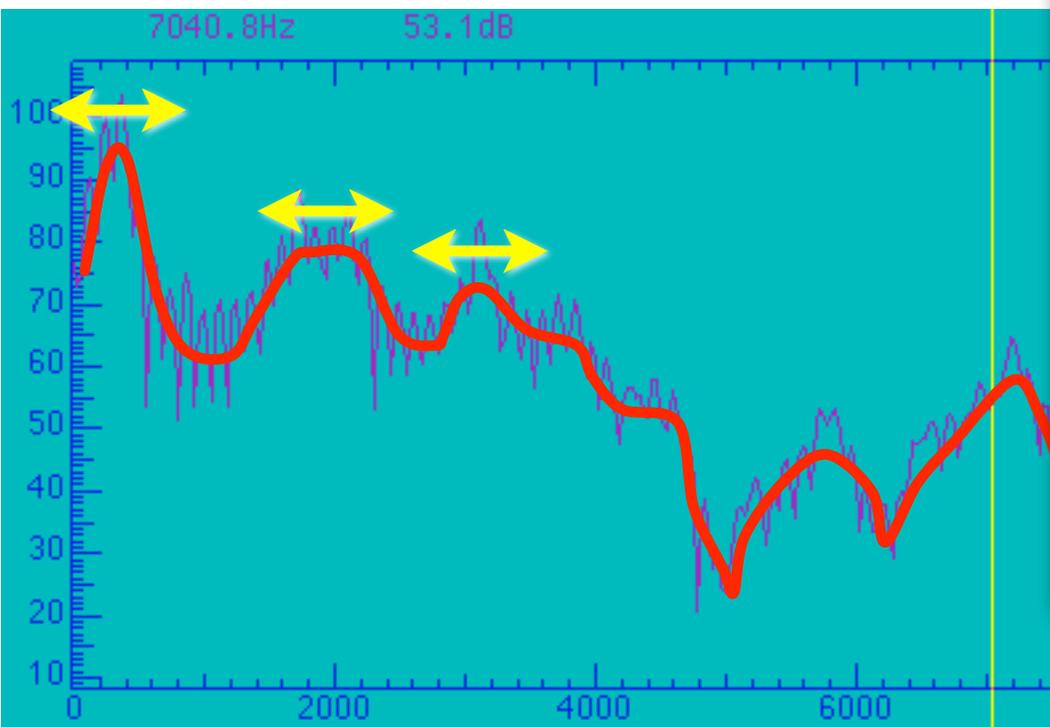


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

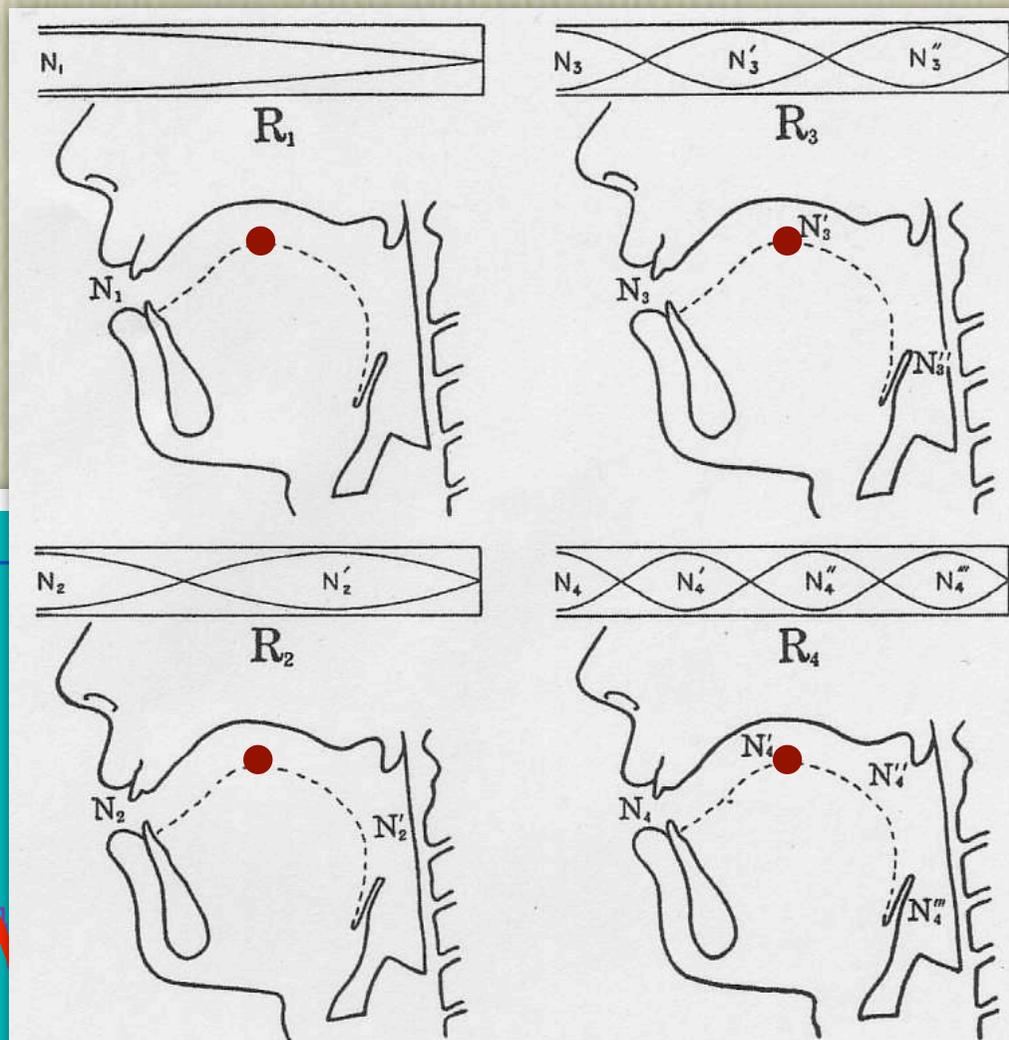
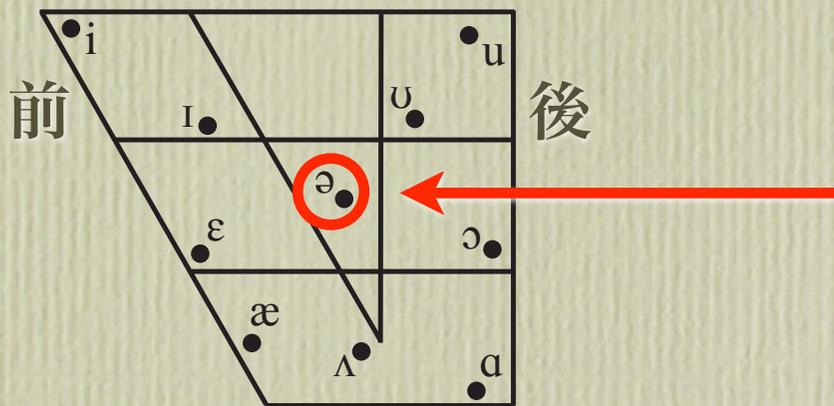
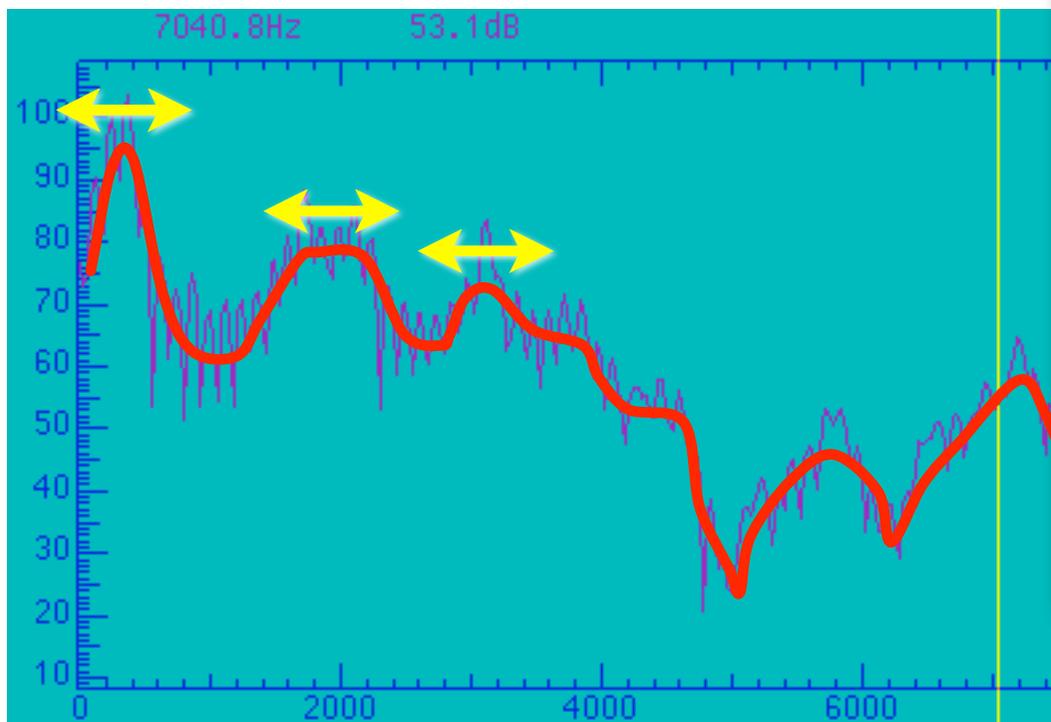


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

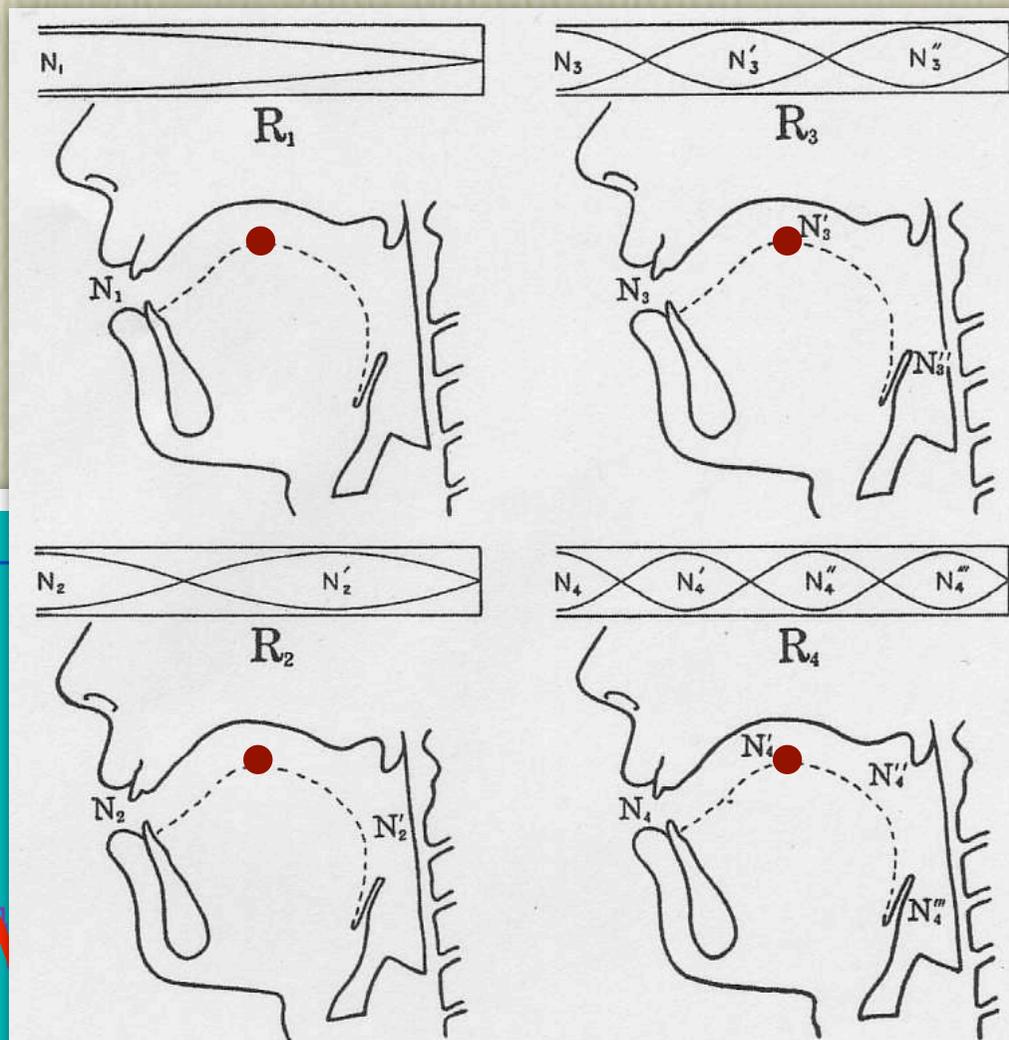
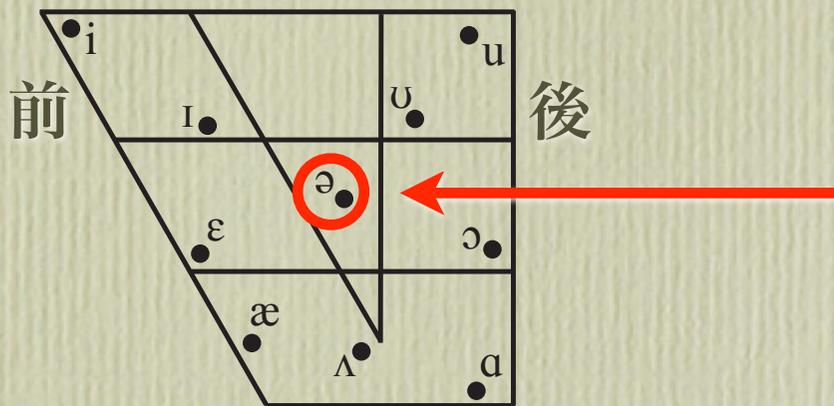
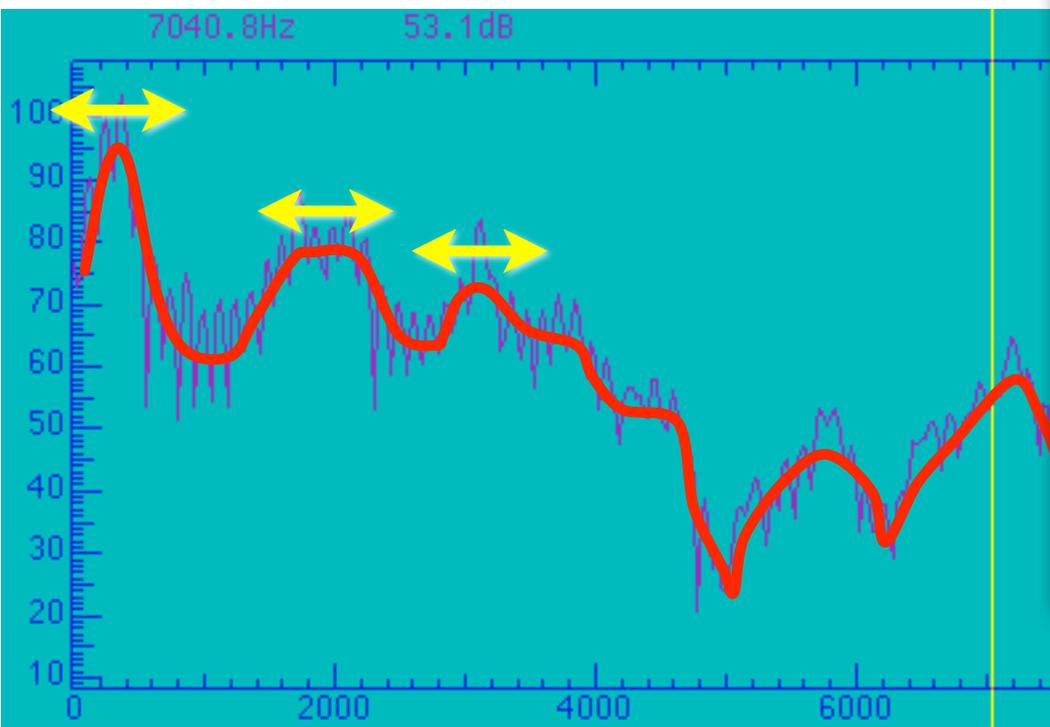


図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

音声の多様性の物理的描写

人間の成長 = 声道長の伸長 = フォルマントの下降

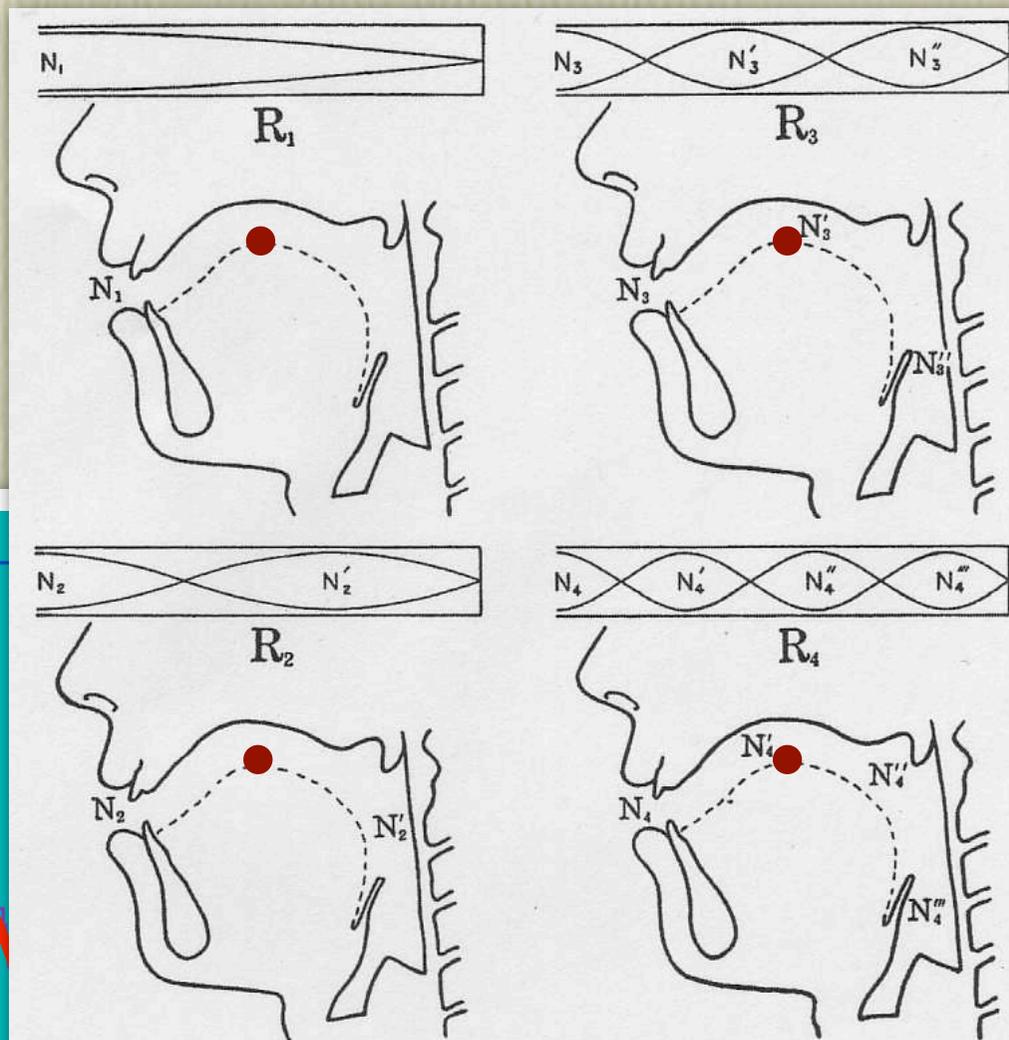
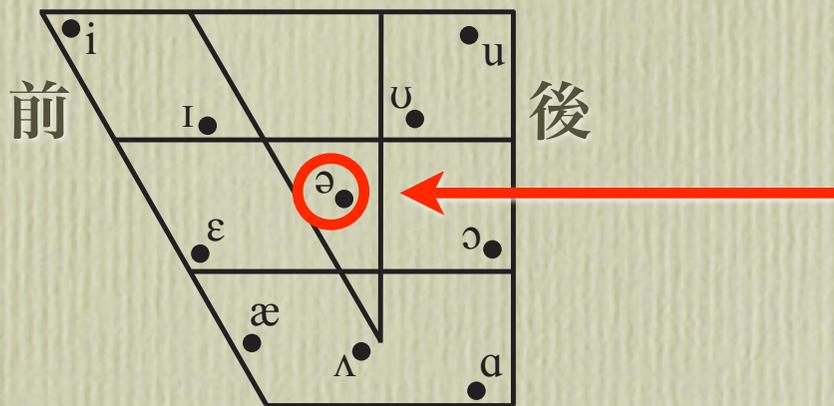
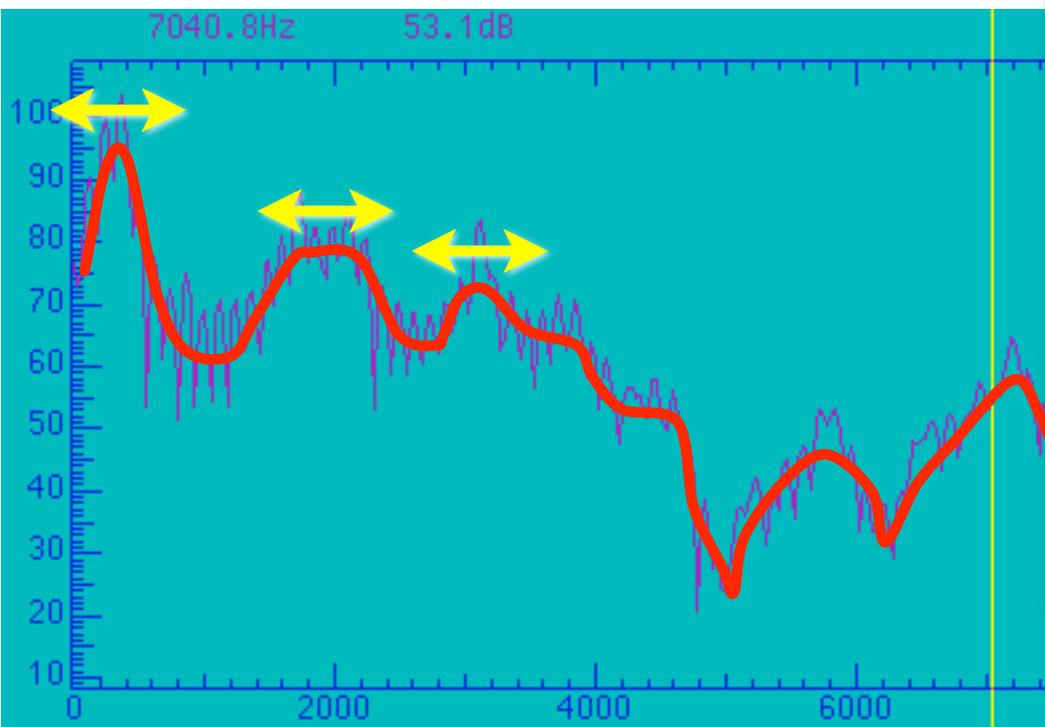


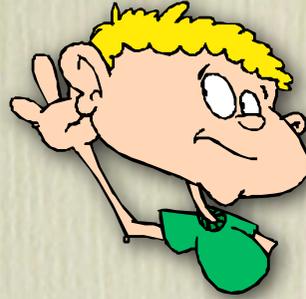
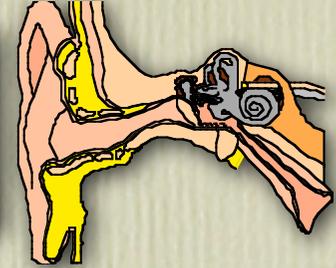
図 93 断面積が均一の音響管または声道における体積流の最大点の分布



$$F_n = \frac{c}{4l} (2n + 1)$$

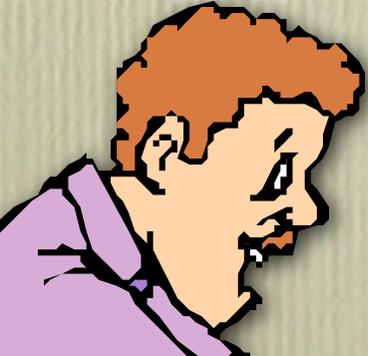
音声の多様性と容易性

音声の音響的側面 = 多様性のるつぼ

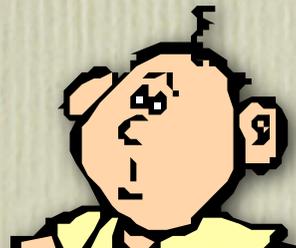


幼児の音声音響的環境 ~Another POS~

- 大部分は母親と父親の音声
- 極端に偏った音声コーパスに基づく超頑健な音声情報処理



おはよう

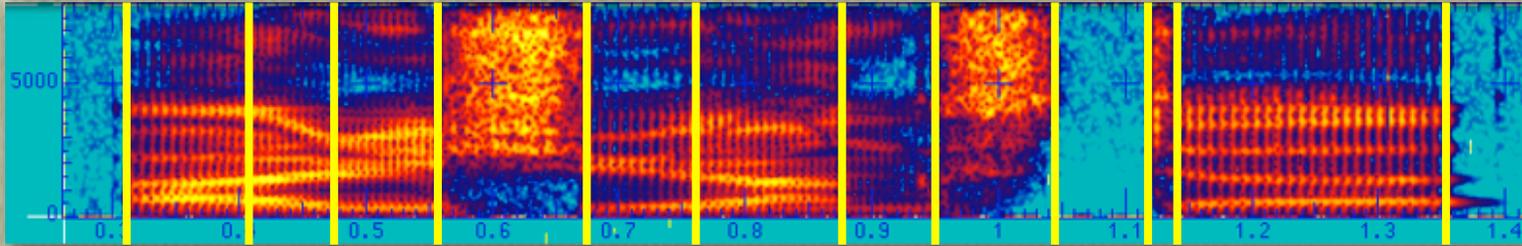


おはよう



音声 = 音シンボル列を音化したもの？

音声 = 音シンボルの並び ～音声学の常識～



[a b c d e f g h i j k]

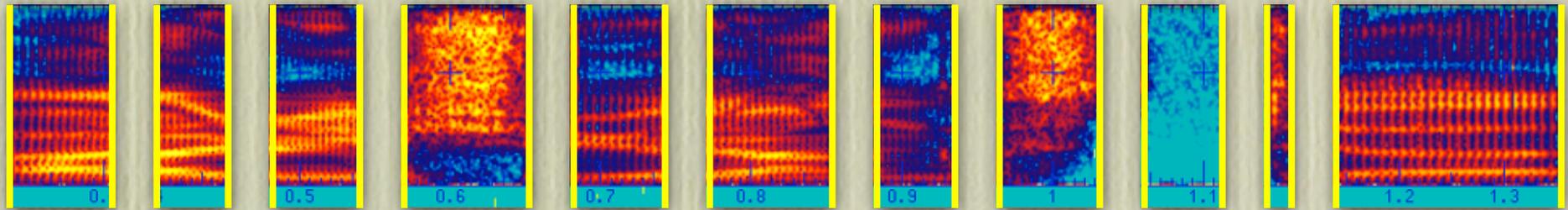
● 音声 ⇔ 音シンボル列 = 音声認識 / 音声合成

● 音シンボル列を各々音に変換したものが音声なのか？

- 音シンボル（音韻）の意識の定着は小学校入学以降
- 単語音声を要素に分割することが困難でも，音声活動を始める
- 「子供は語全体の音形・ゲシュタルトを獲得してから，語を構成する個々の分節音の獲得へと進む（加藤'03，早川'06）」
- 語（発話）全体の音形を音にしたものが音声

音声＝音シンボル列を音化したもの？

音声＝音シンボルの並び ～音声学の常識～



↕
a b c d e f g h i j k

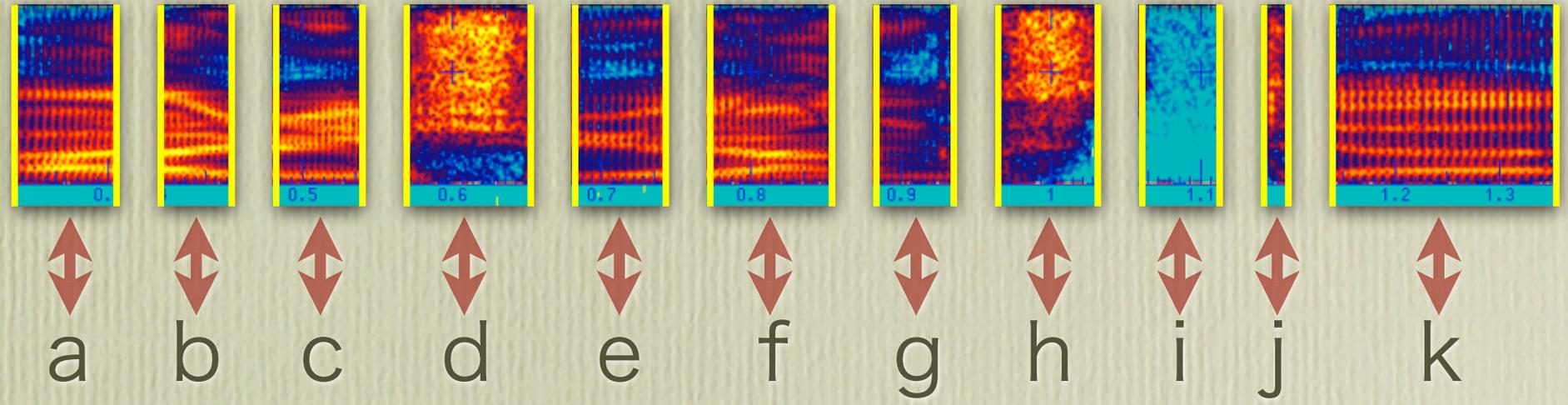
● 音声 ⇔ 音シンボル列 = 音声認識 / 音声合成

音シンボル列を各々音に変換したものが音声なのか？

- 音シンボル（音韻）の意識の定着は小学校入学以降
- 単語音声要素を分割することが困難でも、音声活動を始める
- 「子供は語全体の音形・ゲシュタルトを獲得してから、語を構成する個々の分節音の獲得へと進む（加藤'03, 早川'06）」
- 語（発話）全体の音形を音にしたものが音声

音声 = 音シンボル列を音化したもの？

音声 = 音シンボルの並び ~音声学の常識~



● 音声 ⇔ 音シンボル列 = 音声認識 / 音声合成

● 幼児の音声音響的環境 ~Another POS~

- 大部分は母親と父親の音声
- 極端に偏った音声コーパスに基づく超頑健な音声情報処理



一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



/x, y, z, ..., /
[x, y, z, ...,]

一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



✗ /x, y, z, ..., /
[x, y, z, ...,]

一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



✓ /x, y, z, ..., /
✓ [x, y, z, ...,]



一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



✗ /x, y, z, ..., /
[x, y, z, ...,]

一つの単純な問いかけ

幼児



INTERNATIONAL CONFERENCE ON INFANT STUDIES

JUNE 19-23, 2006
KYOTO, JAPAN

CO-SPONSORED BY ISIS AND JSBS



▶ Home

▶ Conference Program

▶ Shaken Baby PreConf.

▶ Registration

▶ Hotel Reservations

▶ Transportation

▶ Dining

▶ Call For Papers

▶ Sightseeing

▶ Travel Awards

▶ Submissions

WELCOME

Welcome to the website for the International Conference on Infant Studies, to be held June 19-23, 2006 in beautiful Kyoto, Japan.

The goals of the conference are:

- * To further internationalize infancy research.
- * To increase the involvement and impact of early career researchers, that is, pre-mid-tenure faculty, postdocs and graduate students, in infancy research.
- * To have an innovative program that offers high intellectual and collaborative value to attendees.

Please follow the link to the Call For Papers for more details on what is already taking shape as an exciting program in an exquisite location.

TO CONTACT US:

Prof. Joseph J. Campos
Institute of Human Development
University of California, Berkeley
Berkeley, California, USA

Prof. Ichiro Uchiyama
Psychology Department
Doshisha University
Kyoto, JAPAN

E-mail: infancy@berkeley.edu

一つの単純な問いかけ

??? QUESTION ???

We're faced with a simple but difficult question to answer. We believe that this is one of the most fundamental questions of language acquisition. If you have any comments, please let us know.

What in a father's speech is mimicked *acoustically* by his young children?

Young children often mimic their father's speech. Acoustically speaking, however, they are not mimicking the utterance as it is. They are not trying to produce *their father* in their voices. Some may say that the children decode the utterance into a sequence of phonemes and then, each phoneme is generated by the mouth. But the young children don't have good phonemic awareness and it is rather difficult for them to decode speech. In this situation, what is mimicked acoustically? As acoustic characteristics of a speaker are determined by the shape of his articulators, it is certainly impossible to produce their father. But it is true that, if they mimic the utterance acoustically, they must at least *try* to produce their father but we've never seen such children. We still believe that they are mimicking something in the utterance acoustically. What is THAT?



Nobuaki MINEMATSU(mine@gavo.t.u-tokyo.ac.jp) & Tazuko NISHIMURA(nt-tazuko@ams.odn.ne.jp)
If you have any answers, please visit our poster (9:00-10:50 on Wed in Colonade of Miyako Southeast).

一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



✗ /x, y, z, ..., /
[x, y, z, ...,]

一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



~~/x, y, z, ..., /~~
~~[x, y, z, ...,]~~

言語情報
パラ言語情報
非言語情報

言語情報
パラ言語情報
非言語情報

一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



× /x, y, z,.../
[x, y, z,...]

言語情報
パラ言語情報
非言語情報

言語情報
パラ言語情報
非言語情報



一つの単純な問いかけ

幼児は父親の声の何を真似ようとしているのか？



× /x, y, z,.../
[x, y, z,...]

言語情報
パラ言語情報
非言語情報

言語情報
パラ言語情報
非言語情報



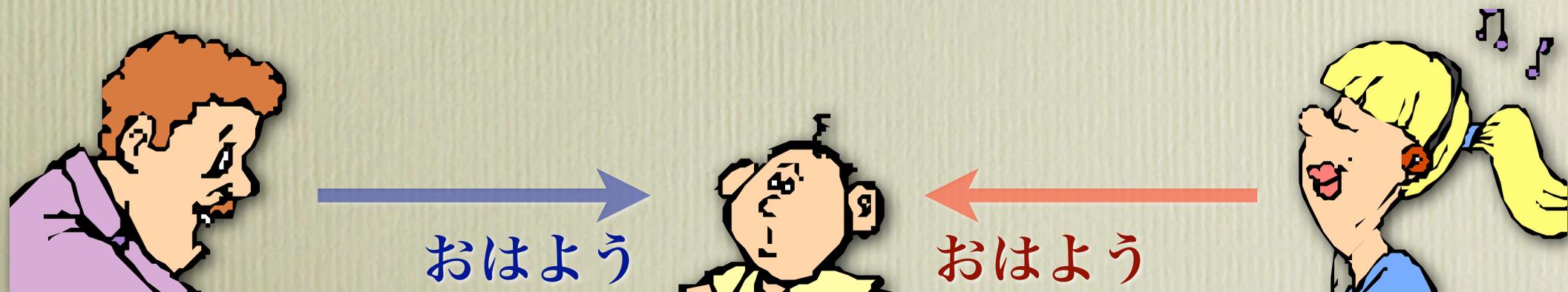
発達心理学・赤ちゃん研究の主張

要素 vs. 全体

- 幼児は単語を構成する個々の分節音（音韻）の獲得の前に，単語全体の音形・語ゲシュタルトを獲得する。

音の同定 vs. 音の区別

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）



発達心理学・赤ちゃん研究の主張

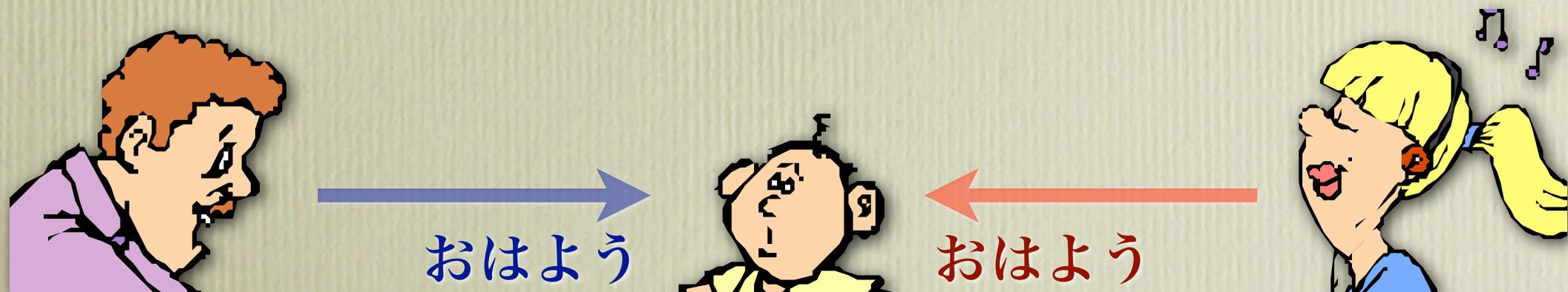
要素 vs. 全体

- 幼児は単語を構成する個々の分節音（音韻）の獲得の前に，単語全体の音形・語ゲシュタルトを獲得する。

音の同定 vs. 音の区別

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

音そのものではなく，
音と音の違いを捉えること



発達心理学・赤ちゃん研究の主張

要素 vs. 全体

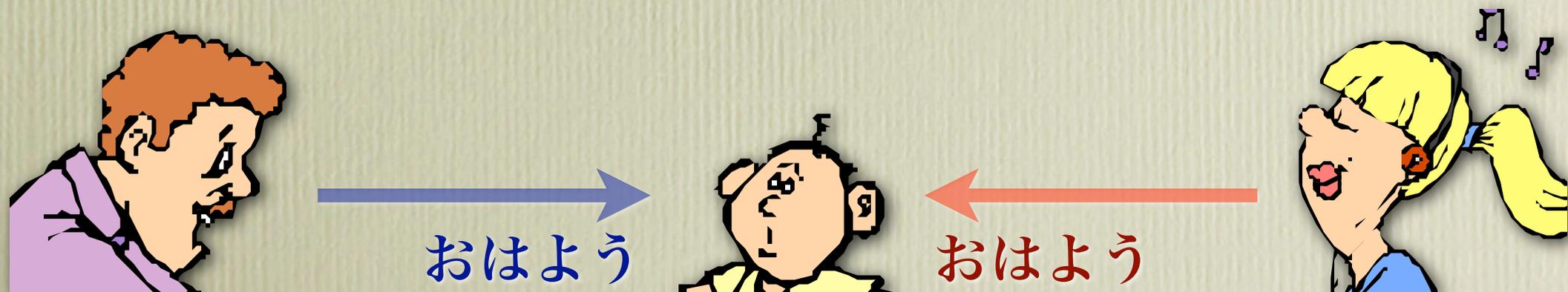
- 幼児は単語を構成する個々の分節音（音韻）の獲得の前に，単語全体の音形・語ゲシュタルトを獲得する。

音の同定 vs. 音の区別

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

音そのものではなく，
音と音の違いを捉えること

語（発話）全体を表現する
音響表象を考えること



発達心理学・赤ちゃん研究の主張

要素 vs. 全体

- 幼児は単語を構成する個々の分節音（音韻）の獲得の前に，単語全体の音形・語ゲシュタルトを獲得する。

音の同定 vs. 音の区別

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

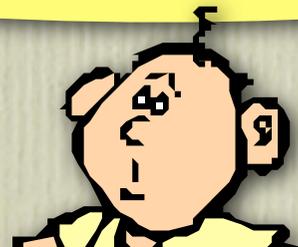
音そのものではなく，
音と音の違いを捉えること

語（発話）全体を表現する
音響表象を考えること

話者・年齢・性別・マイクなどに
不変な音声の全体的表象

おはよう

おはよう



発達心理学・赤ちゃん研究の主張

要素 vs. 全体

- 幼児は単語を構成する個々の分節音（音韻）の獲得の前に，単語全体の音形・語ゲシュタルトを獲得する。

音の同定 vs. 音の区別

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

音そのものではなく，
音と音の違いを捉えること

語（発話）全体を表現する
音響表象を考えること

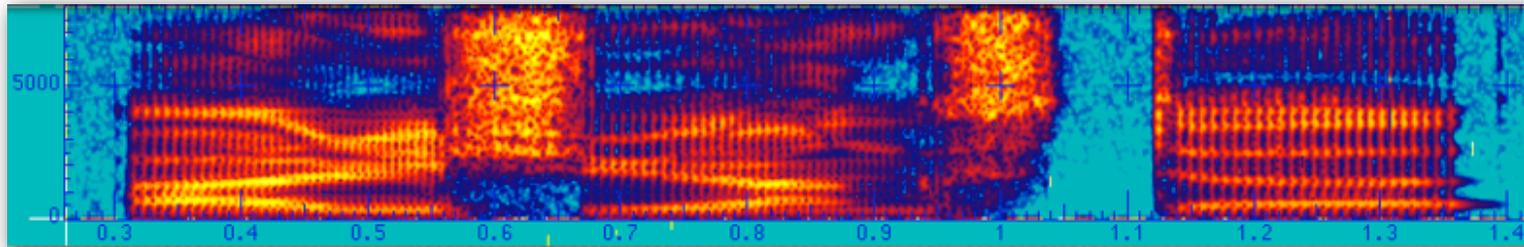
話者・年齢・性別・マイクなどに
不変な音声の全体的表象

おはよう



発達心理学・赤ちゃん研究の主張

要素 vs 全体



単

- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

音そのものではなく，
音と音の違いを捉えること

語（発話）全体を表現する
音響表象を考えること

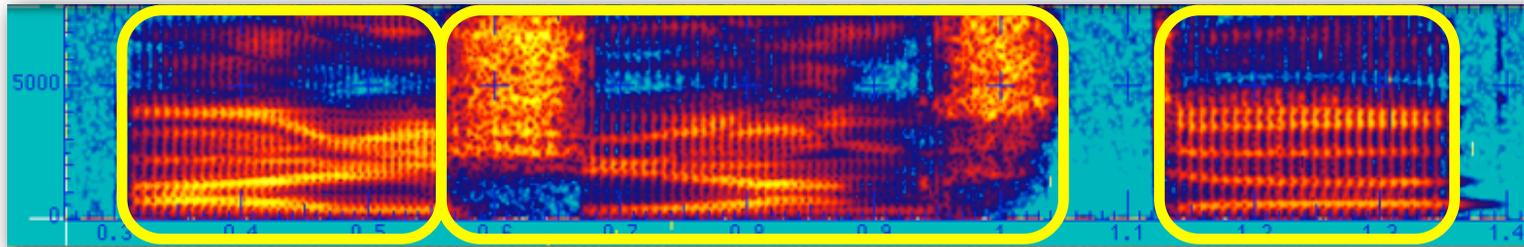
話者・年齢・性別・マイクなどに
不変な音声の全体的表象

おはよう



発達心理学・赤ちゃん研究の主張

要素 vs 全体



単

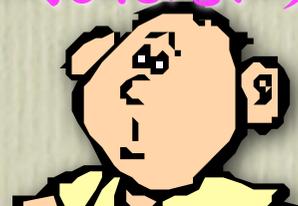
- 音を音韻として同定する以前に，幼児は，音と音の区別がまずできるようになる（例：日本人幼児の [r] と [l] の区別）

音そのものではなく，
音と音の違いを捉えること

語（発話）全体を表現する
音響表象を考えること

話者・年齢・性別・マイクなどに
不変な音声の全体的表象

おはよう



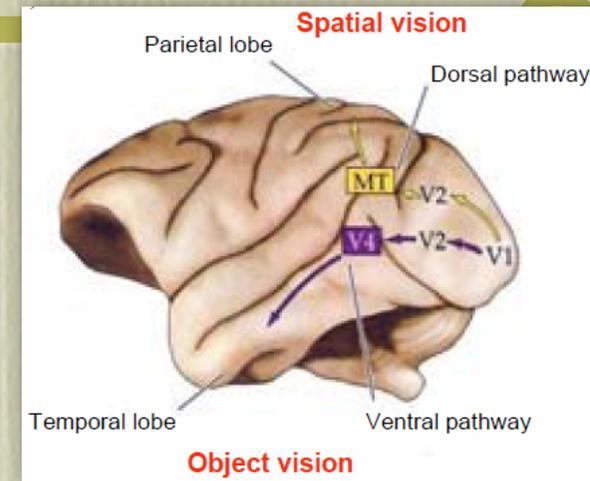
神経生理学・脳科学の知見より

蝸牛の議論から新皮質の議論へ

- 聴覚における蝸牛＝視覚における網膜
- 蝸牛モデル→大脳新皮質モデルへ

複数の情報処理経路による分離処理

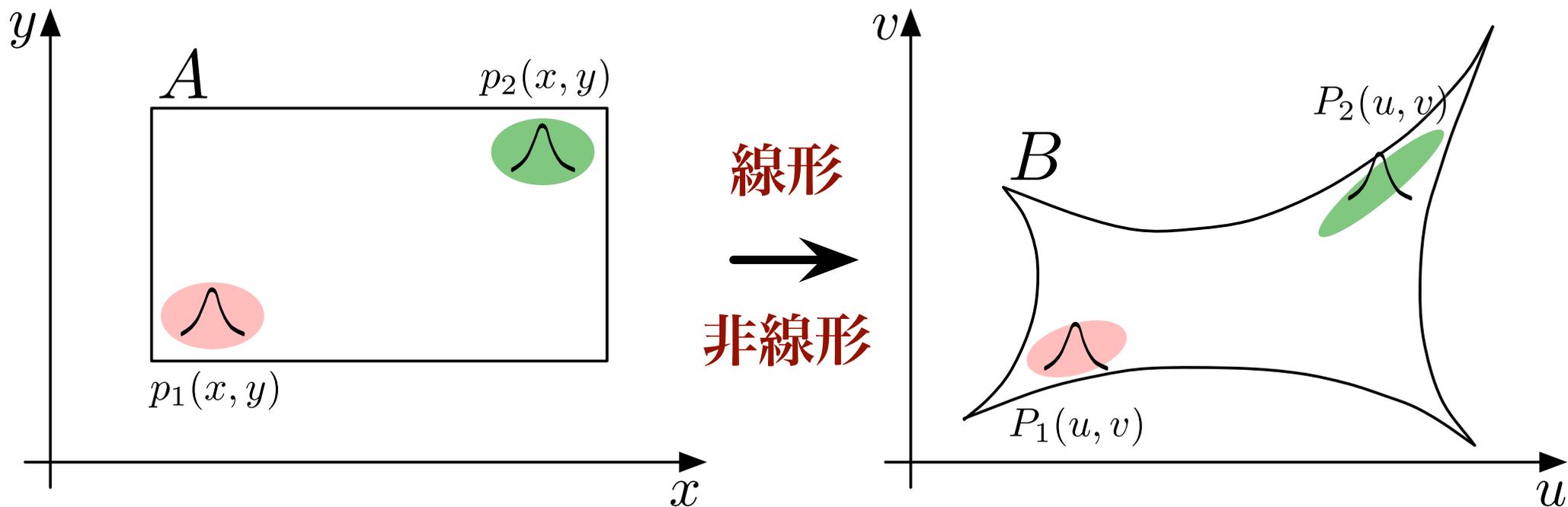
- 視覚：「何 (what)」経路と「如何に (how)」経路
 - 何＝腹側経路＝形，色，キメに関する情報を取得
 - 如何＝背側経路＝動き，位置に関する情報を取得
 - 広く受け入れられている古典的モデル
- 聴覚における情報処理は？ 【言語】と【非言語】の分離？
 - 例えば，Belin & Zatorre (2000)では……
 - 腹側経路＝静的な情報を処理＝非言語情報（話者）の同定
 - 音楽で言えば楽器の同定
 - 背側経路＝動的な情報を処理＝言語情報の同定 → speech motions
 - 音楽で言えばメロディーの同定



話者不変な音響量の数学的探求

二人の話者の音声音響空間において不変なる物理量

- 全ての音響事象は点ではなく、分布として表現される
 - 空間Aの分布 p は空間Bの分布 P へと写像される (一対一)
 - p と P は異なる物理特性を持つ
 - 両空間において不変な物理量はどこにある？



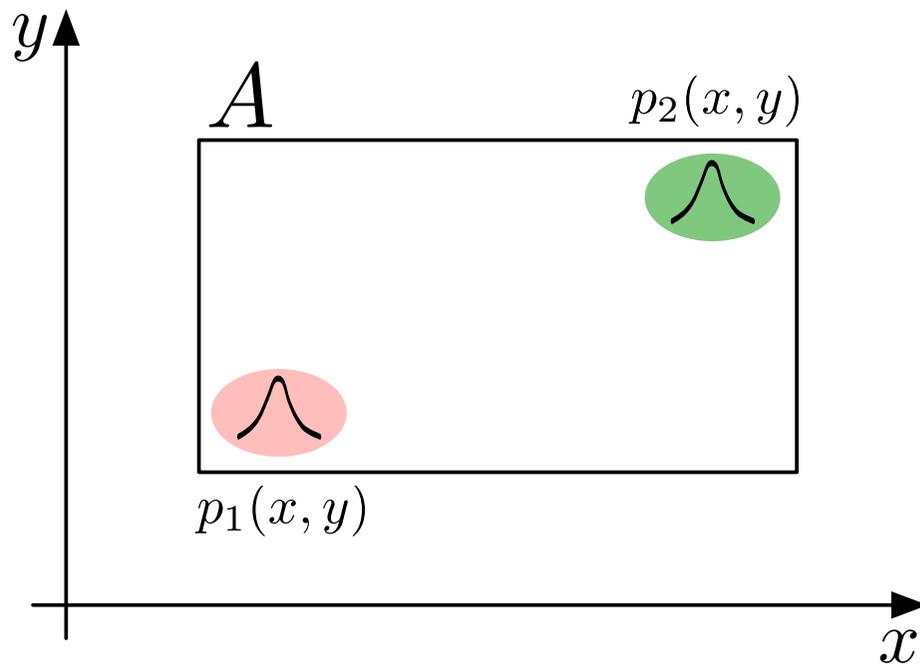
話者不変な音響量の数学的探求

バタチャリヤ距離 (=分布間距離尺度の一つ)

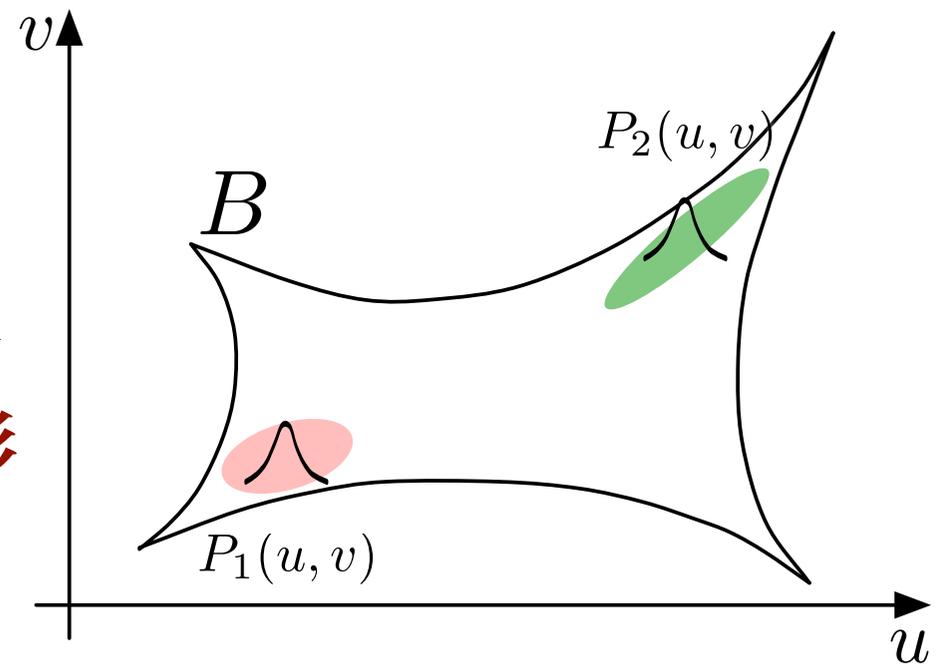
● 座標変換による式の変形 $x = x(u, v)$, $y = y(u, v)$

● $BD(p_1(x, y), p_2(x, y))$

$$= -\log \iint \sqrt{p_1(x, y)p_2(x, y)} dx dy$$



線形
→
非線形



話者不変な音響量の数学的探求

バタチャリヤ距離 (=分布間距離尺度の一つ)

● 座標変換による式の変形 $x = x(u, v)$, $y = y(u, v)$

● $BD(p_1(x, y), p_2(x, y))$



$$= -\log \iint \sqrt{p_1(x, y)p_2(x, y)} dx dy$$

$$= -\log \iint \sqrt{q_1(u, v)q_2(u, v)} |J(u, v)| dudv$$

$$= -\log \iint \sqrt{\frac{q_1(u, v)|J(u, v)|}{|J(u, v)|} \cdot \frac{q_2(u, v)|J(u, v)|}{|J(u, v)|}} dudv$$

$$= -\log \iint \sqrt{P_1(u, v)P_2(u, v)} dudv$$

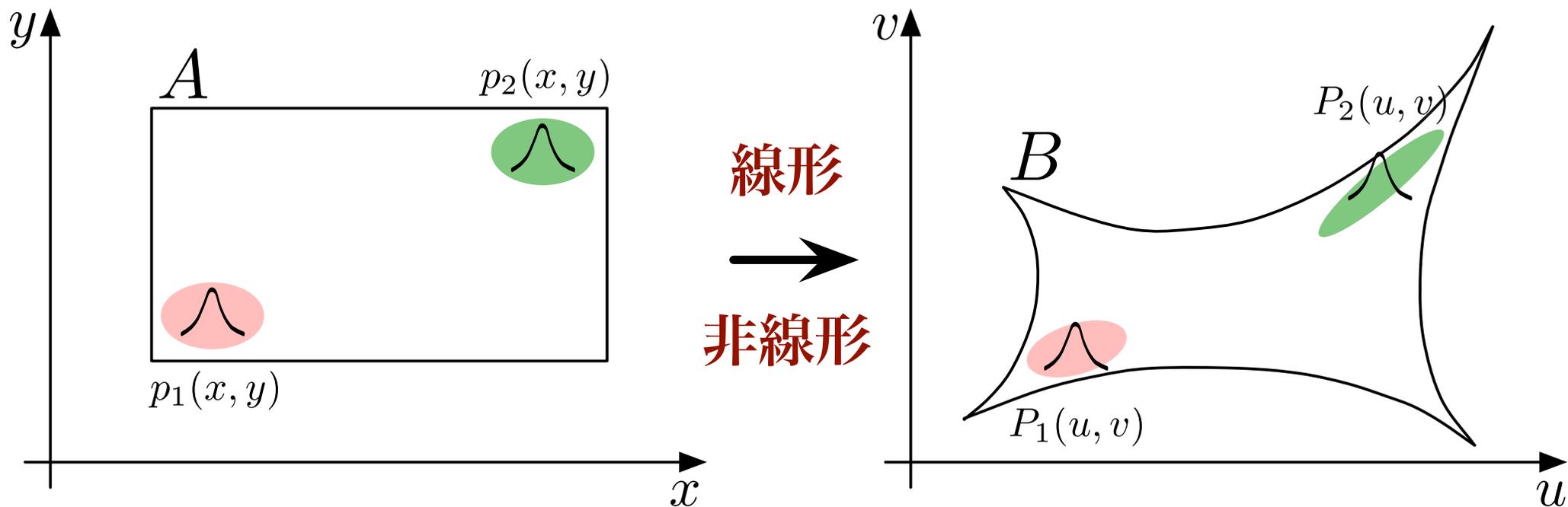
$$= BD(P_1(u, v), P_2(u, v))$$

$$q_1(u, v) = p_1(x(u, v), y(u, v)), \quad J = \text{Jacobian}$$

話者不変な音響量の数学的探求

二人の話者の音声音響空間において不変なる物理量

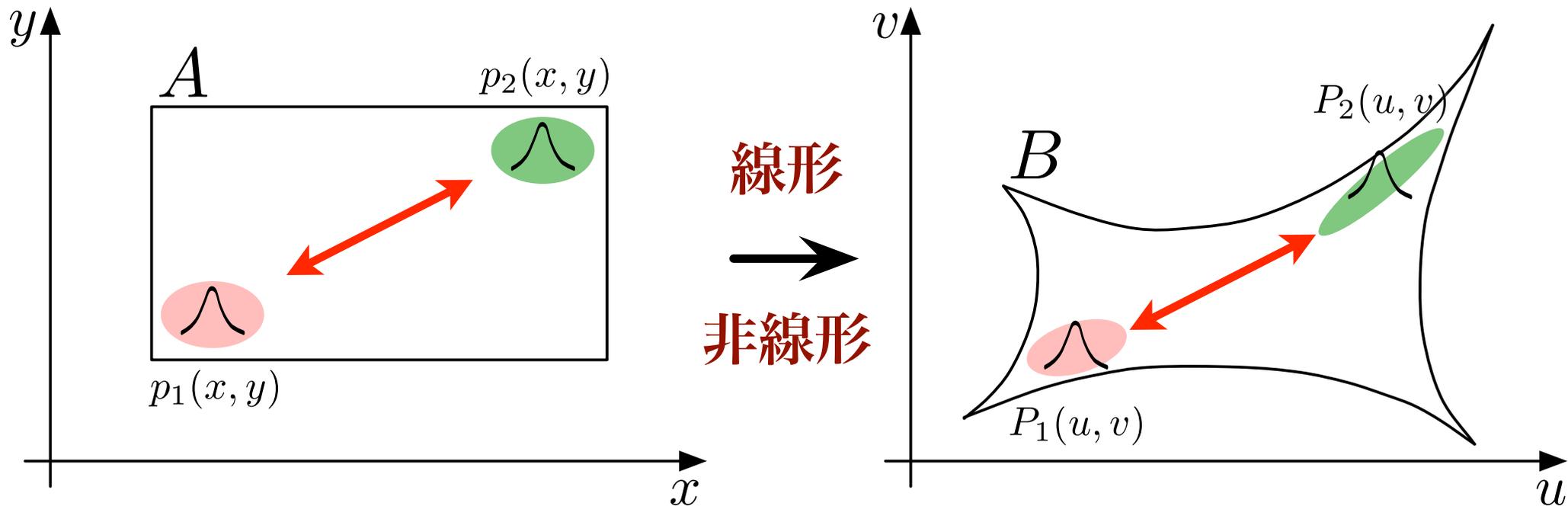
- 全ての音響事象は点ではなく、分布として表現される
 - 空間Aの分布 p は空間Bの分布 P へと写像される (一対一)
 - p と P は異なる物理特性を持つ
 - 両空間において不変な物理量はどこにある？



話者不変な音響量の数学的探求

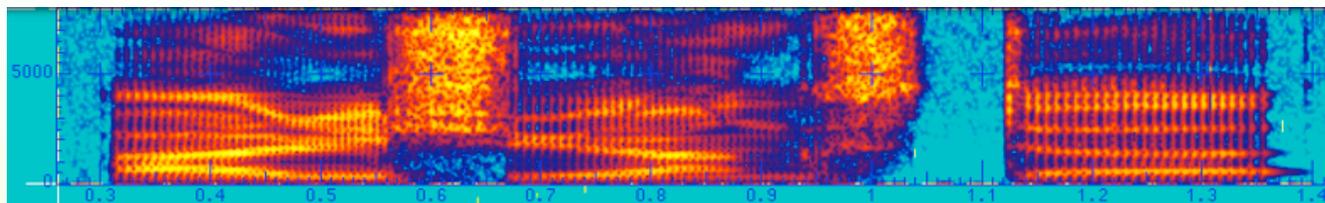
二人の話者の音声音響空間において不変なる物理量

- 全ての音響事象は点ではなく、分布として表現される
 - 空間Aの分布 p は空間Bの分布 P へと写像される (一対一)
 - p と P は異なる物理特性を持つ
 - 両空間において不変な物理量はどこにある？
 - 要素事象は可変, しかし, 要素間距離は不変にならざるをえない

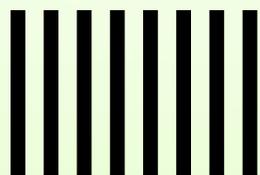
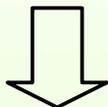


話者不変な音響量の数学的探求

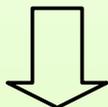
変換（話者・マイク）不変の音声の構造的表象



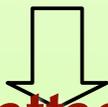
Sequence of spectrum slices



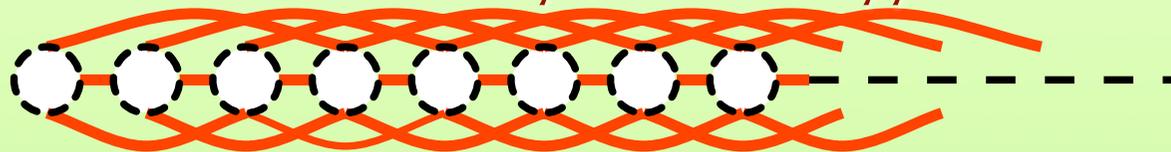
Sequence of cepstrum vectors



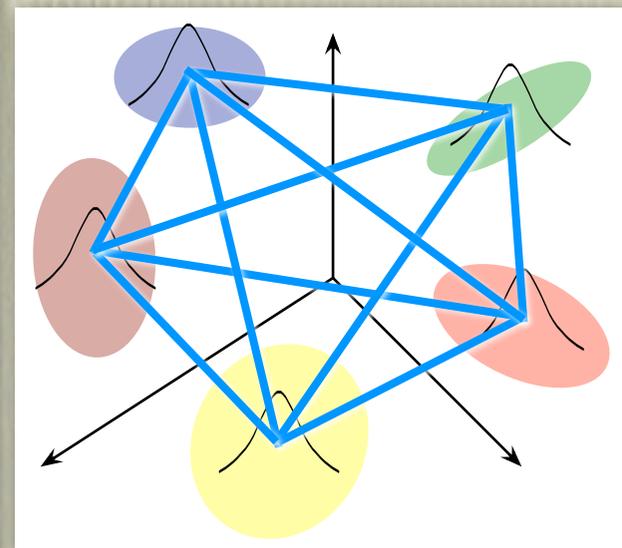
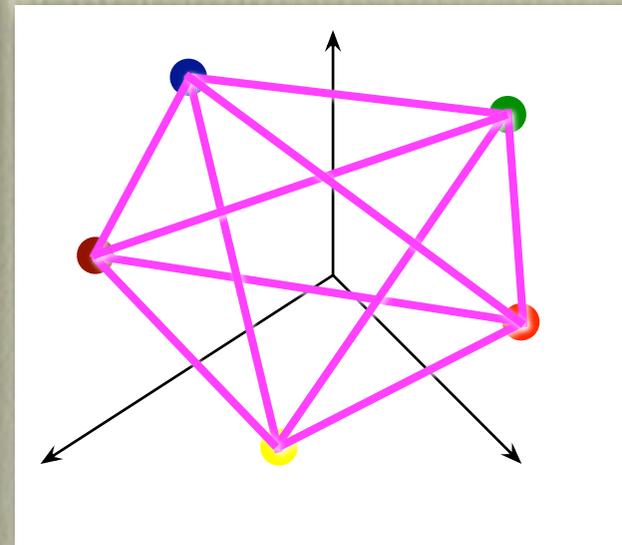
Sequence of distributions



Motions are measured by Bhattacharyya distance.

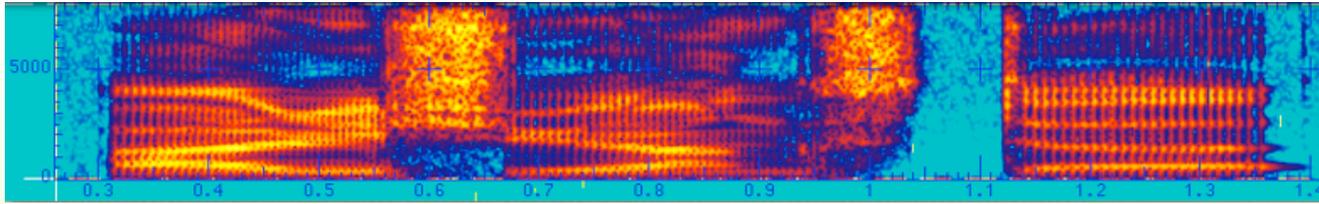


Structuralization by interrelating temporally-distant events

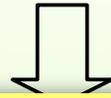


話者不変な音響量の数学的探求

変換（話者・マイク）不変の音声の構造的表象



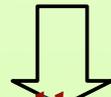
Sequence of spectrum slices



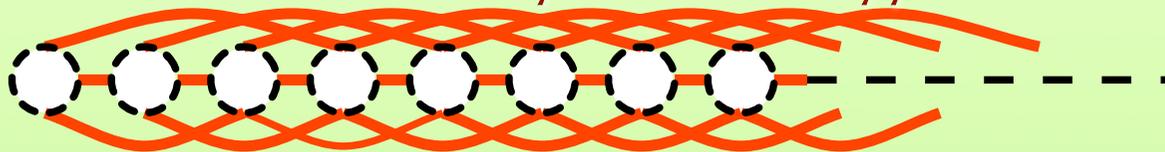
音声 = 音 + 運動

「音」面を捨て、「運動」面のみを抽出

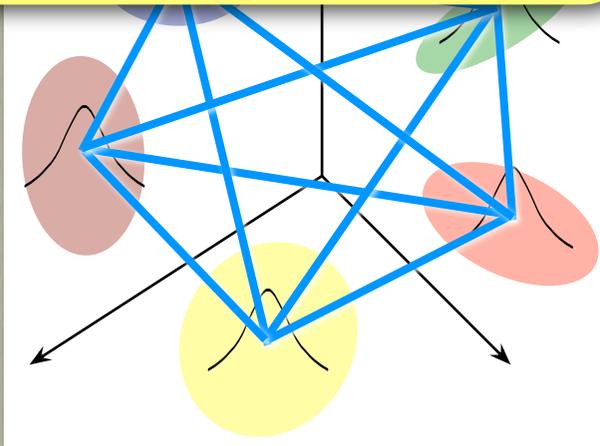
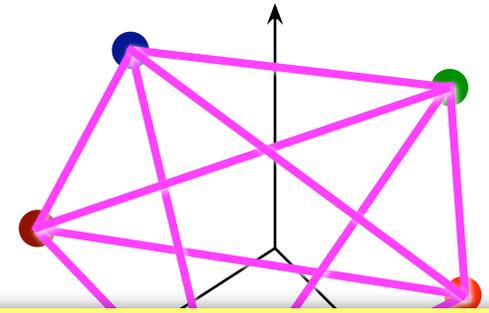
Sequence of distributions



Motions are measured by Bhattacharyya distance.

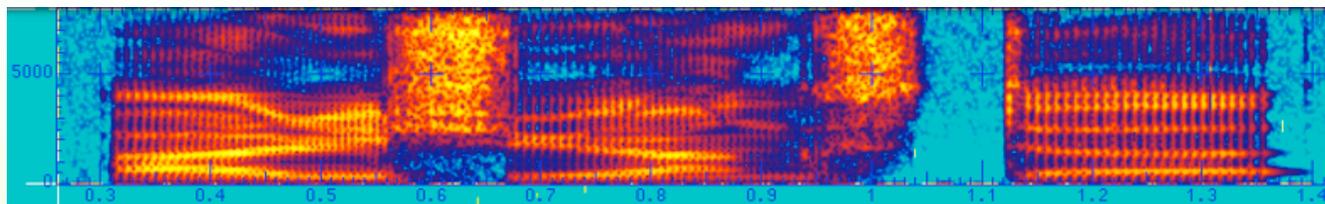


Structuralization by interrelating temporally-distant events

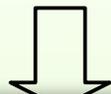


音響的普遍構造

変換（話者・マイク）不変の音声の構造的表象



Sequence of spectrum slices

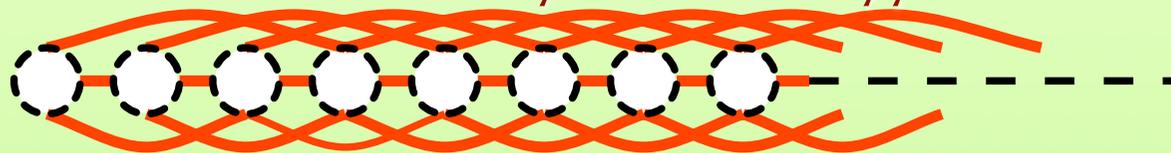


音声 = 音 + 運動

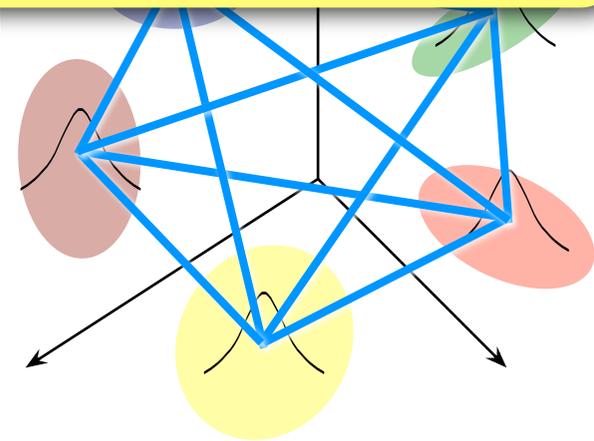
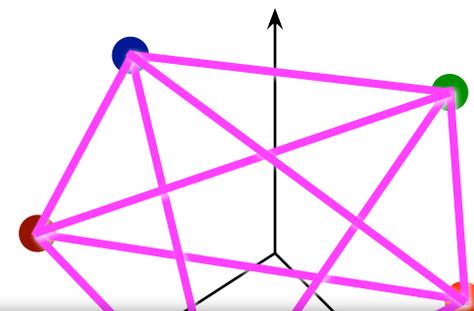
「音」面を捨て、「運動」面のみを抽出

Sequence of distributions

Motions are measured by Bhattacharyya distance.

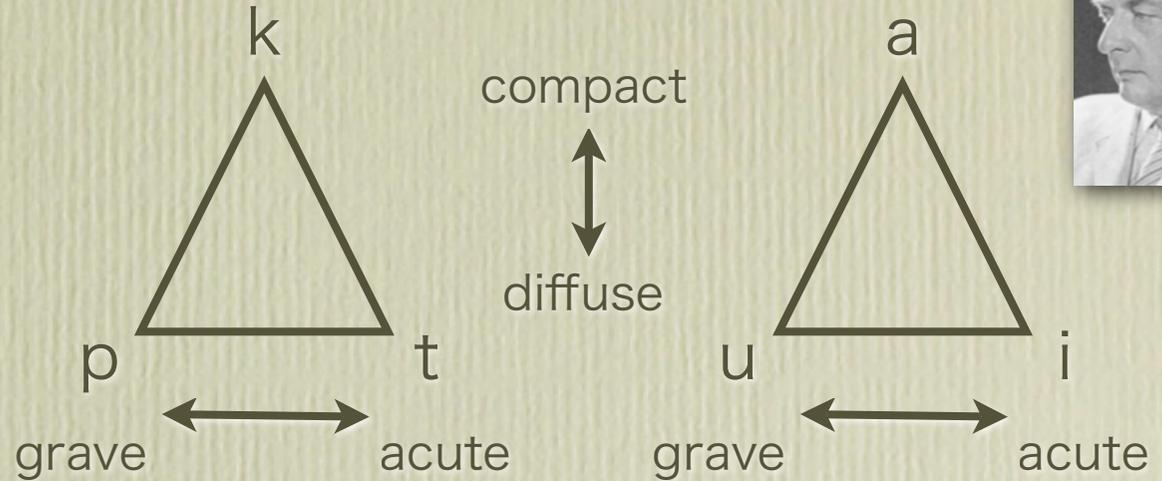
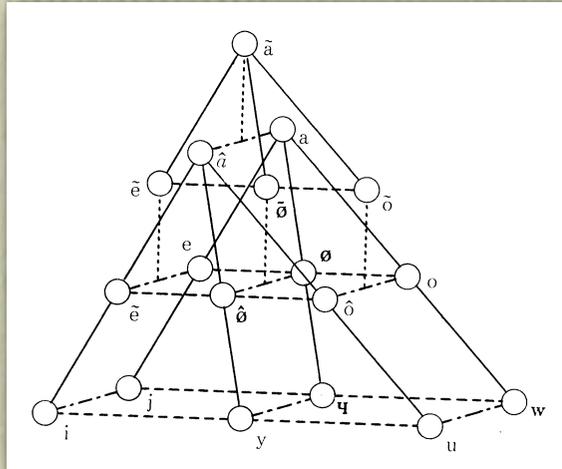


Structuralization by interrelating temporally-distant events



構造音韻論の数学的・物理的実装

構造音韻論 (ヤコブソン)



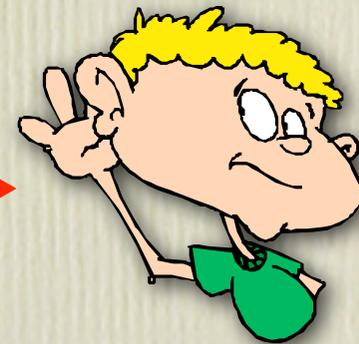
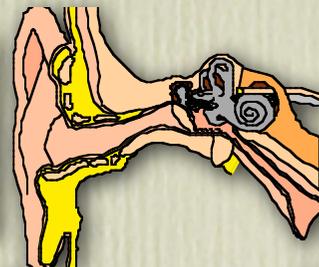
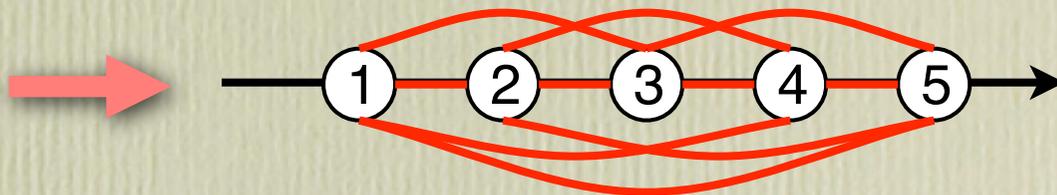
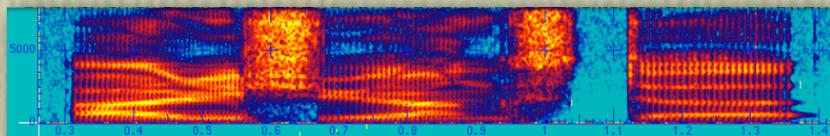
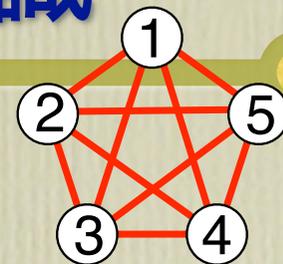
近代言語学の祖・ソシュールの主張

- Language is a system of only conceptual and **phonic differences**.
- The important thing in the word is **not the sound alone but the phonic differences** that make it possible to distinguish this word from the others.
- Similarity between the claim of Saussure and that of neurosciences
 - Importance of the phonic **differences** or **contrasts**
 - Importance of the speech **motions** or **changes**



音響的普遍構造を用いた音声認識

構造（話し手） → 音事象系列 → 構造（聞き手）



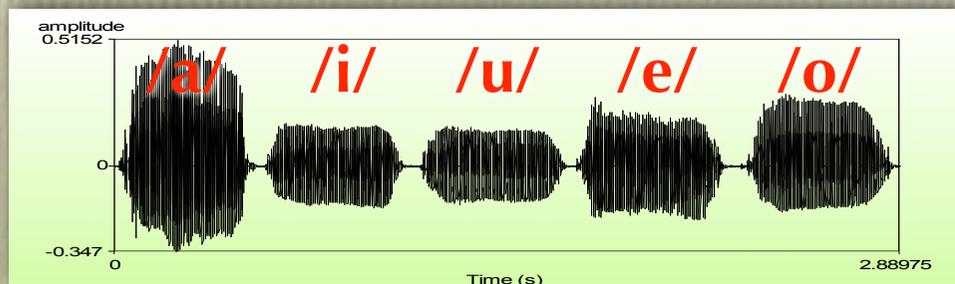
音響的普遍構造を用いた音声認識

認識タスク

- 孤立母音系列 (系列長 = 5) $V_1-V_2-V_3-V_4-V_5$ $V_i \neq V_j$
- V_i = 日本語五母音のいずれか。各母音は一回ずつ発声される
- 語彙数 = $5 \times 4 \times 3 \times 2 \times 1 = 120$, $CL = 1/120 = 0.8\%$

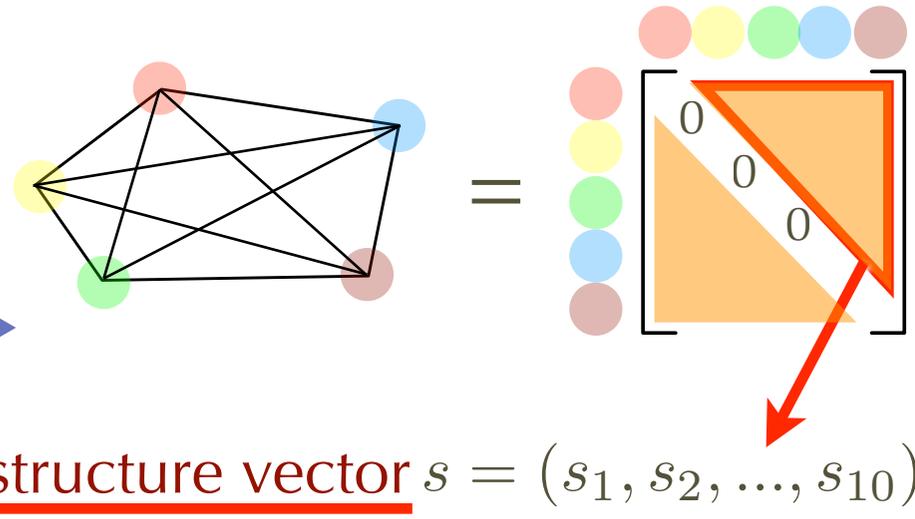
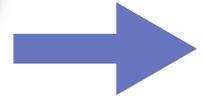
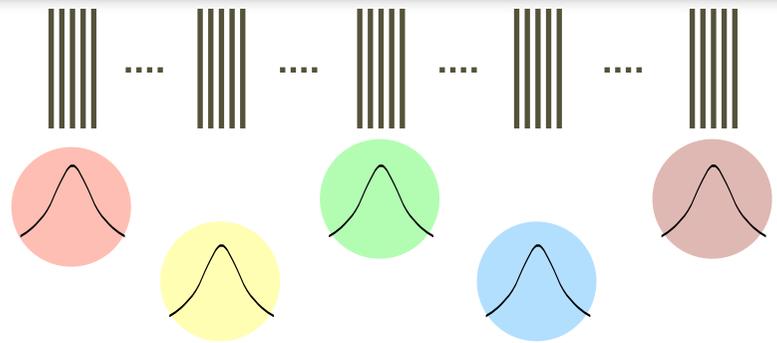
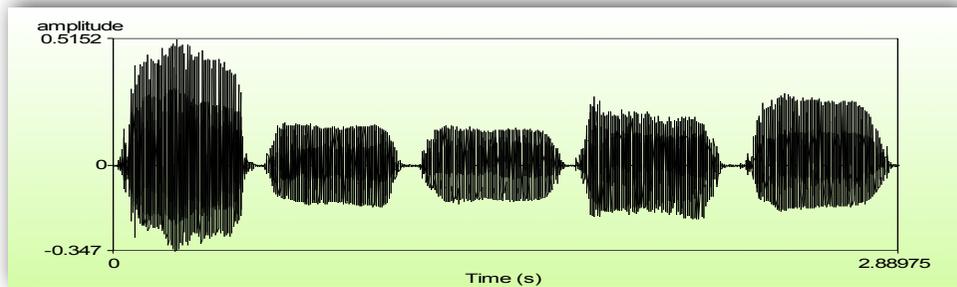
音声の収録

- 評価用音声 = {4男 + 4女} \times {/a/, /i/, /u/, /e/, /o/} \times 5回
 - 各話者, 各系列毎に $5^5 (=3,125)$ サンプルを取得
 - 各系列毎に $25,000 (=3,125 \times 8)$ サンプルを取得
- 学習用音声 = {1男} \times {/a/, /i/, /u/, /e/, /o/} \times 35回
 - 各系列毎に $21,875 (=5^5 \times 7)$ サンプルを取得

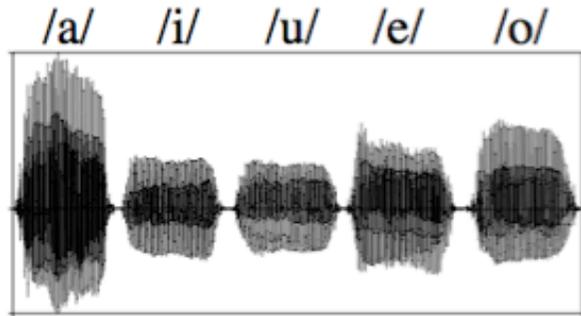


$$\begin{bmatrix} a \\ a \\ a \\ a \\ a \end{bmatrix} \times \begin{bmatrix} i \\ i \\ i \\ i \\ i \end{bmatrix} \times \begin{bmatrix} u \\ u \\ u \\ u \\ u \end{bmatrix} \times \begin{bmatrix} e \\ e \\ e \\ e \\ e \end{bmatrix} \times \begin{bmatrix} o \\ o \\ o \\ o \\ o \end{bmatrix} = 3,125$$

音響的普遍構造を用いた音声認識



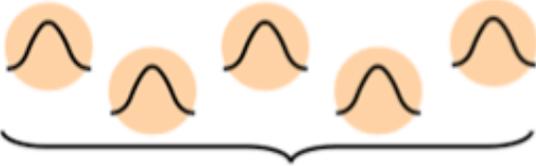
音響的普遍構造を用いた音声認識



CEP sequence



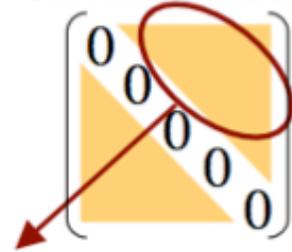
CEP distribution



structure



distance matrix



structure vector

$$S = (s_1, s_2, \dots, s_{10})$$

statistical structure model

/a/-/i/-/u/-/e/-/o/



/a/-/i/-/u/-/o/-/e/



•

•

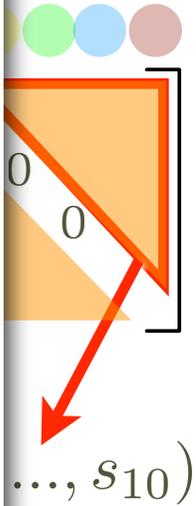
•

•

•

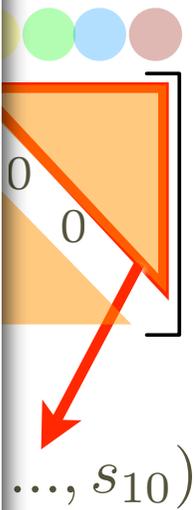
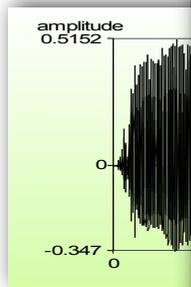
•

/o/-/e/-/u/-/i/-/a/



音響的普遍構造を用いた音声認識

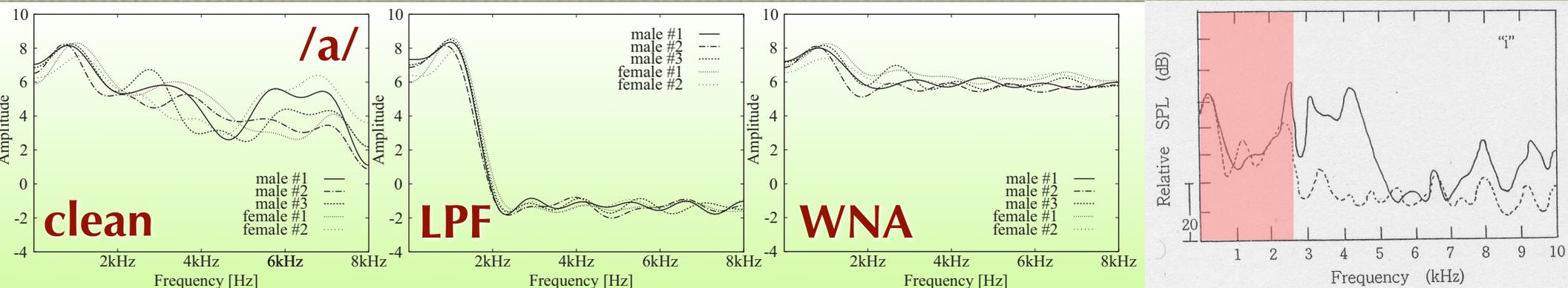
100%



音響的普遍構造を用いた音声認識

前処理の必要性

- 高域に存在する話者性の事前の除去 (LPF or WNA)



認識実験結果

- LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

- WNA の効果

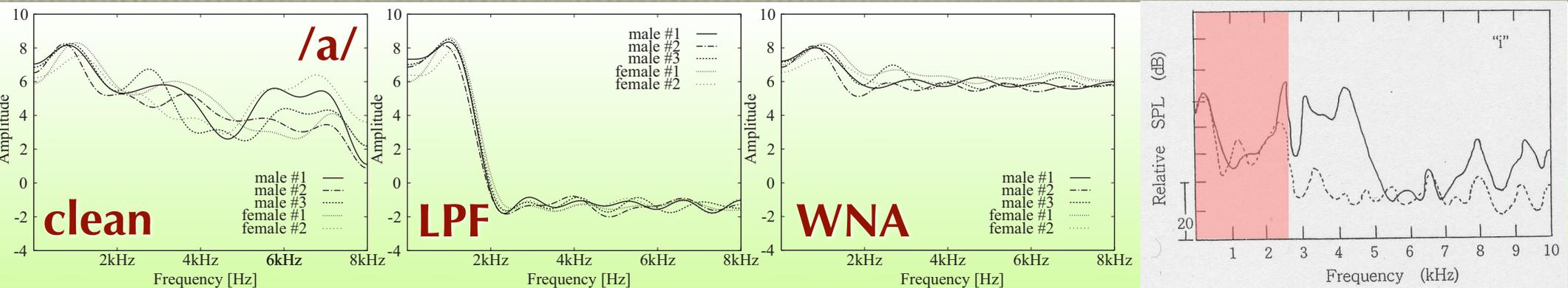
SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

- 4,000人以上の話者から構築したHMMより高い頑健性

音響的普遍構造を用いた音声認識

前処理の必要性

- 高域に存在する話者性の事前の除去 (LPF or WNA)



認識実験結果

- LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

- WNA の効果

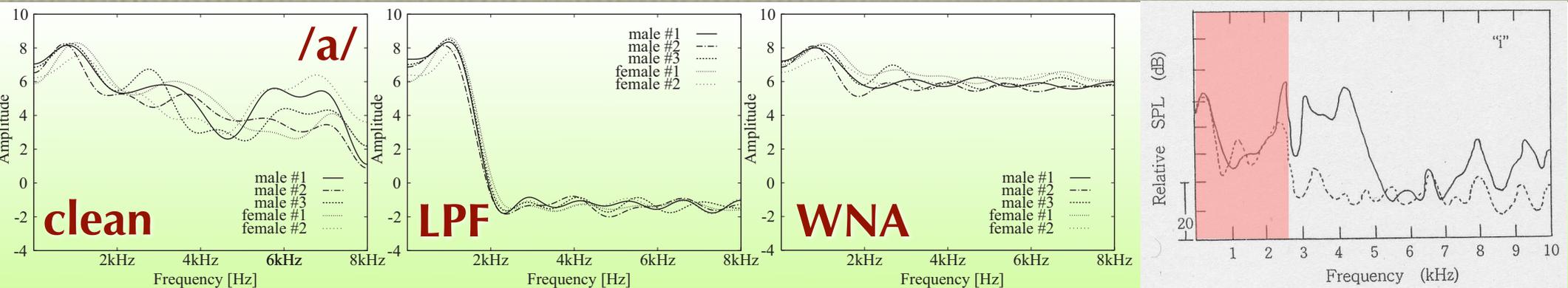
SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

- 4,000人以上の話者から構築したHMMより高い頑健性

音響的普遍構造を用いた音声認識

前処理の必要性

- 高域に存在する話者性の事前の除去 (LPF or WNA)



認識実験結果

- LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

- WNA の効果

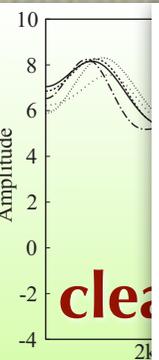
SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

- 4,000人以上の話者から構築したHMMより高い頑健性

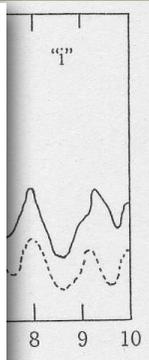
音響的普遍構造を用いた音声認識

前処理の必要性

- 高域に存在する話者性の事前の除去 (LPF or WNA)



methods	#speakers	full-band	tel. band	2kHz LPF
HMM1	4,130	100.0	95.2	87.5
HMM2	260	100.0	93.8	72.3
Proposed	1	100.0	100.0	100.0



認識実験結果

- LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

- WNA の効果

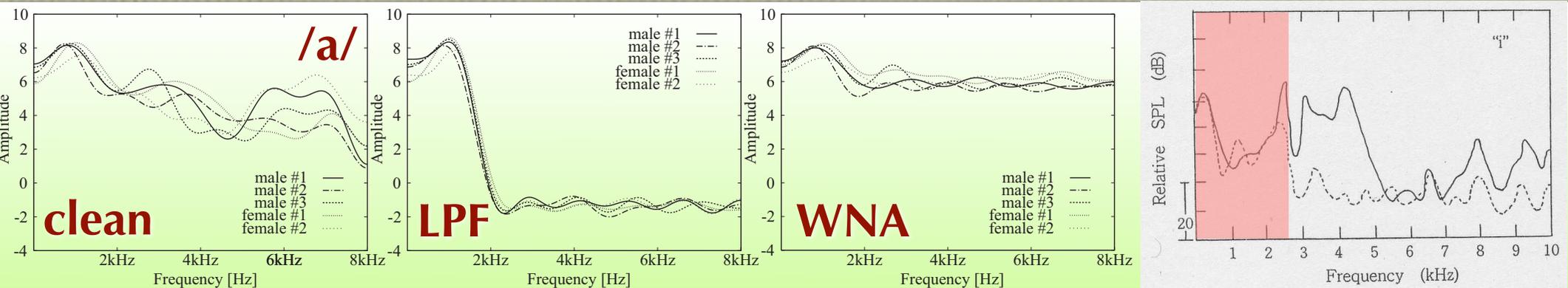
SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

- 4,000人以上の話者から構築したHMMより高い頑健性

音響的普遍構造を用いた音声認識

前処理の必要性

- 高域に存在する話者性の事前の除去 (LPF or WNA)



認識実験結果

- LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

- WNA の効果

SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

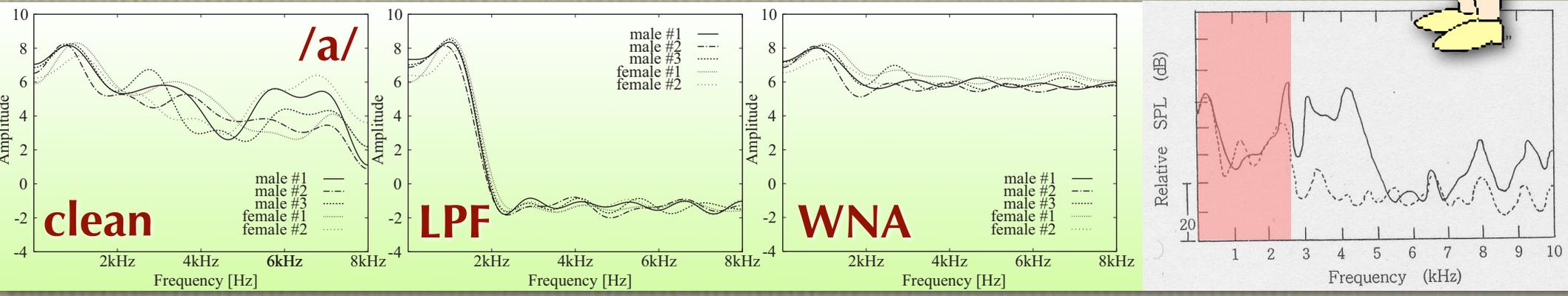
- 4,000人以上の話者から構築したHMMより高い頑健性

音響的普遍構造を用いた音声認識



前処理の必要性

● 高域に存在する話者性の事前の除去 (LPF or WNA)



認識実験結果

● LPF の cut-off と認識率

cut-off [Hz]	8.0	4.0	3.5	3.0	2.5	2.0
accuracy [%]	43.0	62.8	81.8	96.9	80.0	100.0

● WNA の効果

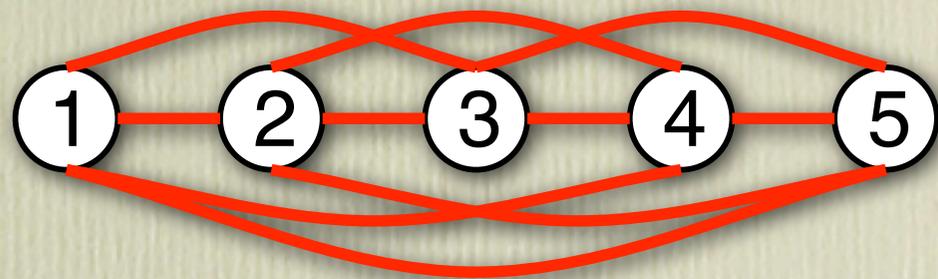
SN ratio [dB]	∞	20.0	10.0
accuracy [%]	70.3	92.9	99.1

● 4,000人以上の話者から構築したHMMより高い頑健性

音響的普遍構造を通して行う考察

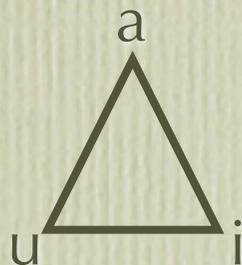
単語の同定が出来れば、音の同定など必要ない

- 音声の構造的&全体的表象を用いた音声認識
- 音響音を音シンボル（ラベル）として一切同定しない単語同定
- 音声を**音楽**として聞く？



音素って何だろう？ ～二種類の音素定義～

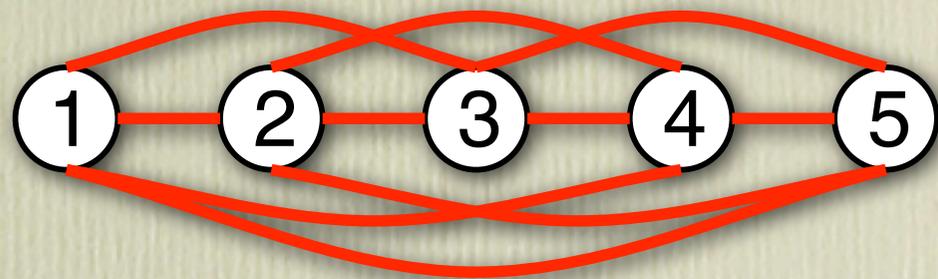
- A phoneme is a class of phonetically-similar sounds. **(absolute)**
- A phoneme is an element in the sound system of a language having characteristic interrelations to the other ones. **(relative)**



音響的普遍構造を通して行う考察

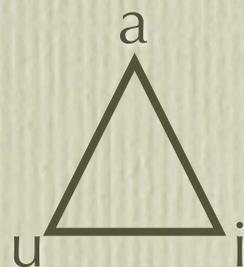
単語の同定が出来れば、音の同定など必要ない

- 音声の構造的&全体的表象を用いた音声認識
- 音響音を音シンボル（ラベル）として一切同定しない単語同定
- 音声を**音楽**として聞く？



音素って何だろう？ ～二種類の音素定義～

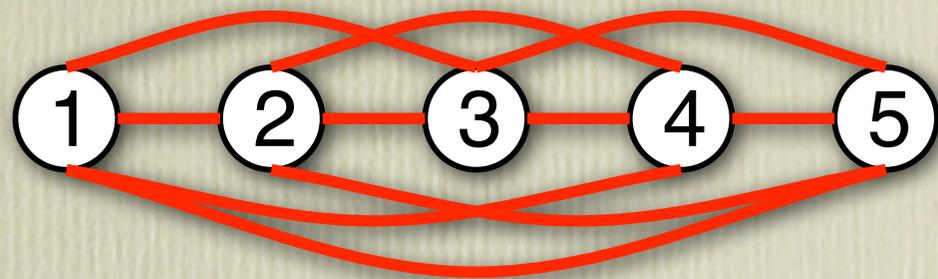
- A phoneme is a class of phonetically-similar sounds. **(absolute)**
- A phoneme is an element in the sound system of a language having characteristic interrelations to the other ones. **(relative)**



音響的普遍構造を通して行う考察

単語の同定が出来れば、音の同定など必要ない

- 音声の構造的&全体的表象を用いた音声認識
- 音響音を音シンボル（ラベル）として一切同定しない単語同定
- 音声を**音楽**として聞く？



音素って何だろう？ ～二種類の音素定義～

- A phoneme is a class of phonetically-similar sounds. **(absolute)**
- A phoneme is an element in the sound system of a language having characteristic interrelations to the other ones. **(relative)**

音名

階名



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす



● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下屬音, 屬音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす



● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下屬音, 屬音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす



● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下屬音, 屬音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー



● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー

● ソーミソドー ラードドソー ソードレミーレドレー (絶対音感)

● レーシレソー ミーソソレー レーソラシーラソラー (絶対音感)



● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー

● ソーミソドー ラードドソー ソードレミーレドレー (絶対音感)

● レーシレソー ミーソソレー レーソラシーラソラー (絶対音感)

● ソーミソドー ラードドソー ソードレミーレドレー (相対音感)

● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー

● ソーミソドー ラードドソー ソードレミーレドレー (絶対音感)

● レーシレソー ミーソソレー レーソラシーラソラー (絶対音感)

● **ソーミソドー ラードドソー ソードレミーレドレー (相対音感)**

● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー

● ソーミソドー ラードドソー ソードレミーレドレー (絶対音感)

● レーシレソー ミーソソレー レーソラシーラソラー (絶対音感)

● **ソーミソドー ラードドソー ソードレミーレドレー (相対音感)**

● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

● 楽曲をドレミに落とす

● ラーララー ラーララー ラーララーララー

● ソーミソドー ラードドソー ソードレミーレドレー (絶対音感)

● レーシレソー ミーソソレー レーソラシーラソラー (絶対音感)

● ● ソーミソドー ラードドソー ソードレミーレドレー (相対音感)

● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ

● 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,

● 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド

● ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…

● 長調の主音＝ド, 短調の主音＝ラ

● 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる

● 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…

● 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

全体から攻めて、個々の要素を同定する方法論

要素の物理的等価性に基づく同一性認知

要素の機能的相対的等価性に基づく同一性認知

- 2種類の音の命名：音名（固定ド）と階名（移動ド）
 - 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ
 - 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,
 - 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド
 - ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下屬音, 屬音…
 - 長調の主音＝ド, 短調の主音＝ラ
 - 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる
 - 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…
 - 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

多様性に対して超頑健な音列刺激の同定技術

要素の物理的等価性に基づく同一性認知

要素の機能的相対的等価性に基づく同一性認知

- 2種類の音の命名：音名（固定ド）と階名（移動ド）
 - 音名：音の絶対的&物理的特性に対する命名＝ハニホヘトイロハ
 - 音名は言語によって異なる。独語＝CDEFGAH, 伊語＝Do,Re,Mi,,,,
 - 階名：その音の曲に対する機能に対する命名＝ドレミファソラシド
 - ド, レ, ミ, ファ, ソ…＝主音, 上主音, 中音, 下主音, 属音…
 - 長調の主音＝ド, 短調の主音＝ラ
 - 曲全体（調性）の把握があって初めて、個々の音の機能は感覚できる
 - 如何なる鍵盤＝ド／ある鍵盤＝ドレミファ…
 - 単独提示された音に階名は付与できない



音響的普遍構造を通して行う考察

多様性に対して超頑健な音列刺激の同定技術

要素の物理的等価性に基づく同一性認知

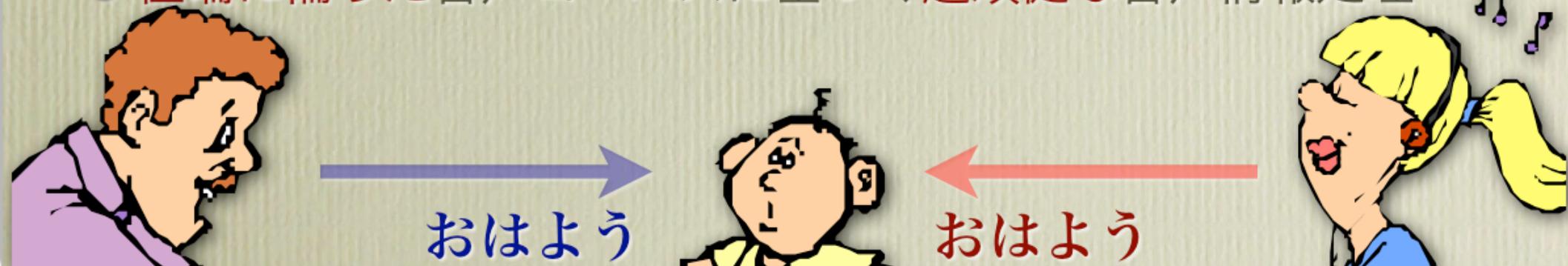
要素の機能的相対的等価性に基づく同一性認知

● 2種類の音の命名：音名（固定ド）と階名（移動ド）

● 幼児の音声音響的環境 ~another POS~

● 大部分は母親と父親の音声

● 極端に偏った音声コーパスに基づく超頑健な音声情報処理

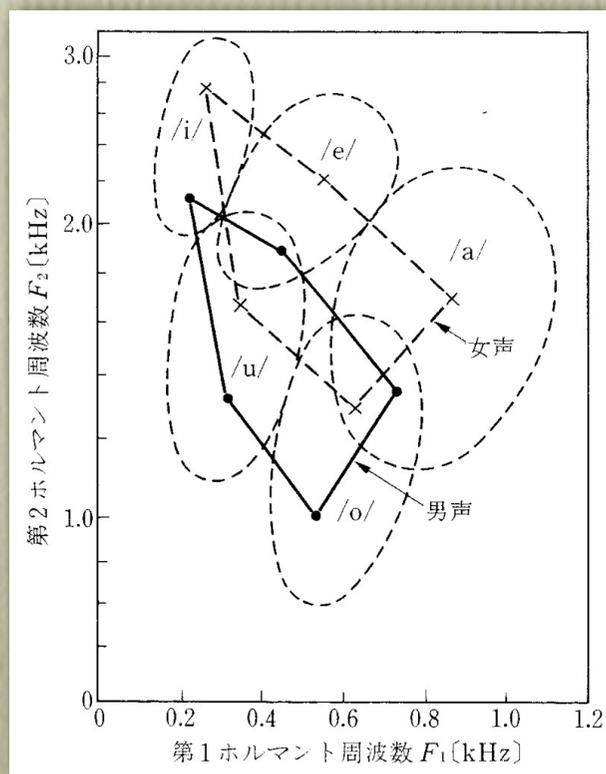


音響的普遍構造を通して行う考察



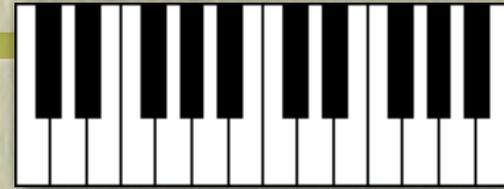
階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
 - 「音シンボル：物理実体＝1：1」が成立するのか否か
 - 成立→音シンボル＝音名的， 不成立→音シンボル＝階名的



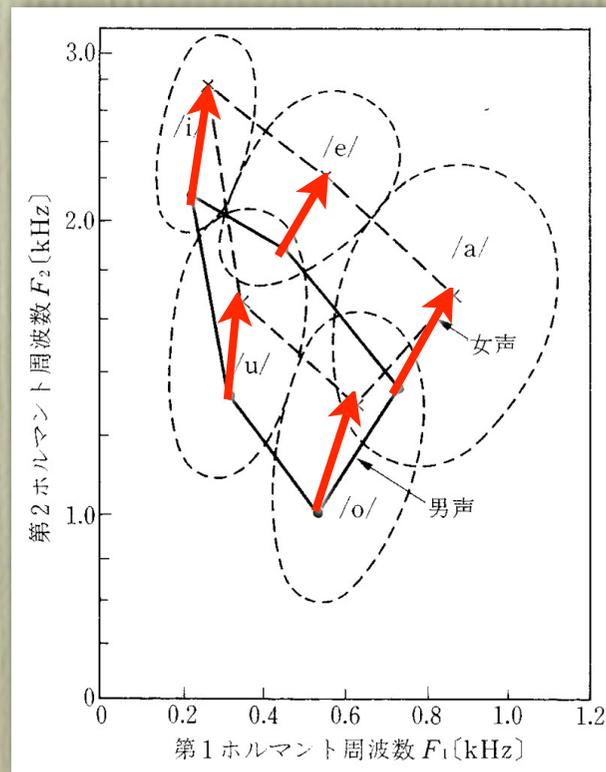
- 日本語＝絶対音感（音名）的， 英語＝相对音感（階名）的？

音響的普遍構造を通して行う考察



階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
 - 「音シンボル：物理実体 = 1 : 1」が成立するの否か
 - 成立 → 音シンボル = 音名的, 不成立 → 音シンボル = 階名的



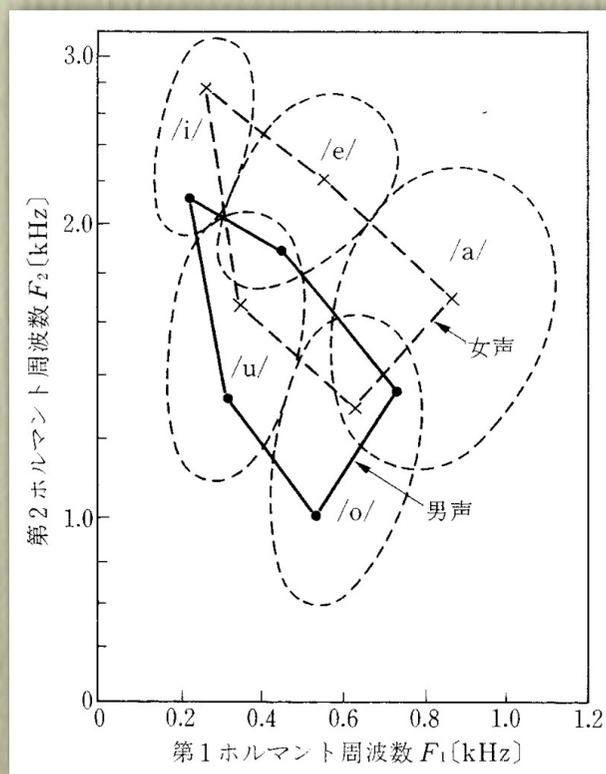
- 日本語 = 絶対音感（音名）的, 英語 = 相対音感（階名）的？

音響的普遍構造を通して行う考察



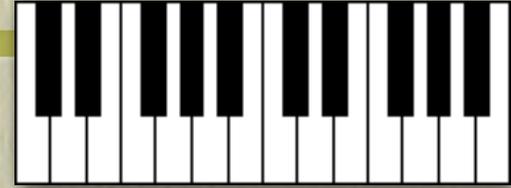
階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
 - 「音シンボル：物理実体＝1：1」が成立するの否か
 - 成立→音シンボル＝音名的， 不成立→音シンボル＝階名的



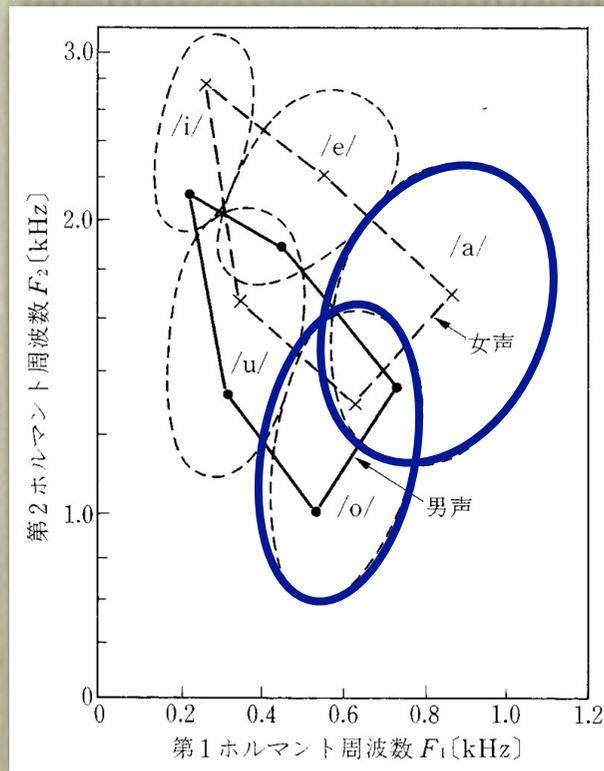
- 日本語＝絶対音感（音名）的， 英語＝相対音感（階名）的？

音響的普遍構造を通して行う考察



階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
- 「音シンボル：物理実体＝1：1」が成立するの否か
- 成立→音シンボル＝音名的， 不成立→音シンボル＝階名的



- 日本語＝絶対音感（音名）的， 英語＝相对音感（階名）的？

音響的普遍構造を通して行う考察

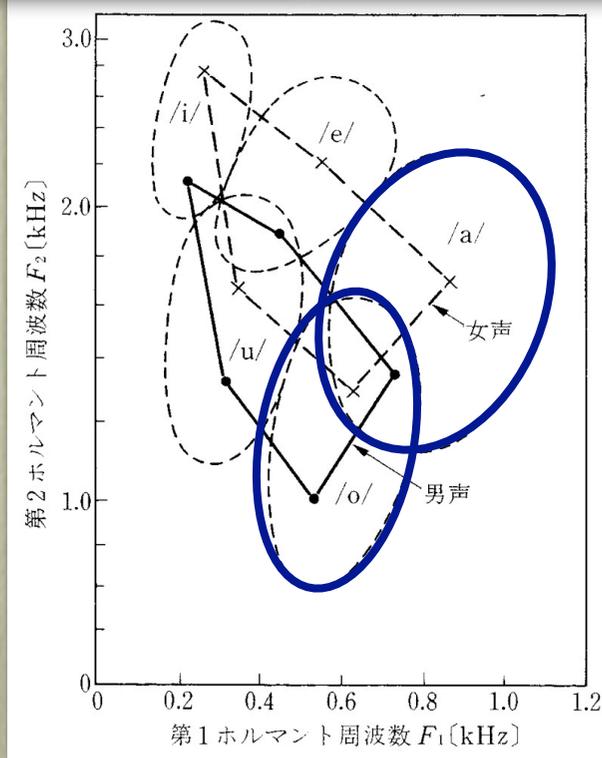


の「あいうえお」

の差異 ～絶対性と相対性～

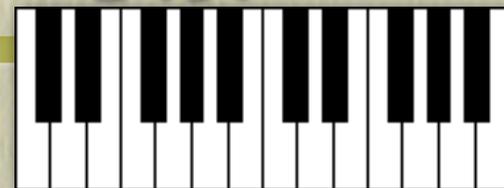
「1 : 1」が成立するのか否か

的, 不成立→音シンボル=階名的



● 日本語 = 絶対音感 (音名) 的, 英語 = 相対音感 (階名) 的?

音響的普遍構造を通して行う考察

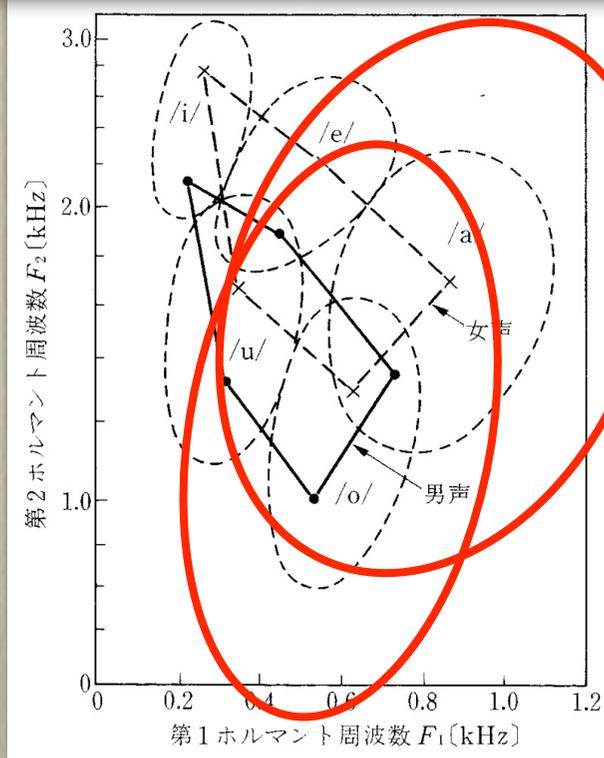


の「あいうえお」

の差異 ～絶対性と相対性～

「1 : 1」が成立するのか否か

的, 不成立→音シンボル=階名的



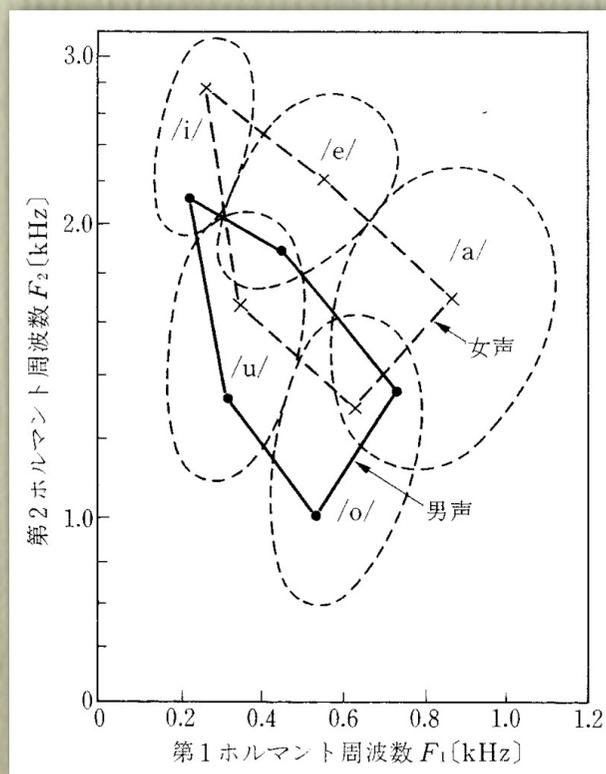
● 日本語 = 絶対音感 (音名) 的, 英語 = 相対音感 (階名) 的?

音響的普遍構造を通して行う考察



階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
- 「音シンボル：物理実体＝1：1」が成立するの否か
- 成立→音シンボル＝音名的， 不成立→音シンボル＝階名的



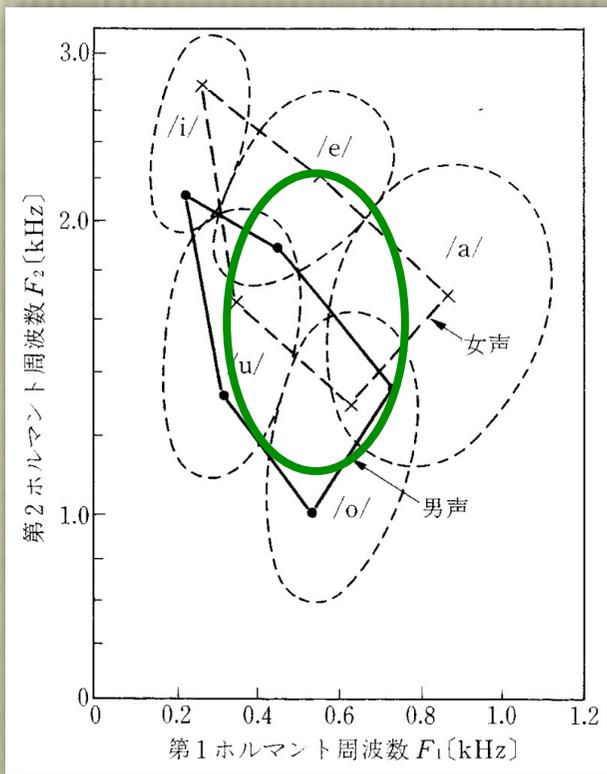
- 日本語＝絶対音感（音名）的， 英語＝相对音感（階名）的？

音響的普遍構造を通して行う考察



階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
 - 「音シンボル：物理実体 = 1 : 1」が成立するの否か
 - 成立 → 音シンボル = 音名的, 不成立 → 音シンボル = 階名的



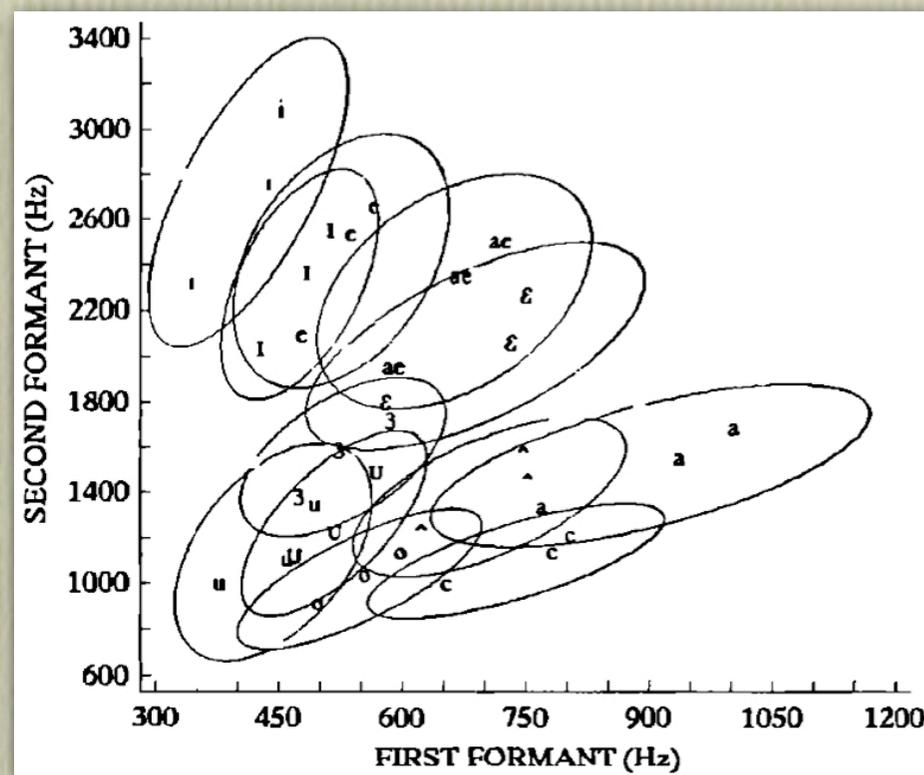
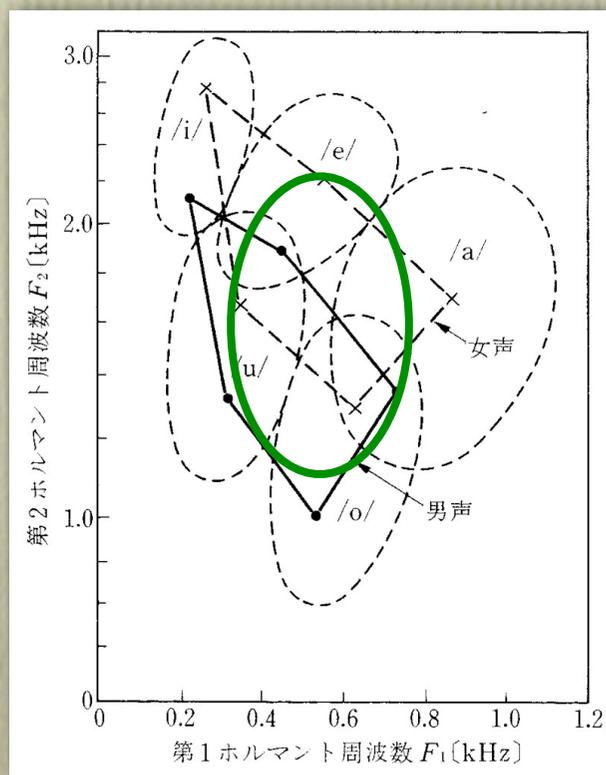
- 日本語 = 絶対音感（音名）的, 英語 = 相对音感（階名）的？

音響的普遍構造を通して行う考察



階名（相对音感）としての「あいうえお」

- 男女における「あいうえお」の差異 ～絶対性と相対性～
- 「音シンボル：物理実体＝1：1」が成立するの否か
- 成立→音シンボル＝音名的， 不成立→音シンボル＝階名的



- 日本語＝絶対音感（音名）的， 英語＝相对音感（階名）的？

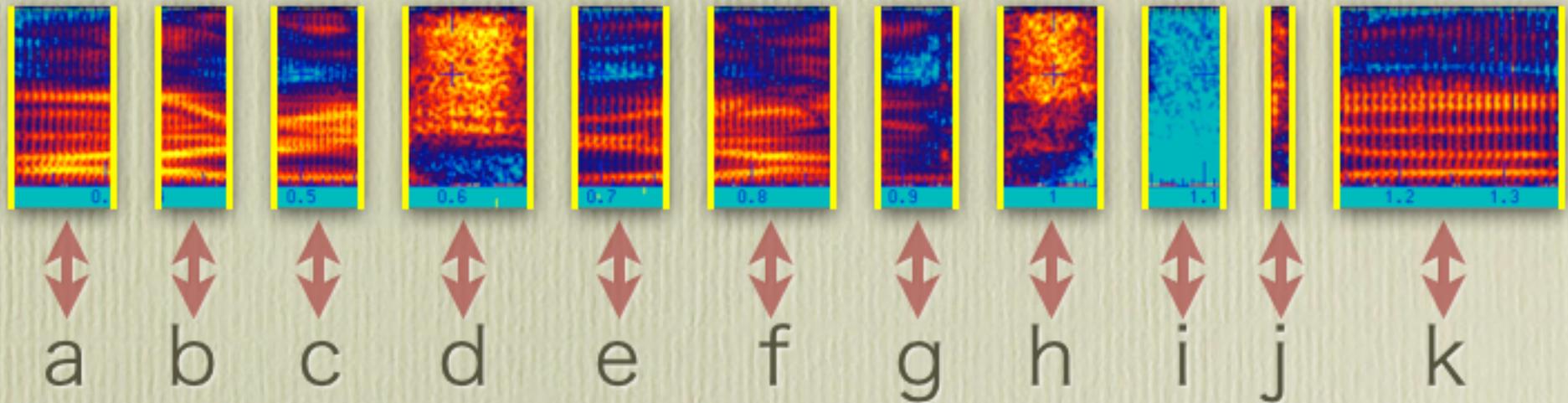
音響的普遍構造を通して行う考察

音響的普遍構造に基づいた単語同定処理の持つ特性

- 幼児の音声認知がそのまま引き継がれている
 -
- 語発声はそれで一つである。個々の音に分割することが困難
 -
- 音を「カテゴリー」として知覚・同定することが困難
 -
- 音と音の差は、健常者以上に感覚することができる
 -
- 音声言語の支障は軽いが、文字言語は非常に困難
 -
- 全体像を捉える、という意味では左脳の優位性は低い
 -
- 知的レベルが高い人がいても不思議ではない

音響的普遍構造を通して行う考察

音声 = 音シンボルの並び ~音声学の常識~



- 音と音の差は，健常者以上に感覚することができる
- 音声言語の支障は軽いが，文字言語は非常に困難
- 全体像を捉える，という意味では左脳の優位性は低い
- 知的レベルが高い人がいても不思議ではない

音響的普遍構造を通して行う考察

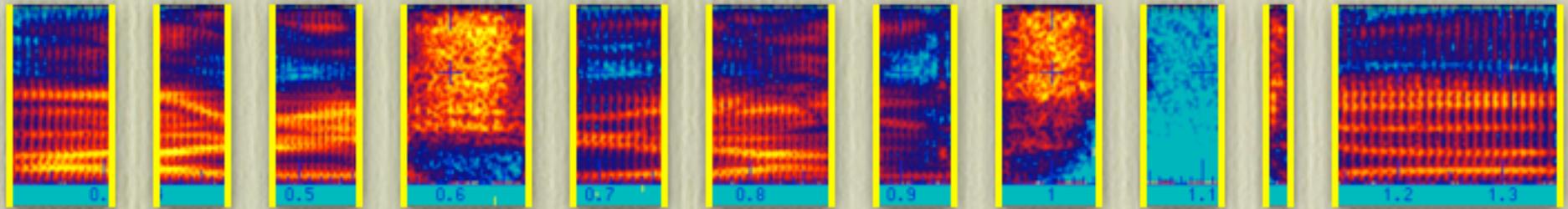
音響的普遍構造に基づいた単語同定処理の持つ特性

- 幼児の音声認知がそのまま引き継がれている
 - 単語の分節化は4歳より。これが生じない（「Dyslexiaのすべて」）
- 語発声はそれで一つである。個々の音に分割することが困難
 - 失読症者が持つ最も共通した特性（「Dyslexiaのすべて」）
- 音を「カテゴリー」として知覚・同定することが困難
 - 範疇的知覚が起こりにくい（Serniclaes'2003）
- 音と音の差は、健常者以上に感覚することができる
 - 例えば異音の音響的差異は検知可能（Serniclaes'2004）
- 音声言語の支障は軽いが、文字言語は非常に困難
 - 頭は良いのに、本が読めない（「Dyslexiaのすべて」）
- 全体像を捉える、という意味では左脳の優位性は低い
 - 言語野の左脳優位性が明確に現れない（多くの脳科学論文より）
- 知的レベルが高い人がいても不思議ではない

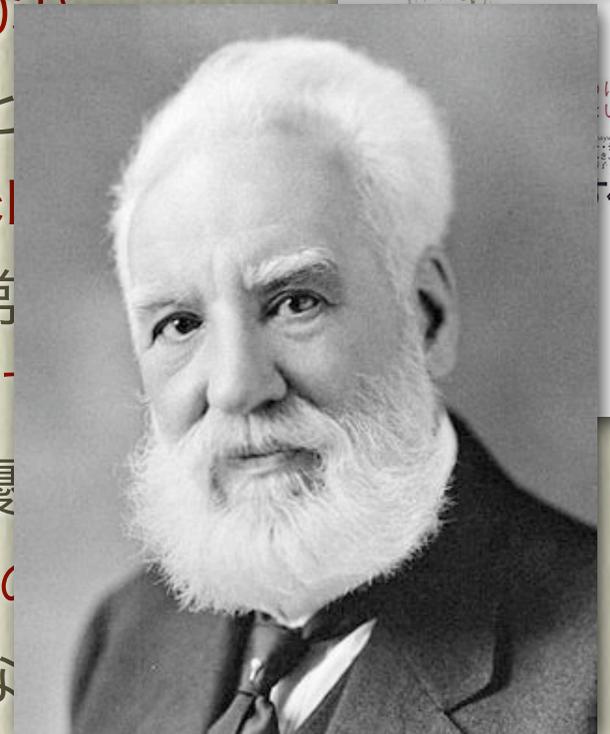


遍て

音声 = 音シンボルの並び ~音声学の常識~



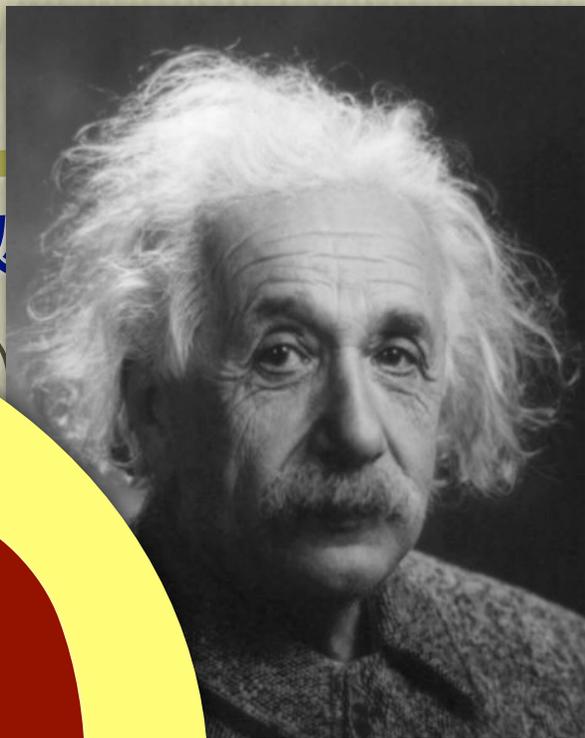
↕ a ↕ b ↕ c ↕ d ↕ e ↕ f ↕ g ↕ h ↕ i ↕ j ↕ k



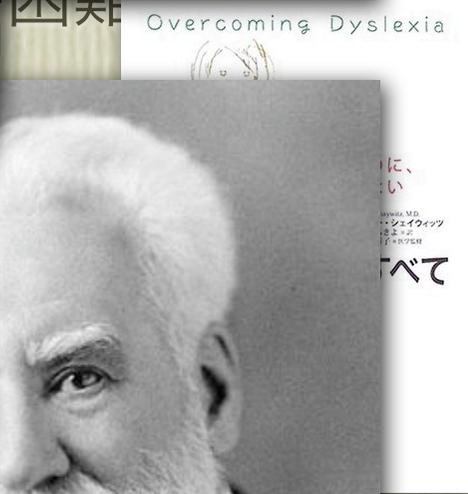
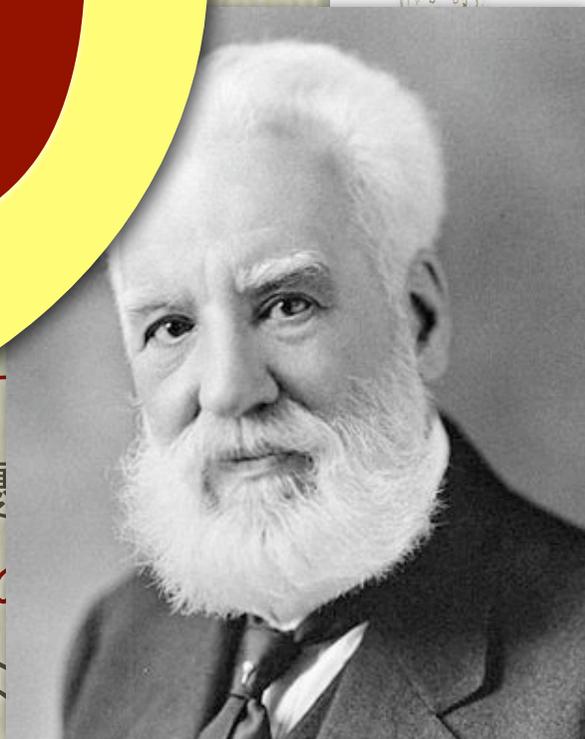
常
音
的
軽
が
と
性
人

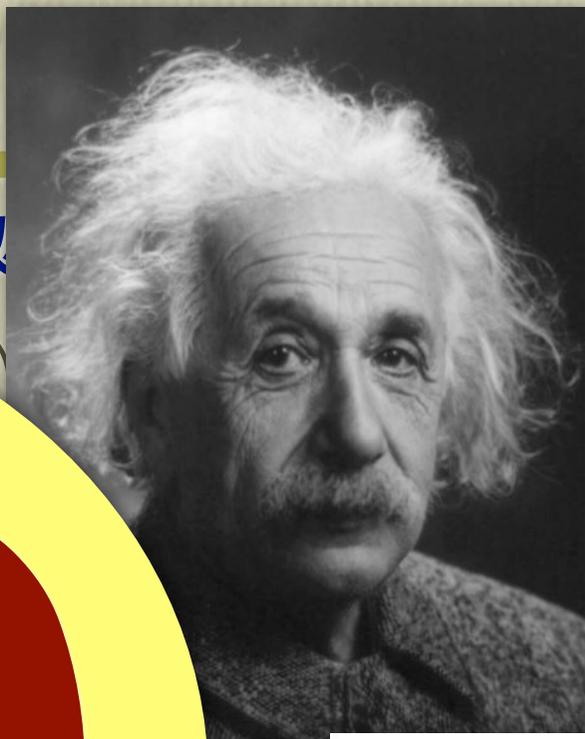
こと
nich
非常
の
優
く
な

に、
い
Hertz, M.D.
ヘンリー・ヘルトツ
1857年12月
1887年

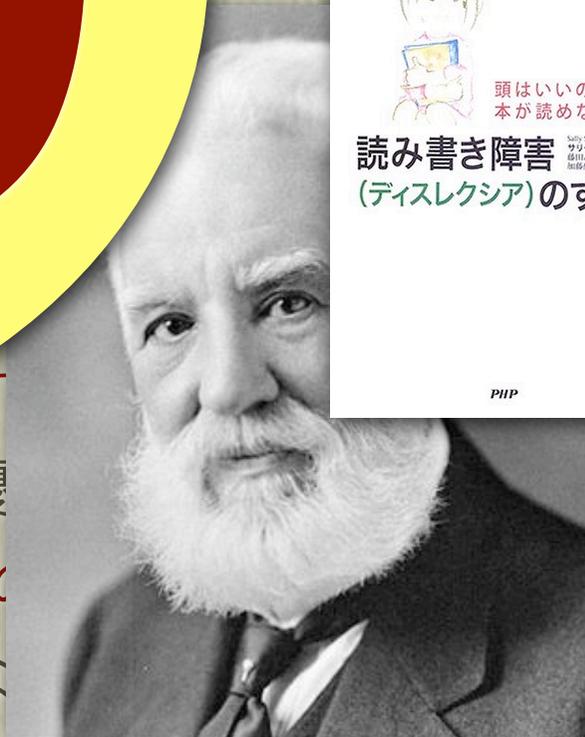


20





20



Overcoming Dyslexia



頭はいいのに、
本が読めない

**読み書き障害
(ディスレクシア)のすべて**

Sally Shaywitz, M.D.
サリー・シェイクワッツ
博士による
最新の「ノウハウ」

PHP

音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。

音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。

音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



音響的普遍構造を通して行う考察

サルの聴覚特性 ～乏しい相対音感～

- オマキザル (D'Amato M. R.'1988)
 - 2旋律の弁別を訓練。移調等価性は観測されず。
- ニホンザル (Izumi A.'2001)
 - 局所的な移調等価性のみ観測された。
 - 全体的特徴と局所的特徴では、後者の方に注意が向く傾向あり。



サルからヒトへ ～違うのに同じもの～

- サルの乏しい相対音感
 - 物理的絶対項への着眼／違う音は違う音
 - その意味において絶対音感者（採譜は当然不可能）
- 「パターンとしての同一性認知」はヒトが持つ能力
 - 「違うもの」を「同一のもの」として感覚すること＝「意識」
 - 逆に「ものの差異」に対する感覚が麻痺してきた。
 - ヒトの場合、乳児は「絶対音感」者。やがて、それが麻痺する。



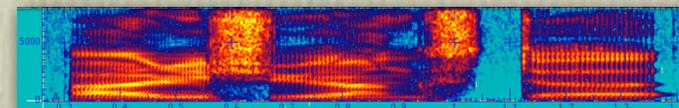
音響的普遍構造を通して行う考察



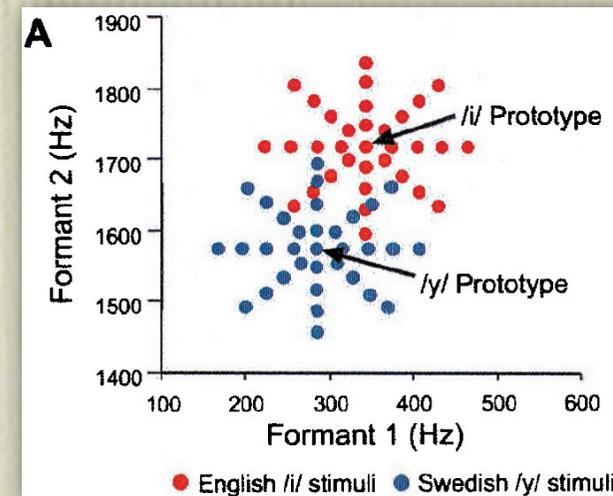
時々受ける質問 ～勘違いの学生～

- 全ての長調の曲が八長調に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - ハ（音名）で終わる＝八長調， ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（ハ）」と捉える勘違い
 - 「ド～」と聞こえるから物理的に同一音のはず、と考える勘違い

音響音声学／音声工学における「あ」

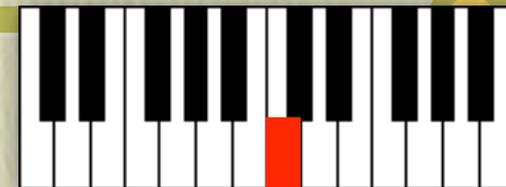


- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



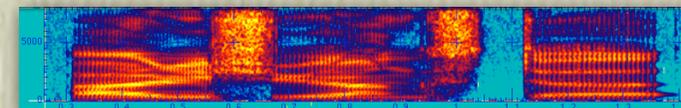
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

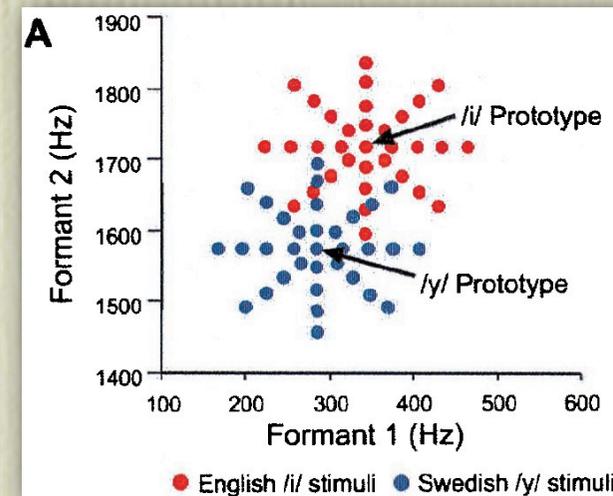


- 全ての長調の曲が**八長調**に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - ハ（音名）で終わる＝八長調， ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（ハ）」と捉える**勘違い**
 - 「ド～」と聞こえるから物理的に同一音のはず、と考える**勘違い**

音響音声学／音声工学における「あ」



- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



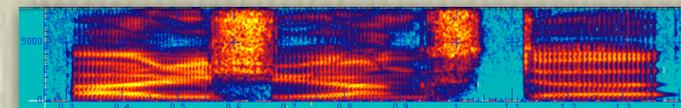
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

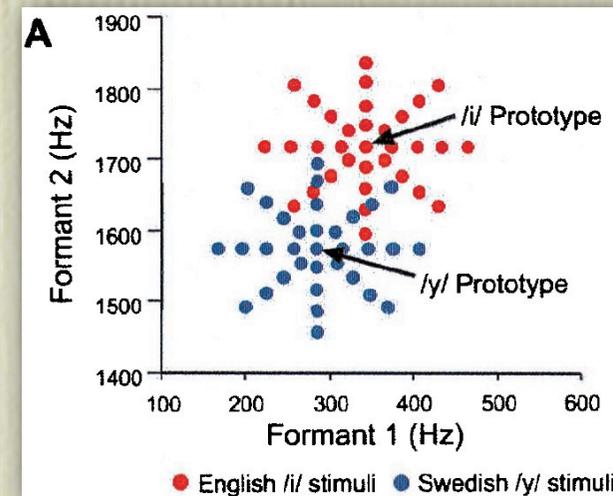


- 全ての長調の曲が**八長調**に聞こえるのですが、
● 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - 八（音名）で終わる＝八長調，ト（音名）で終わる＝ト長調
- 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（八）」と捉える**勘違い**
 - 「ド～」と聞こえるから物理的に同一音のはず、と考える**勘違い**

音響音声学／音声工学における「あ」

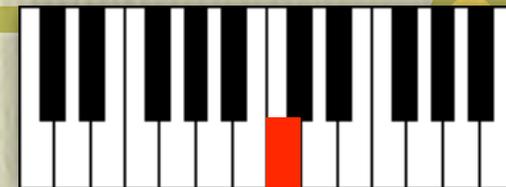


- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



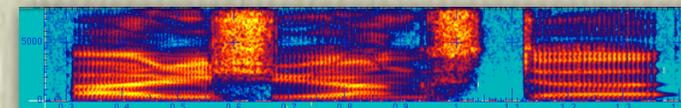
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

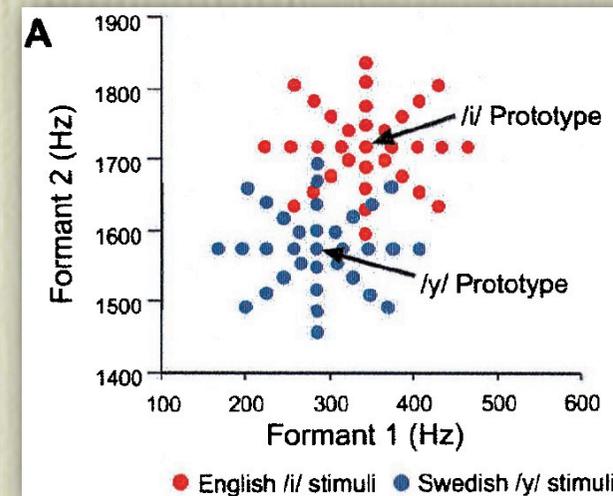


- 全ての長調の曲が**八長調**に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - 八（音名）で終わる＝八長調，ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（八）」と捉える**勘違い**
 - 「ド～」と聞こえるから物理的に同一音のはず、と考える**勘違い**

音響音声学／音声工学における「あ」

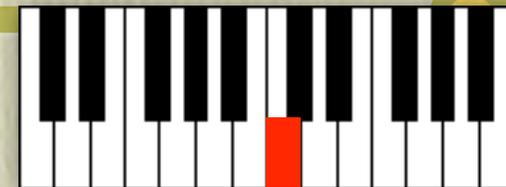


- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



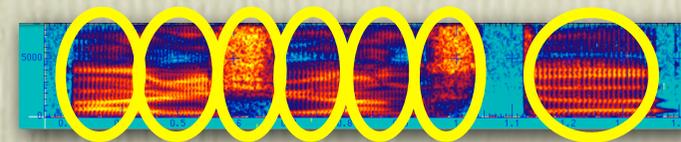
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

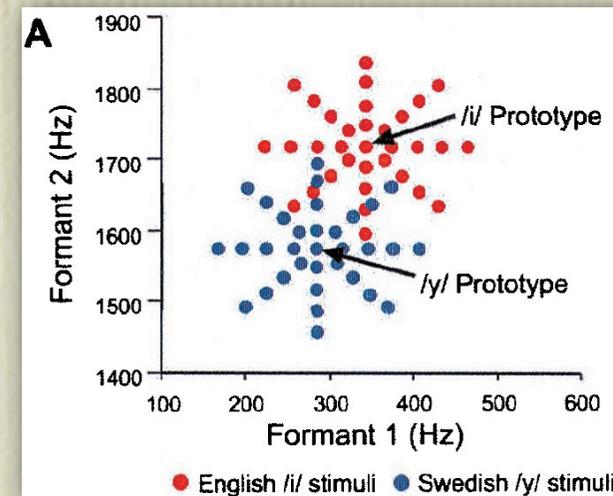


- 全ての長調の曲が**八長調**に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - ハ（音名）で終わる＝八長調， ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（ハ）」と捉える**勘違い**
 - 「ド～」と聞こえるから物理的に同一音のはず、と考える**勘違い**

音響音声学／音声工学における「あ」

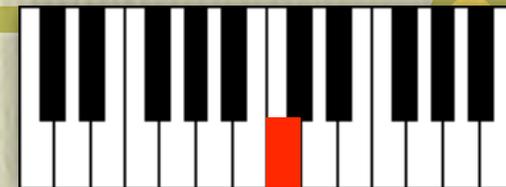


- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



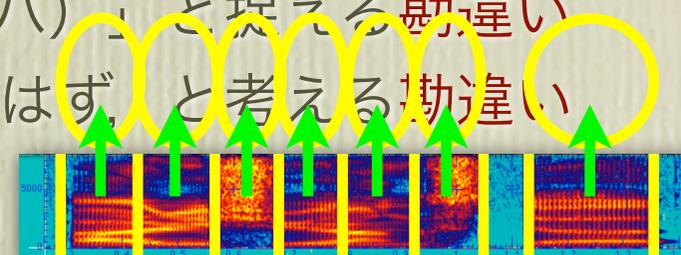
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

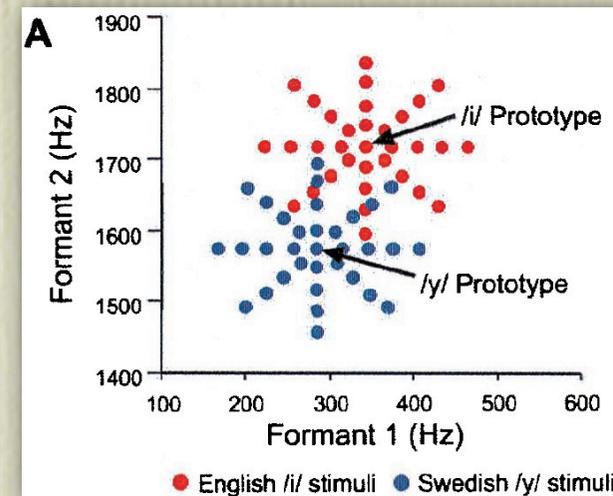


- 全ての長調の曲が**八長調**に聞こえるのですが、
● 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - 八（音名）で終わる＝八長調，ト（音名）で終わる＝ト長調
- 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（八）」と捉える勘違い
 - 「ド～」と聞こえるから物理的に同一音のはずと考える勘違い

音響音声学／音声工学における「あ」

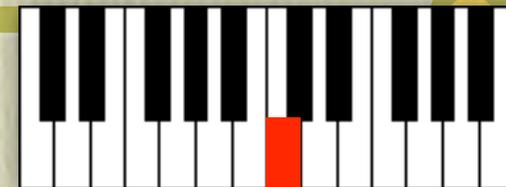


- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



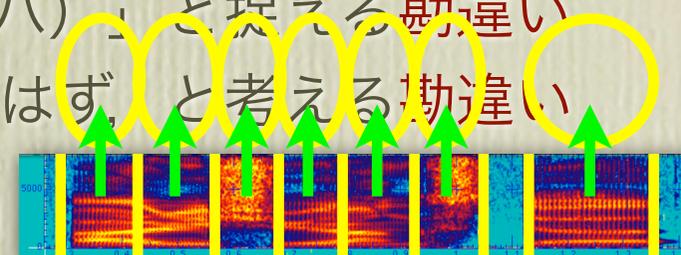
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

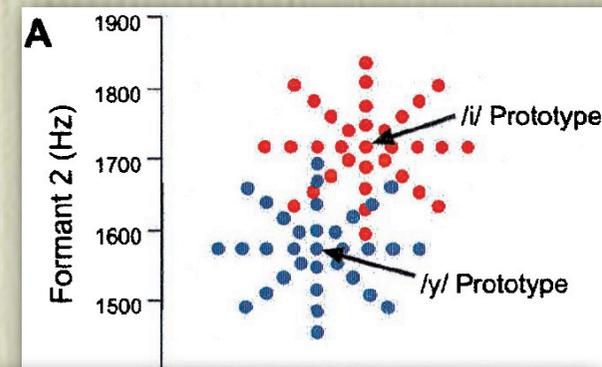


- 全ての長調の曲が八長調に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - ハ（音名）で終わる＝八長調， ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（ハ）」と捉える勘違い
 - 「ド～」と聞こえるから物理的に同一音のはずと考える勘違い

音響音声学／音声工学における「あ」



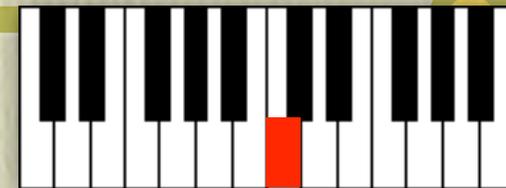
- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



分節音を獲得し、個々の音韻意識を有する研究者の

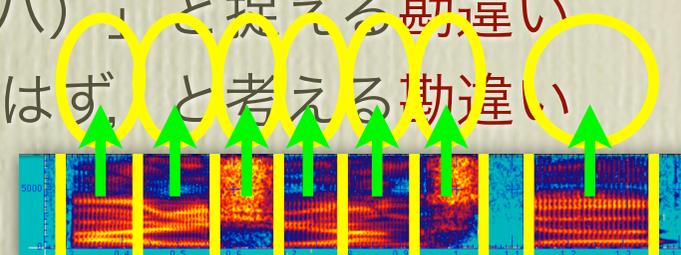
音響的普遍構造を通して行う考察

時々受ける質問 ～勘違いの学生～

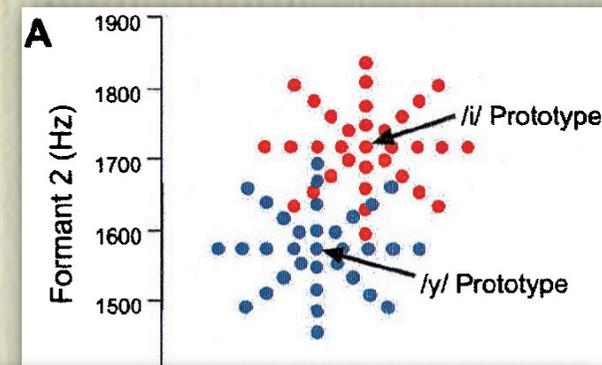


- 全ての長調の曲が八長調に聞こえるのですが、
 - 多くの曲は「主音」で終わる（長調の主音＝階名としてのド）
 - ハ（音名）で終わる＝八長調， ト（音名）で終わる＝ト長調
 - 全ての長調の曲の最後の音が「ド～」と聞こえる
 - 「階名としてのド」を「音名としてのド（ハ）」と捉える勘違い
 - 「ド～」と聞こえるから物理的に同一音のはずと考える勘違い

音響音声学／音声工学における「あ」



- 「あ～」と聞こえるから物理的同一性があるはず
 - 数万人の「あ」の音を集める&分布化
- プロトタイプという考え方
 - 60億の「あ」のうち一番正しい「あ」を定める
 - 人は分節音の前に語全体の音形を獲得する



分節音を獲得し、個々の音韻意識を有する研究者の勘違い

音響的普遍構造を通して行う考察

絶対音感者の憂鬱

- 「基準音 = 440Hz」として記憶した絶対音感者
 - しかし、基準音の物理特性はオーケストラ依存。442Hz, 445Hz, …
- 440Hz 以外の音を「A」として認知することが困難
 - 適応が困難（適応処理ができれば、それは絶対音感ではない）
- 要素的音刺激の絶対的物理特性を記憶する
 - 違う音は違う音 / 異なる音を同一と認知する枠組みが無い

音声の究極の絶対音感者

- 話者Aの[おはよう]と話者Bの[おはよう]の同一性が感覚不可
 - 話者Aの[あ]が「あ」という図形に対応するならば、話者Bの[あ]は決して「あ」ではない。
 - 特定の話者の音声だけ言語メッセージ（文字言語）へ変換可能

音響的普遍構造を通して行う考察

絶対音感者の憂鬱

- 「基準音 = 440Hz」として記憶した絶対音感者
 - しかし、基準音の物理特性はオーケストラ依存。442Hz, 445Hz, …
- 440Hz 以外の音を「A」として認知することが困難
 - 適応が困難（適応処理ができれば、それは絶対音感ではない）
- 要素的音刺激の絶対的物理特性を記憶する
 - 違う音は違う音 / 異なる音を同一と認知する枠組みが無い

音声の究極の絶対音感者

- 話者Aの[おはよう]と話者Bの[おはよう]の同一性が感覚不可
 - 話者Aの[あ]が「あ」という図形に対応するならば、話者Bの[あ]は決して「あ」ではない。
 - 特定の話者の音声だけ言語メッセージ（文字言語）へ変換可能
 - 音声コミュニケーションは間違い無く、困難を極める。

音響的普遍構造を通して行う考察

絶対音感者の憂鬱

- 「基準音 = 440Hz」として記憶した絶対音感者
 - しかし、基準音の物理特性はオーケストラ依存。442Hz, 445Hz, …
- 440Hz 以外の音を「A」として認知することが困難
 - 適応が困難（適応処理ができれば、それは絶対音感ではない）
- 要素的音刺激の絶対的物理特性を記憶する
 - 違う音は違う音 / 異なる音を同一と認知する枠組みが無い

音声の究極の絶対音感者

- 話者Aの[おはよう]と話者Bの[おはよう]の同一性が感覚不可
 - 話者Aの[あ]が「あ」という図形に対応するならば、話者Bの[あ]は決して「あ」ではない。
 - 特定の話者の音声だけ言語メッセージ（文字言語）へ変換可能
 - 音声コミュニケーションは間違い無く、困難を極める。
- 自閉症

音響的普遍構造を通して行う考察

要素刺激の絶対的物理特性の記憶に基づく情報処理

- 要素間の関係ではなく，要素そのものを覚え込む。
 -
 -
 -
- 要素群の関係に基づくパターン・構造の抽出に困難を示す。
 -
 -
 -
 -
- 優れた音感を持つが，音声コミュニケーションが苦手
 -
 -
 -
 -

音響的普遍構造を通して行う考察

要素刺激の絶対的物理特性の記憶に基づく情報処理

- 要素間の関係ではなく、要素そのものを覚え込む。
 - 情報の整理整頓ができず、最終的に散在した形で情報を取得する。
 - 但し、記憶力は良い。関連性の無い数字／単語群を記憶できる。
 - 絶対音感者。ド（C）で始まらないと「カエルの歌」じゃない！
- 要素群の関係に基づくパターン・構造の抽出に困難を示す。
 - 見えるもの／聞こえるものをそのまま受け入れる。抽象化が困難
 - ゲシュタルト知覚が起きにくい。表情の認識が困難。錯覚が起き難い。
 - 抽象的パターンに基づく同一性認知が困難。全てが違うもの＝不安
 - 自己と他者／他者と他者などの関係・役割の理解が困難
- 優れた音感を持つが、音声コミュニケーションが苦手
 - 母親の声しか言語メッセージにならない。特定話者音声認識器
 - あるCDが好き。「このCDのファの音が好き」
 - 文字言語の入力を求める。音声は元々苦手。電話なんて論外
 - 音声から抽出できるのは文字面情報だけ。意図／感情は抽出困難

音響的普遍構造を通して行う考察

要素刺激の絶対的物理特性の記憶に基づく情報処理

- 要素間の関係ではなく、要素そのものを覚え込む。

自閉症者の音感

II

音声認識器の音感

- あるCDが好き。「このCDのファの音が好き」
- 文字言語の入力を求める。音声は元々苦手。電話なんて論外
- 音声から抽出できるのは文字面情報だけ。意図／感情は抽出困難

音響的普遍構造を通して行う考察

ロボット研究と自閉症 ～フレーム問題～

- 些細な環境変化に対する脆弱性
 - ある部屋の情報を全てをプログラムし，ロボットに埋め込む。しかし猫が入ってきただけで，ロボットはパニックを起こす。
 - ある部屋のことを全て自閉症児に教えても，花瓶が「あるべき場所」に無いことで，彼らはパニックを起こす。
- 世界＝多様性の海
 - 世界を「静的なもの」として完全に記述することは不可能
 - 世界＝多様性の海，その多様性の海をいかに乗り切るのか？
 - 環境に潜む不変項＝アフォーダンス（生態心理学）

音響的普遍構造を通して行う考察

ロボット研究と自閉症 ～フレーム問題～

The 4th IEEE International Conference on Development and Learning - From Interaction to Cognition -

Dates: July 19-21, 2005

Location: INTEX Osaka, Japan

[Program is now available!](#)

[Instructions for presentation were updated!](#)

Note: The registration desk will be open from 13:00 on July 19th, 2005.

Note: The on-site registration can be made in Japanese Yen CASH!! ONLY. Please register online by July 8th if you prefer to use your credit card.

The 4th IEEE International Conference on Development and Learning (ICDL-05) will be held in Osaka, Japan at [INTEX Osaka](#), July 19-21, 2005. The theme of ICDL-05, "From Interaction To Cognition," refers not only to the principle of development and learning in natural or artificial systems, but also to the social process of building interdisciplinary bridges among researchers. There will be invited sessions by prominent researchers from different fields that represent the study of learning and development in humans, animals, and automats. Tutorials and workshops are scheduled before the conference. The conference is to be held right after [the Ninth RoboCup International Competitions and Conferences, RoboCup-05](#) (at the same site) and right before [the Fifth International Workshop on Epigenetic Robotics \(EpiRob-05, in Nara\)](#). Several joint sessions/events between these meetings will be organized.

音響的普遍構造を通して行う考察

ロボット研究と自閉症 ～フレーム問題～

The 4th IFFF International Conference on Development and Learning

Workshop & Tutorial

Workshop (Tuesday, July 19th, 9:30-15:00) [workshop page](#)

"Social Cognition: From Humans to Robots"

Confirmed speakers:

Prof. Javier Movellan (UCSD)

Dr. Yukiyasu Kamitani (ATR)

Dr. Thierry Chaminade (ATR/UCL)

Tutorial (Tuesday, July 19th, 9:30-15:00)

"The developing child with autism: evidences, speculations and vexed questions"

Jacqueline Nadel, PhD, Research director

(French National Centre of Scientific Research, L'Hôpital de la Salpêtrière, Paris, France)

conference. The conference is to be held right after [the Ninth RoboCup International Competitions and Conferences, RoboCup-05](#) (at the same site) and right before [the Fifth International Workshop on Epigenetic Robotics \(EpiRob-05, in Nara\)](#). Several joint sessions/events between these meetings will be organized.



音響的普遍構造を通して行う考察

ロボット研究と自閉症 ～フレーム問題～

The 4th IFFF International Conference on Development and Learning

Workshop & Tutorial

Workshop (Tuesday, July 19th, 9:30-15:00) [workshop page](#)

"Social Cognition: From Humans to Robots"

Confirmed speakers:

Prof. Javier Movellan (UCSD)

Dr. Yukiyasu Kamitani (ATR)

Dr. Thierry Chaminade (ATR/UCL)

Tutorial (Tuesday, July 19th, 9:30-15:00)

"The developing child with autism: evidences, speculations and vexed questions"

Jacqueline Nadel, PhD, Research director

(French National Centre of Scientific Research, L'Hôpital de la Salpêtrière, Paris, France)

conference. The conference is to be held right after [the Ninth RoboCup International Competitions and Conferences, RoboCup-05](#) (at the same site) and right before [the Fifth International Workshop on Epigenetic Robotics \(EpiRob-05, in Nara\)](#). Several joint sessions/events between these meetings will be organized.



シンボル・グラウンディング問題の謎？

三種類のシグナル (C. S. Peirce)

アイコン

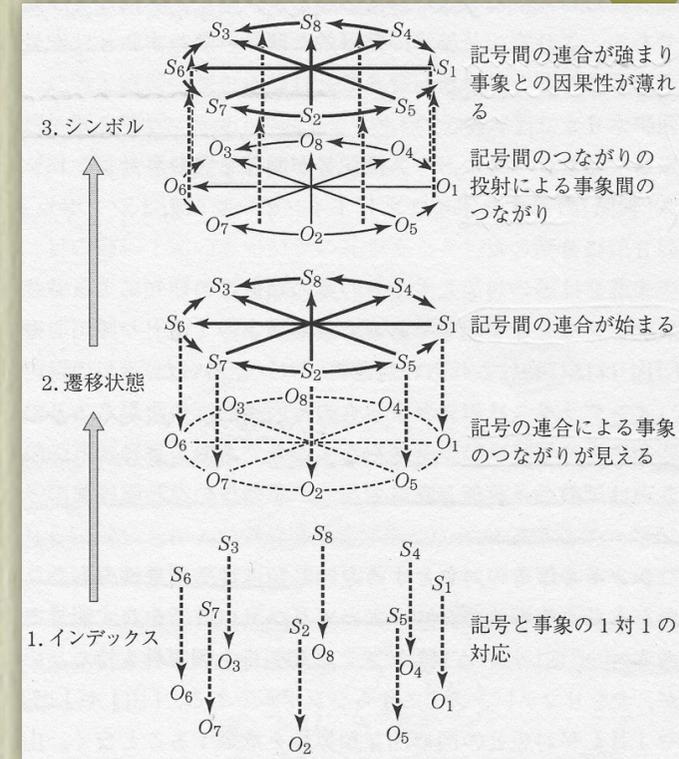
- 対応物と物理的・視覚的に類似したシグナル
 - ジェスチャー

インデックス

- 物理的類似性はない。対応物と一対一対応
 - 「ボタン」と「ジュース」

シンボル

- インデックス群の関係に基づいた階層的表現形式
 - 文法の誕生／再帰構造の誕生



シンボル・グラウンディング問題の謎？

三種類のシグナル (C. S. Peirce)

アイコン

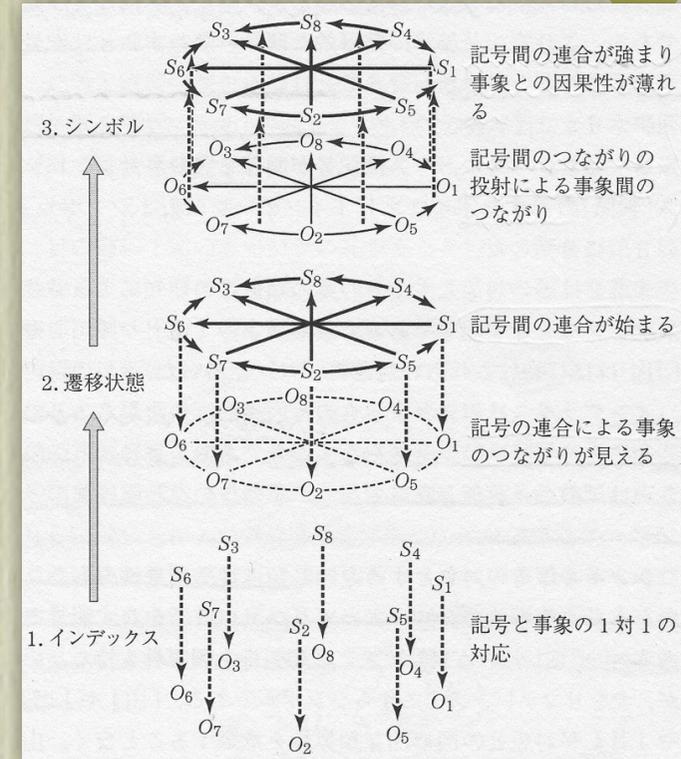
- 対応物と物理的・視覚的に類似したシグナル
 - ジェスチャー

インデックス

- 物理的類似性はない。対応物と一対一対応
 - 「ボタン」と「ジュース」

シンボル

- インデックス群の関係に基づいた階層的表現形式
 - 文法の誕生／再帰構造の誕生



シンボル・グラウンディング問題の謎？

「音群構造化＝不変項としての語」から統語構造まで



sequence of acoustic events

シンボル・グラウンディング問題の謎？

「音群構造化=不変項としての語」から統語構造まで

sequence of dialogues



sequence of sentences



sequence of phrases



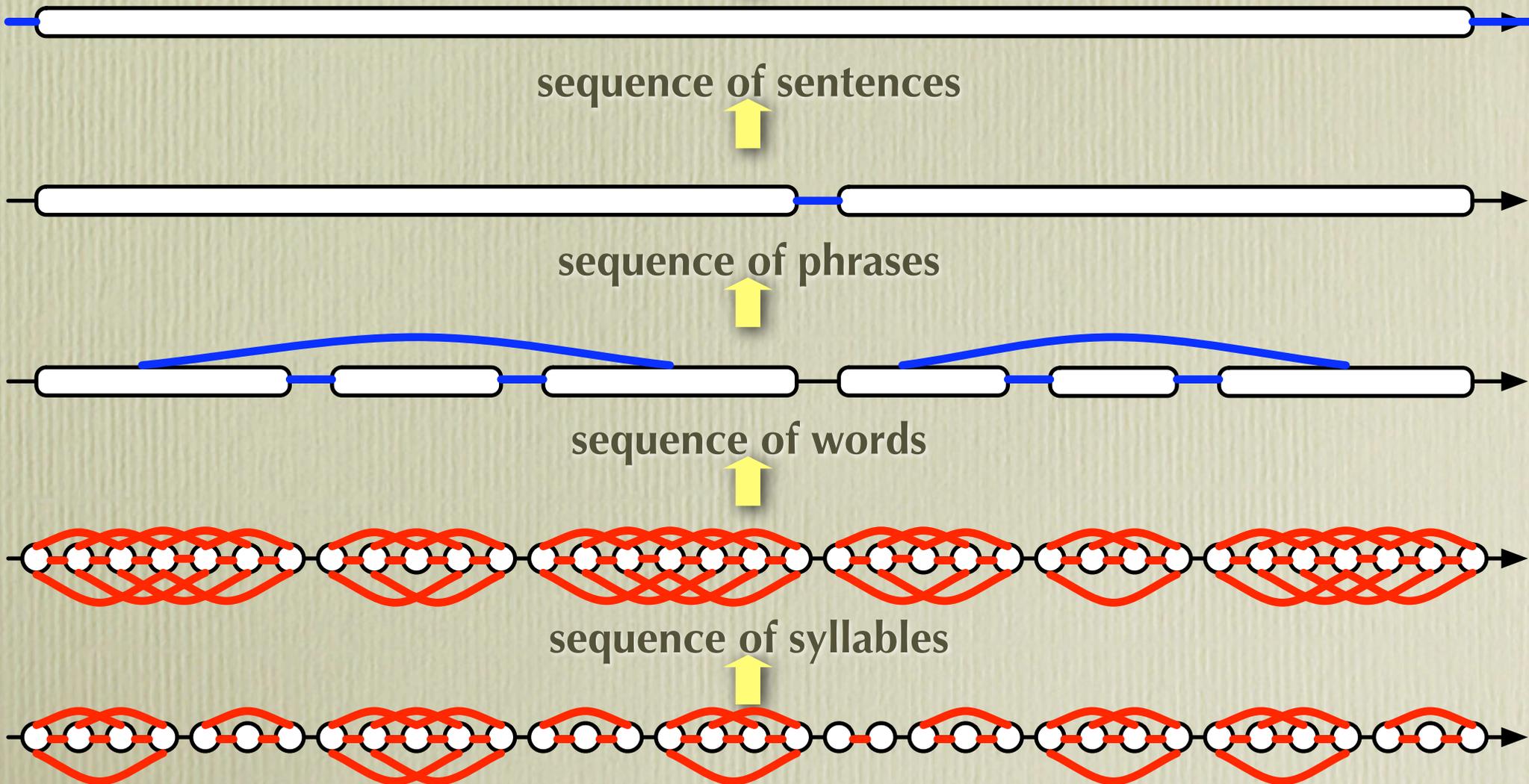
sequence of words



sequence of syllables

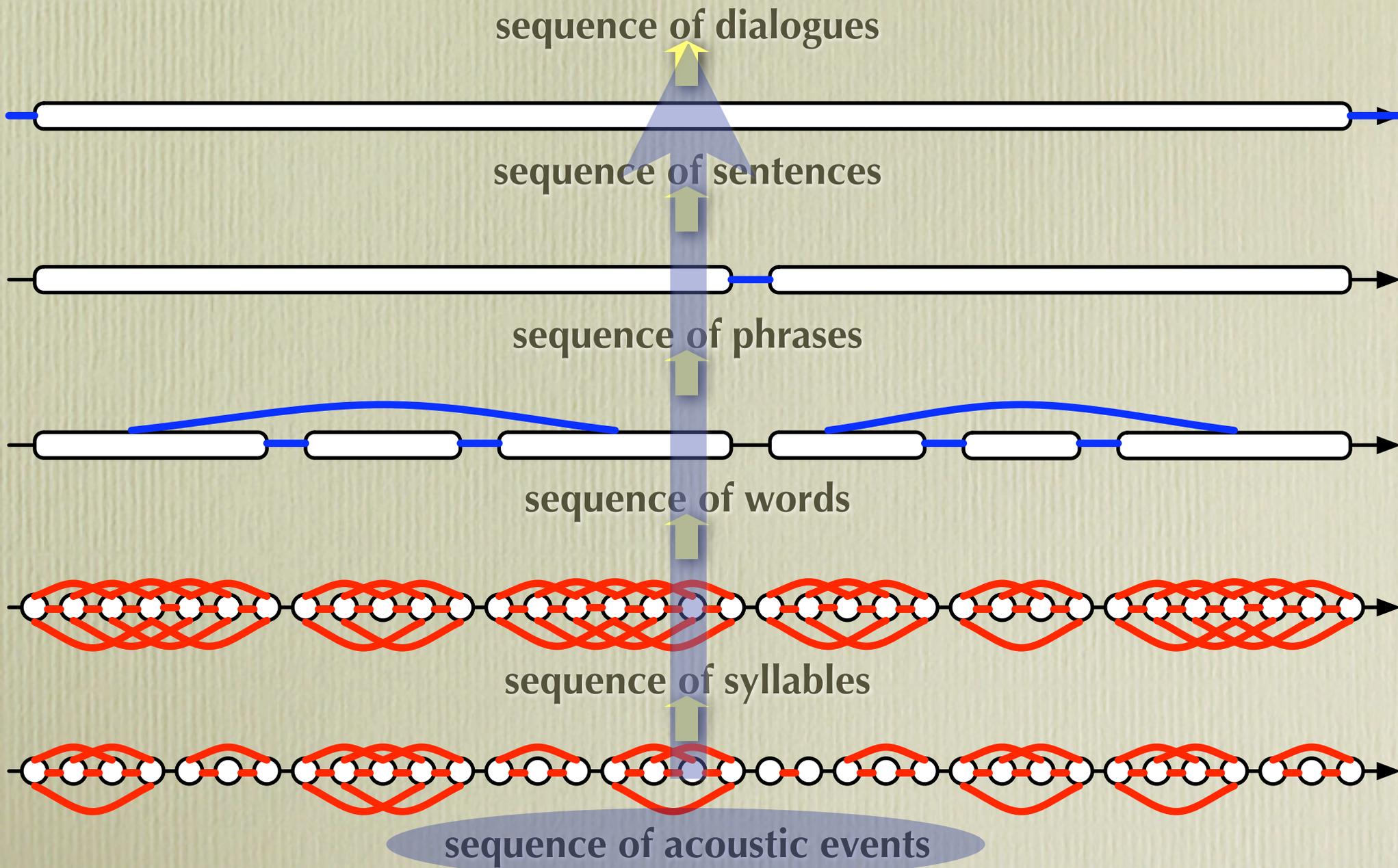


sequence of acoustic events



シンボル・グラウンディング問題の謎？

「音群構造化=不変項としての語」から統語構造まで



シンボル・グラウンディング問題の謎？

「音群構造化＝不変項としての語」から統語構造まで

sequence of dialogues

sequence of sentences

sequence of phrases



シンボル・グラウンディング問題の謎？

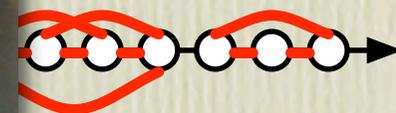
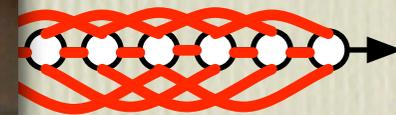
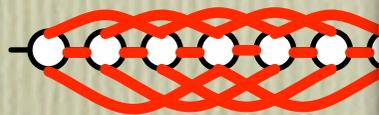
Syntax in music

「単語としての語」から統語構造まで

sequence of dialogues

sequence of sentences

sequence of phrases



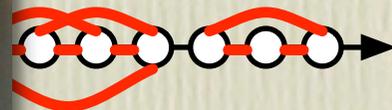
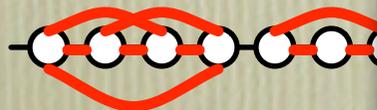
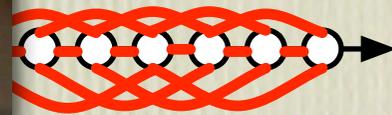
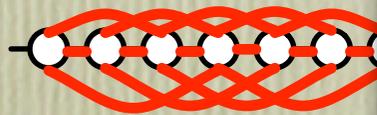
シンボル・グラウンディング問題の謎？

Syntax in music

Syntax in language

sequence of sentences

sequence of phrases



シンボル・グラウンディング問題の謎？

Syntax in music

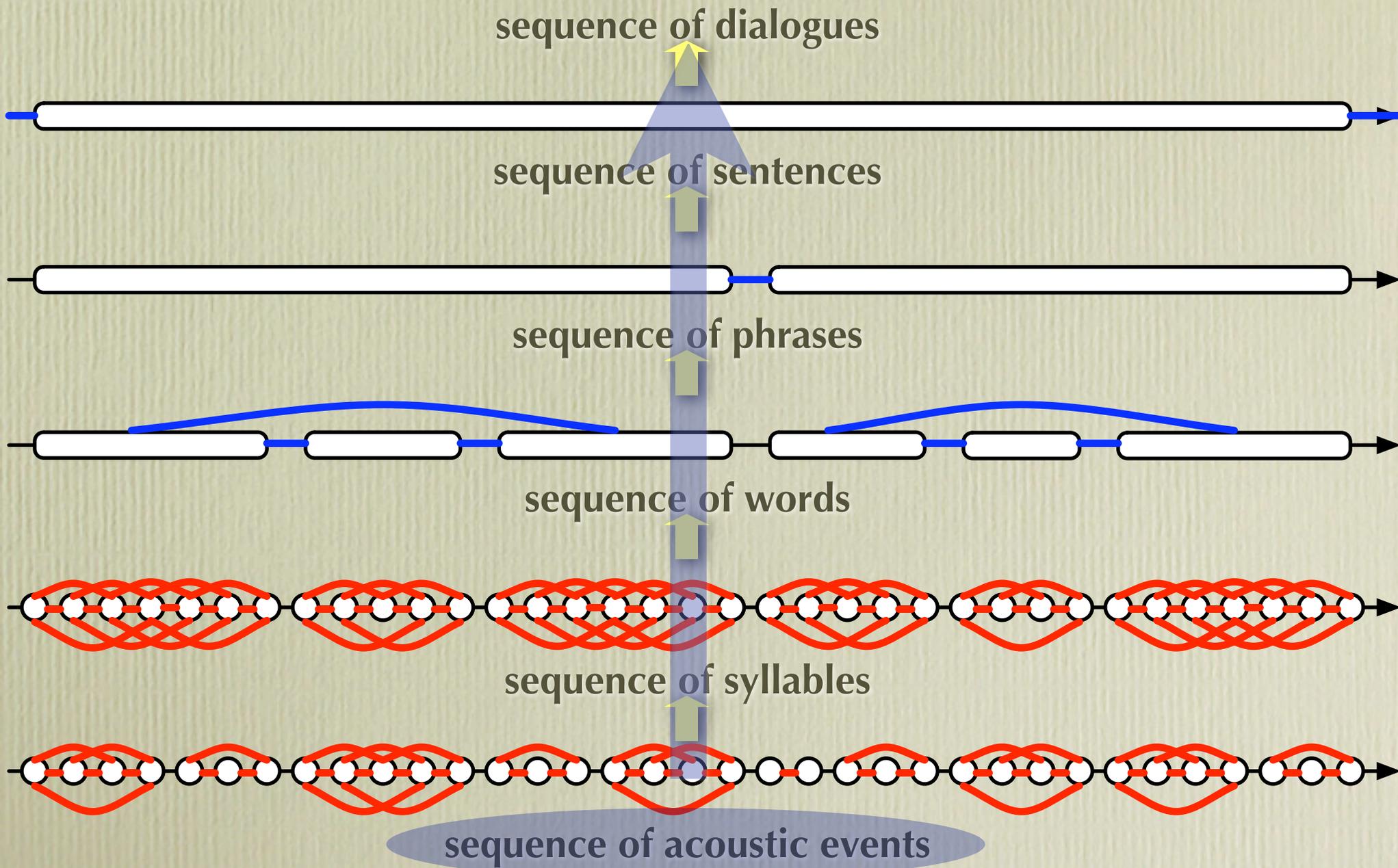
=

Syntax in language



シンボル・グラウンディング問題の謎？

「音群構造化=不変項としての語」から統語構造まで



シンボル・グラウンディング問題の謎？

言語の統語構造 = 音楽の音構造？

- 文の統合構造の処理を担当する脳部位
- 音楽の音構造の処理を担当する脳部位
- 両者は（左右差はあるが）類似した脳リソースを利用

幼児の文知覚 = 「長い単語」としての知覚

Take-your-shirt-off.
Wash-your-hands.
You-want-more-milk?

心とことばの起源を探る
文化認知
The Cultural Origins of Human Cognition

マイケルトマセロ Michael Tomasello
大塚奇夫・中澤和子・西村義樹・本多啓郎

シリーズ **認知科学** 第004巻



シンボル・グラウンディング問題の謎？

言語の統語構造 = 音楽の音構造？

- 文の統合構造の処理を担当する脳部位
- 音楽の音構造の処理を担当する脳部位
- 両者は（左右差はあるが）類似した脳リソースを利用

幼児の文知覚 = 「長い単語」としての知覚

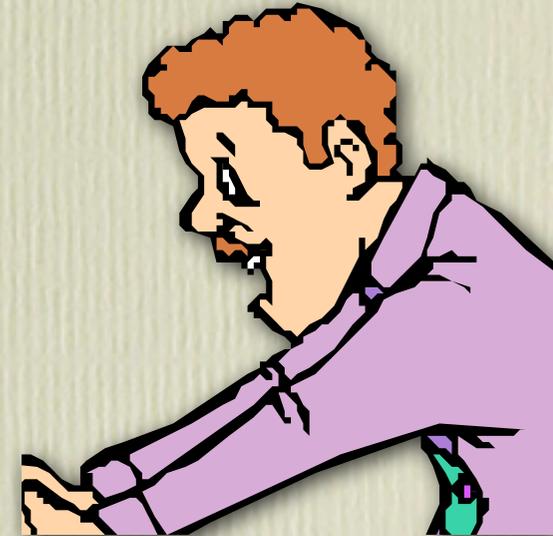
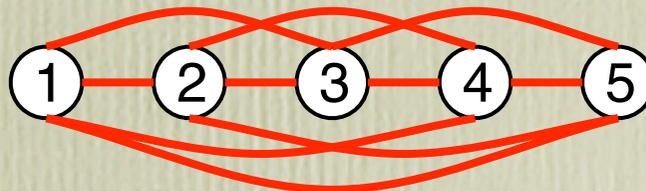
心とことばの起源を探る
文化と認知
The Cultural Origins of Human Cognition

マイケルトマセロ Michael Tomasello
大塚尚夫・中澤和子・西村義樹・本多啓郎

シリーズ **認知科学** 第004号



Take-your-shirt-off.
Wash-your-hands.
You-want-more-milk?



シンボル・グラウンディング問題の謎？

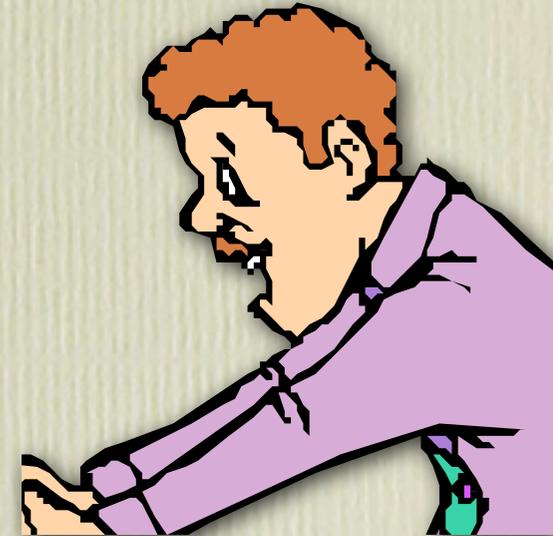
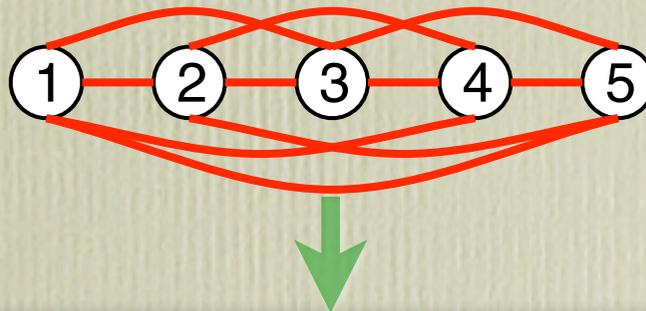
言語の統語構造 = 音楽の音構造？

- 文の統合構造の処理を担当する脳部位
- 音楽の音構造の処理を担当する脳部位
- 両者は（左右差はあるが）類似した脳リソースを利用

幼児の文知覚 = 「長い単語」としての知覚



Take-your-shirt-off.
Wash-your-hands.
You-want-more-milk?



心とことばの起源を探る
文化認知
The Cultural Origins of Human Cognition

マイケルトマセロ Michael Tomasello
大塚奇夫・中澤和子・西村義樹・本多啓郎

シリーズ **認知科学** 第04巻

初めての統語構造 = 音系列の構造解析？

シンボル・グラウンディング問題の謎？

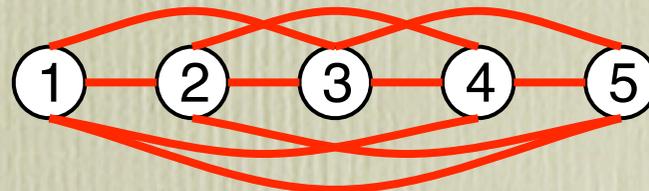
心とことばの起源を探る



幼児の又知見ー「長い単語」としての知見



Take-your-shirt-off.
Wash-your-hands.
You-want-more-milk?



初めての統語構造 = 音系列の構造解析？

まとめ

幼児の言語獲得の持つ不思議さ

- Another POS (Poverty of Stimulus)
- 幼児は親の声の何を真似ようとしているのか？

話者不変の音声の構造的表象 ～音響的普遍構造～

- 峯松の声から峯松であることを除去する秘伝
- 階名としての「あいうえお」と音名としての「あいうえお」

普遍なる不変構造に基づく音声認識

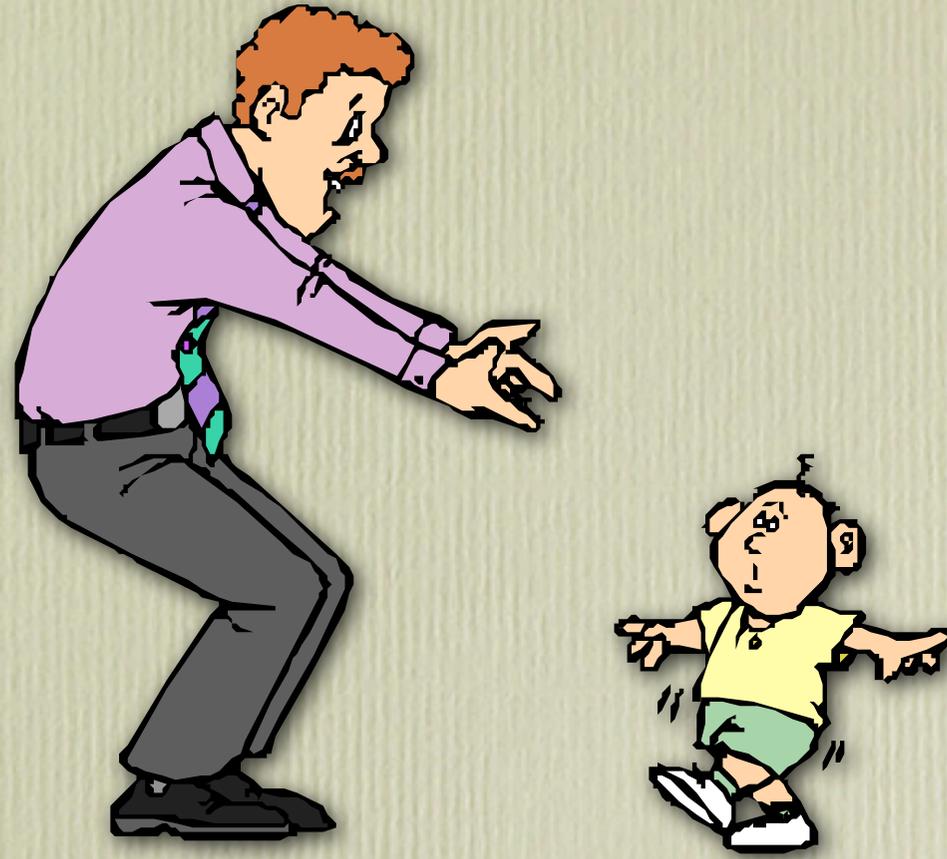
- 一人の声で全ての人の声を認識する!?

言語学／音楽学／脳科学／障害学／心理学／人類学・・・

- 失読症と自閉症 ～音声言語は勘違いの上に成立する～
- ヒトが進化のプロセスを通して獲得したコモンセンス

まとめ

まとめ



まとめ



まとめ



物理的等価性から 機能的・相対的等価性へ



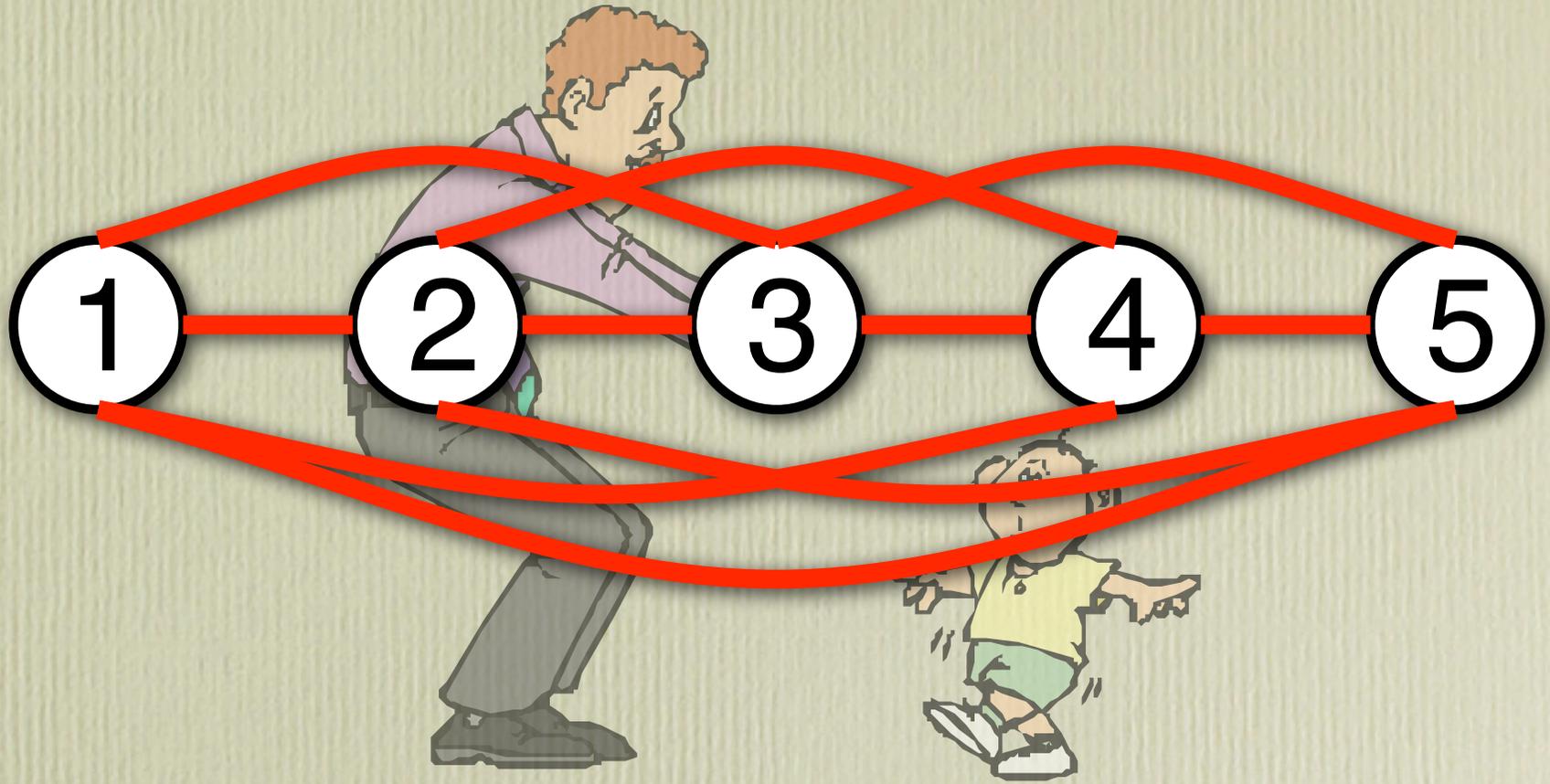
まとめ



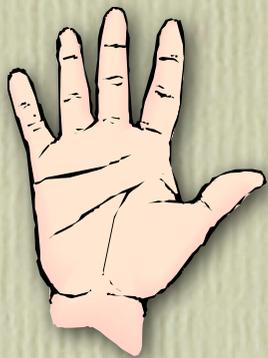
音の差異が成す不変項 (音声アフォーダンス)



まとめ



ご清聴, 有り難うございました



ご清聴, 有り難うございました

