＜SCOPE最終成果発表＞

音声中の非言語情報の生成・知覚の特性解析と多言語間コミュニケーションへの応用

2010年6月11日，学術情報センター

赤木　正人

北陸先端科学技術大学院大学 (JAIST)
情報科学研究科　人間情報処理領域　教授

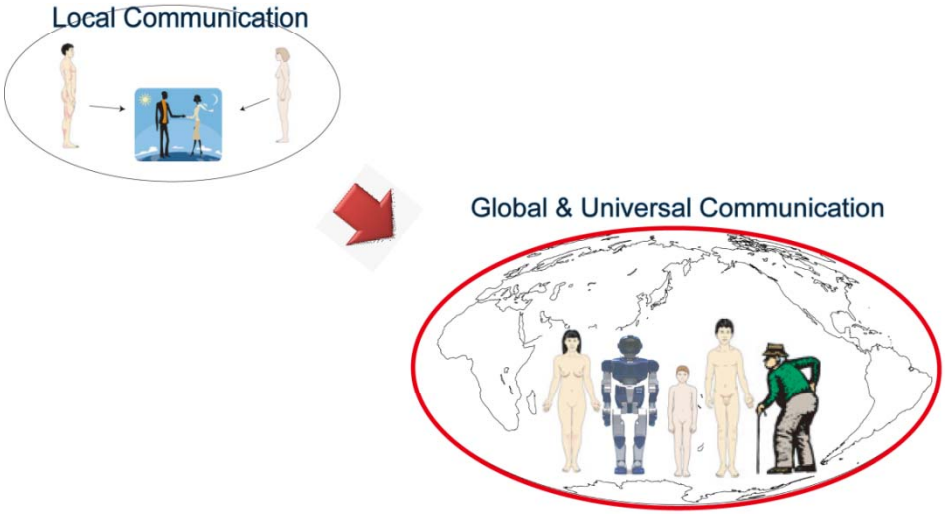
JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY,
HOKURIKU 1990

ユニバーサルな音声コミュニケーション

- 言語・民族・文化を越えた（＝**グローバルな**），また，言語・民族・文化のみならず老人，幼児，あるいは障害者との障壁のない（＝**ユニバーサルな**）コミュニケーションの重要性が増している

- 音声コミュニケーションでは，「何を話しているか」という**言語情報**だけではなく，これ以外の情報，たとえば個人性（性別，年齢），感情・健康状態，声質などの**非言語情報**が多数送受される

- 言語を越えた**グローバルでユニバーサルな音声コミュニケーション環境の構築**
  ← **非言語情報についての音声コミュニケーションを解明**することが一助となる



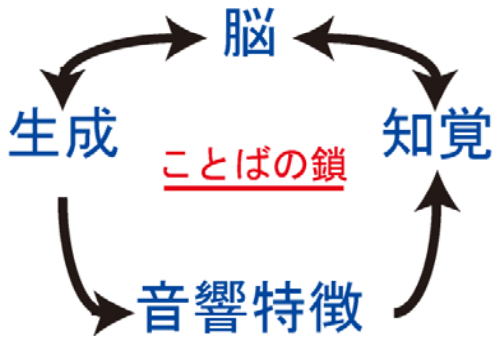
ユニバーサルな音声コミュニケーション(2)

- 音声における非言語情報のコミュニケーションにおいて，言語・民族・文化を越えた**ユニバーサルコミュニケーション環境**を構築するためには
  - 非言語情報の生成・知覚において，言語・民族・文化によらない**ヒトの生物学的「共通要素」**，すなわち，
    - 生成のための万国共通の構音運動，
    - 共通の構音運動から作り出される共通の音声特徴，
    - 音声特徴を呈示することにより生起される共通の知覚特徴・脳活動，そして，
    - この上に立つ人間の共通の行動

    が存在しなければならない．
  - 非言語情報の知覚・生成について「何が本質なのか」を考察し，言語・民族・文化によらない**ヒトの生物学的「共通要素」の存在を明らか**にする，そして，
  - 「共通要素」を多言語間での非言語情報の付加（**音声合成への応用**），非言語情報の認識（**音声認識、音声対話理解への応用**）に応用する

研究概要

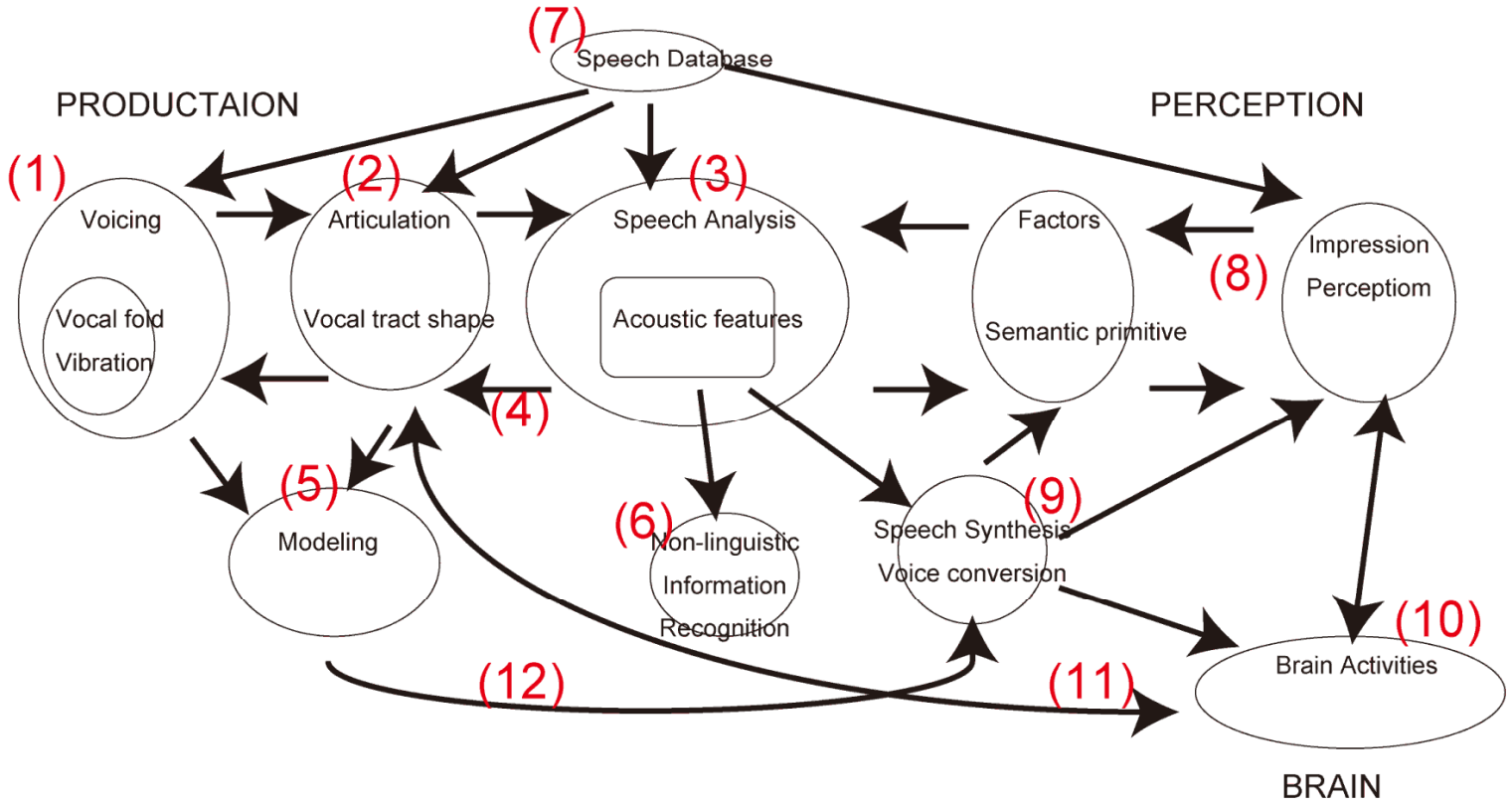
1. 脳と音声生成の相互作用、脳と音声知覚の相互作用、生成から音声特徴を経て知覚への経路，それぞれの中で，**非言語情報の生成・知覚機構の解明**を目指し，言語・民族・文化によらない**共通要素**とは何かについて検討
2. 共通要素を核として，**非言語情報（感情音声）の合成・認識**を試みる



研究方略

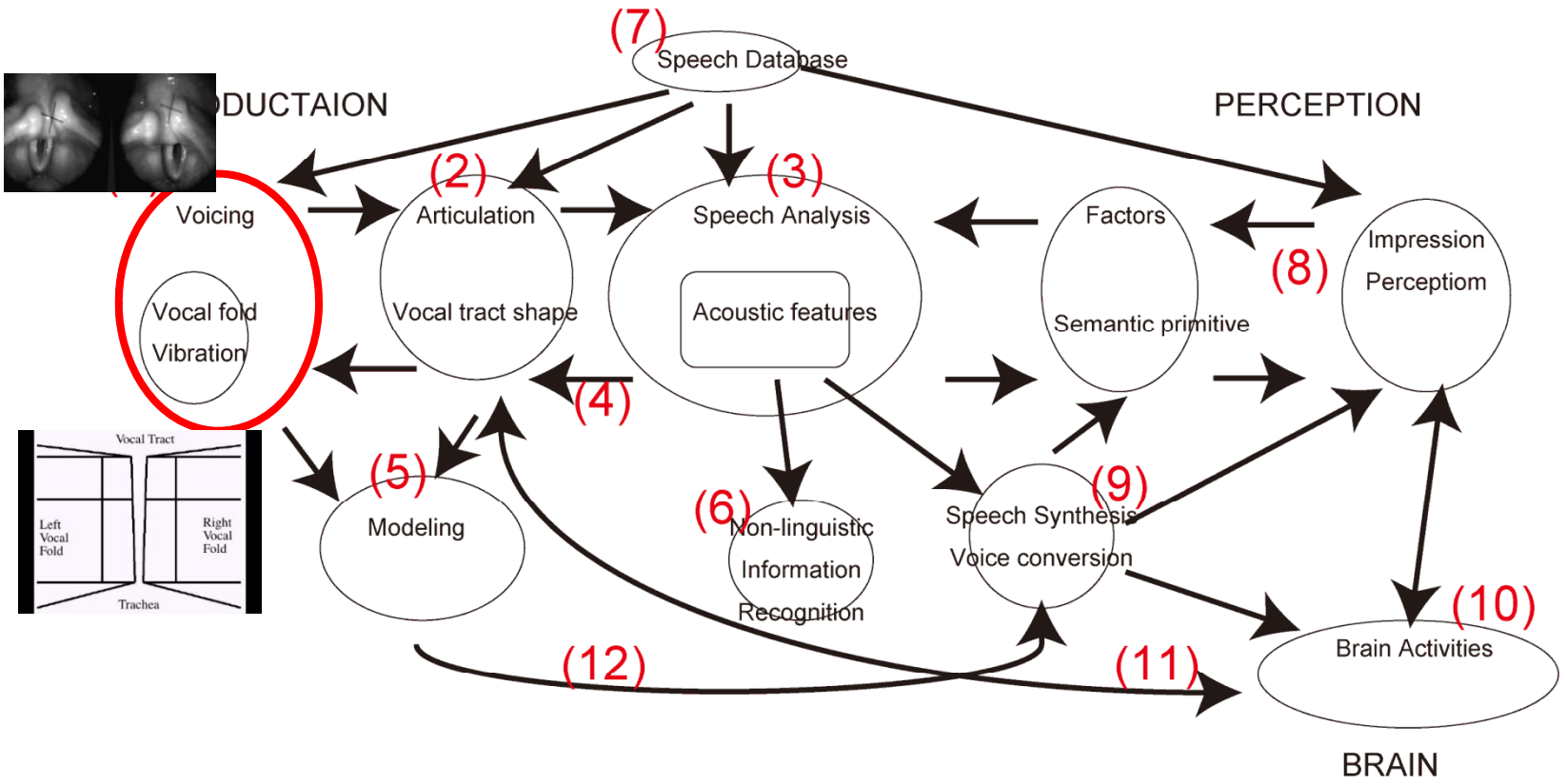
ヒトの観測　→　モデル化　→　工学システムの実現

研究概要(2)



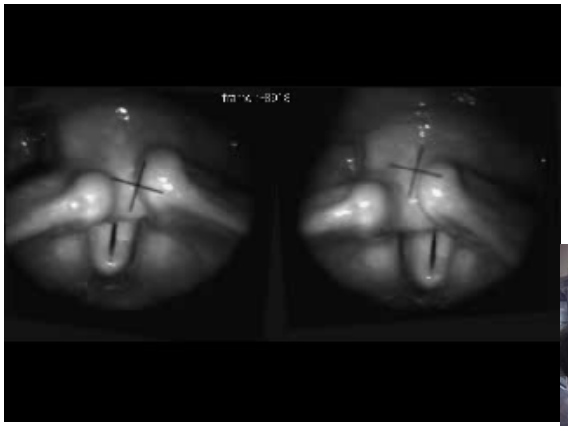
Production (1, 2, 4, 5, 7, 12): Jianwu Dang, Isao Tokuda, Atsuo Suemitsu, and Xugang Lu (JAIST); Tatsuya Kitamura (Konan University); Ken-ichi Sakakibara (Health Sciences University of Hokkaido)
Acoustic features (3): Masashi Unoki, Junfeng Li, and Xugang Lu (JAIST)
Perception (6, 8, 9): Masato Akagi, and Jianwu Dang (JAIST); Donna Erickson (Showa Academia Musicae)
Brain, Interaction between production and perception (10, 11): Masato Akagi, and Jianwu Dang (JAIST); Tatsuya Kitamura (Konan University)
海外共同研究機関：グルノーブル大（仏），オハイオ大，トレド大（米），清華大，中国科学院（中），等

音声発声（Voicing）



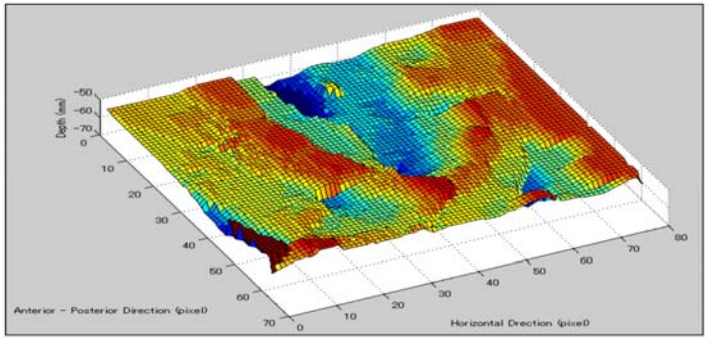
音声発声の観測とモデル化

観測



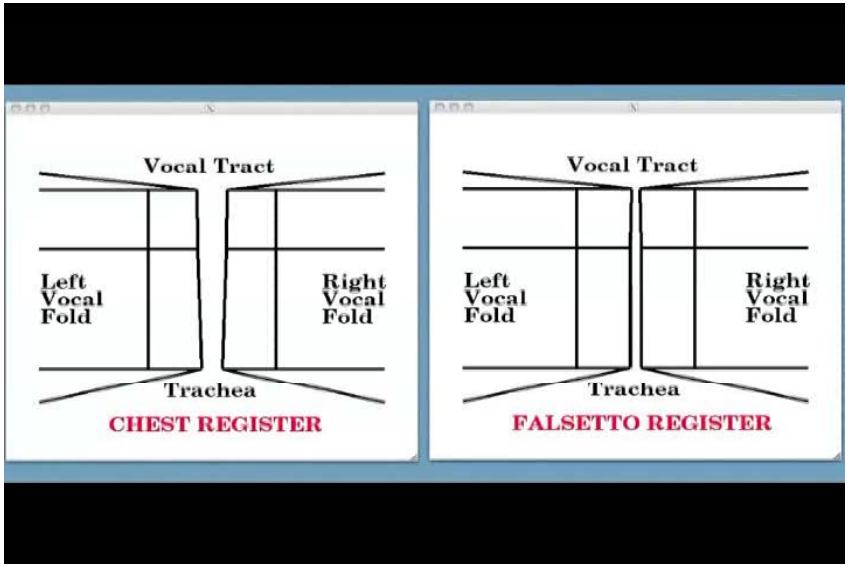
ステレオ計測

歌声発生時の計測（地声 vs. 裏声）

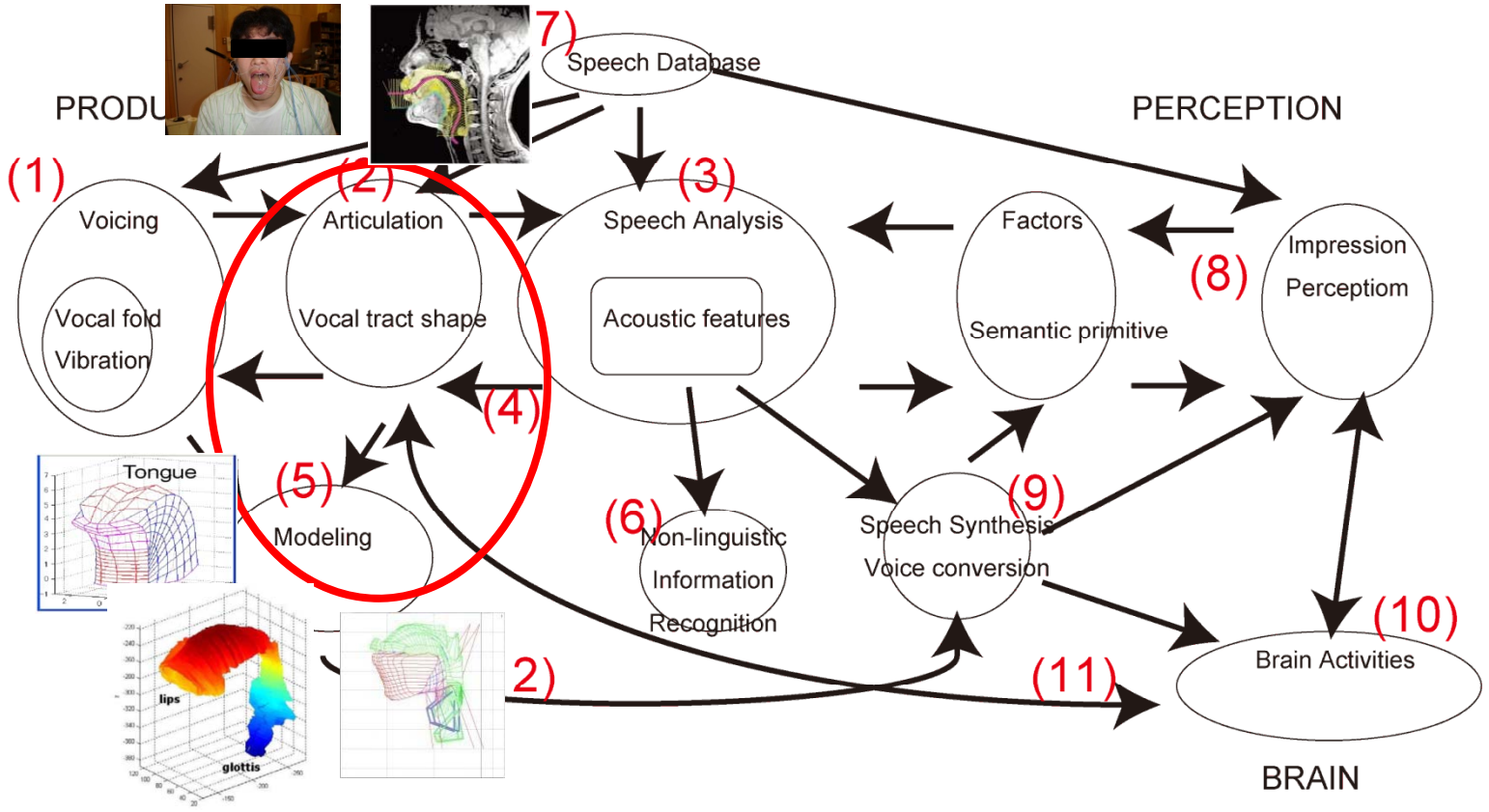


声帯の3次元計測

数理モデル



調音 (Articulation)



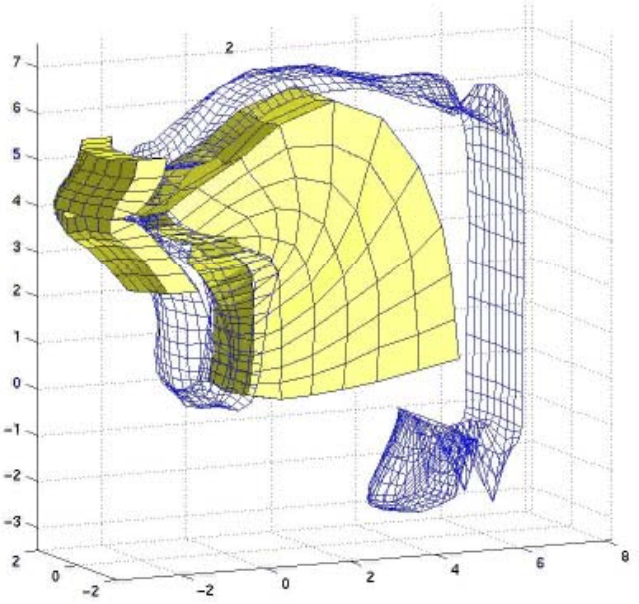
調音の観測とモデル化

観測

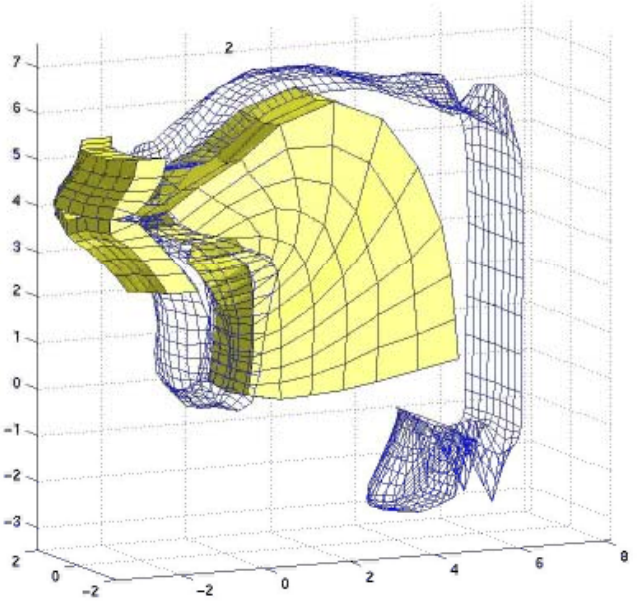
モデル化

応用

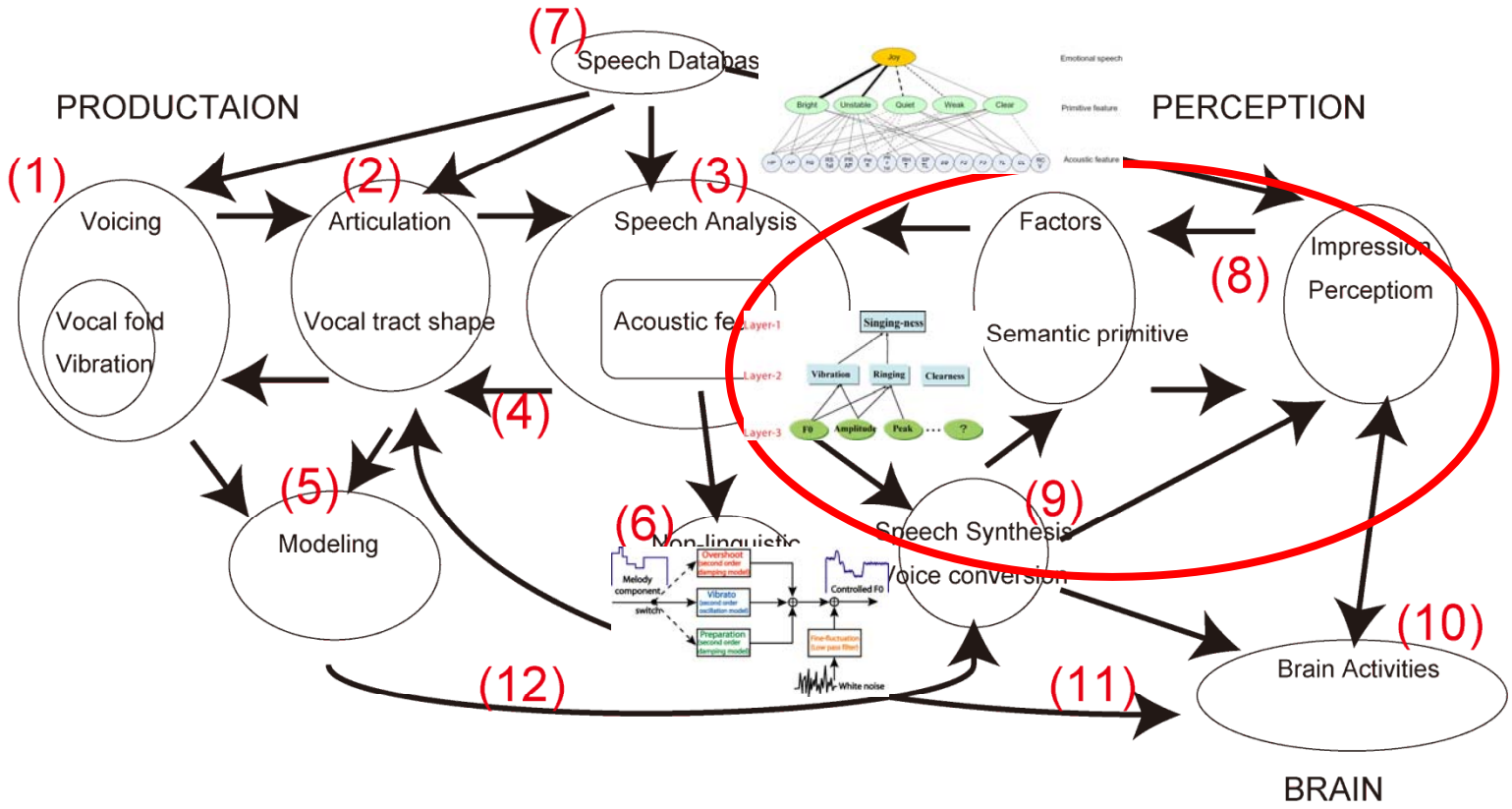
普通の発話



強調した発話

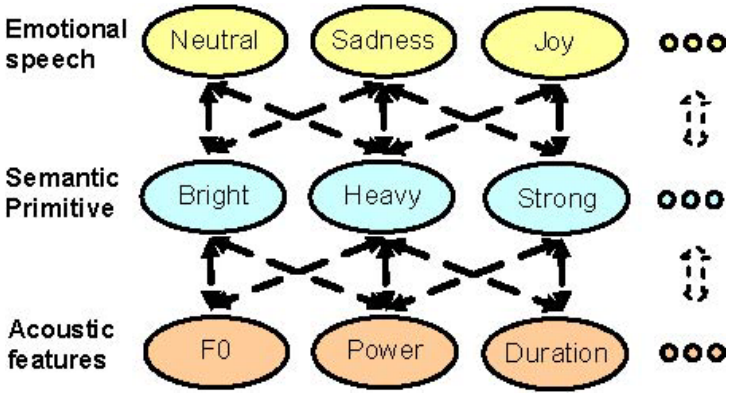


非言語情報知覚 (Perception)

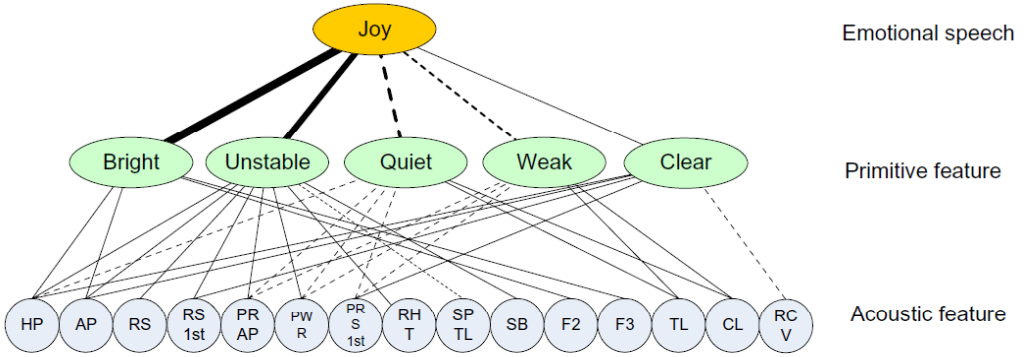


感情音声知覚のモデル化と応用

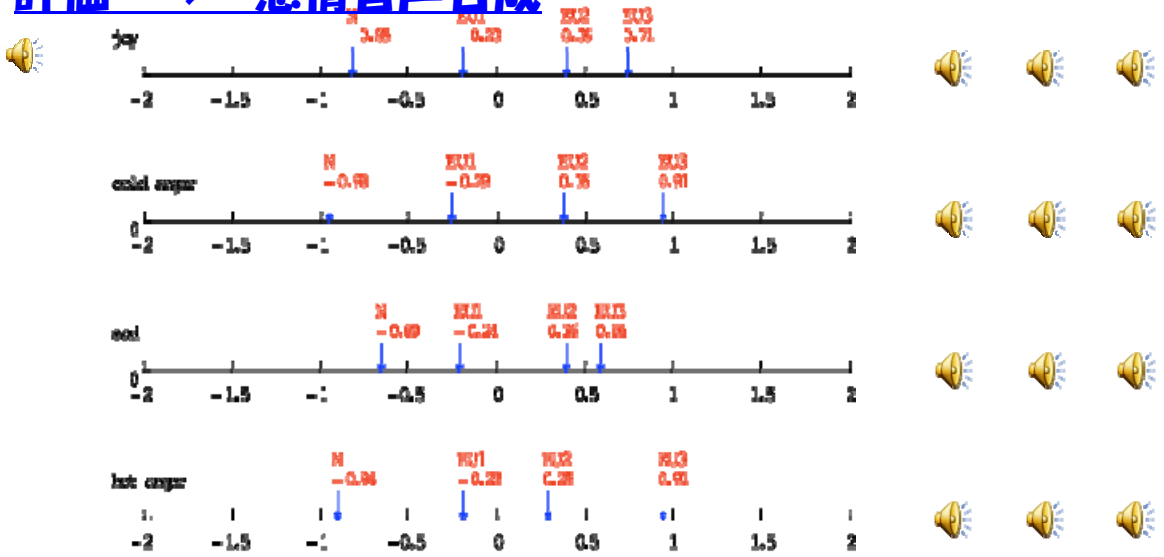
モデル化：三層知覚モデル



例： Joy



評価 → 感情音声合成



中国語話者と日本語話者の比較：
共通要素は存在するか？

JAIST
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY HOKURIKU

中国語話者(a)と日本語話者(b)により選ばれた *semantic primitive*の比較

(a)

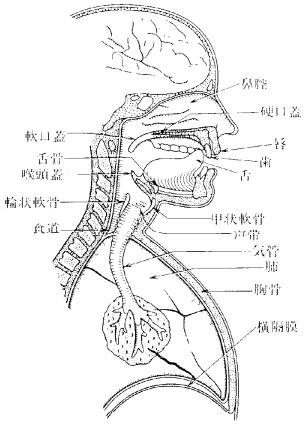
Neutral		Joy		Cold Anger		Sadness		Hot Anger	
PF	S	PF	S	PF	S	PF	S	PF	S
dull	-0.181	weak	-0.254	fluent	-0.498	bright	-0.122	muddy	-0.26
heavy	-0.117	low	-0.185	bright	-0.121	smooth	-0.295	heavy	-0.186
bright	0.115	clear	0.178	slow	0.164	heavy	0.179	fluent	0.118
clear	0.234	calm	0.288	weak	0.212	strong	0.181	unstable	0.17
smooth	0.256	smooth	0.44	muddy	0.384	raucous	0.267	hard	0.325

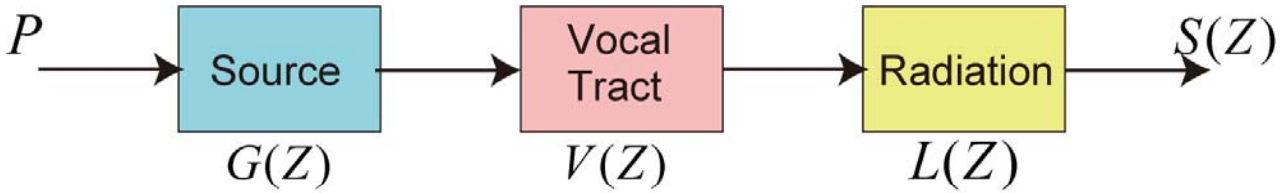
(b)

Neutral		Joy		Cold Anger		Sadness		Hot Anger	
PF	S	PF	S	PF	S	PF	S	PF	S
heavy	-0.329	quiet	-0.039	sharp	-0.079	slow	-0.231	calm	-0.063
weak	-0.181	weak	-0.036	strong	-0.049	monotono us	-0.073	quiet	-0.047
calm	0.103	clear	0.034	quiet	0.044	well-modulated	0.091	sharp	0.103
clear	0.127	unstable	0.063	weak	0.074	fast	0.153	unstable	0.12
monotonous	0.27	bright	0.101	heavy	0.061	heavy	0.197	well-modulated	0.124

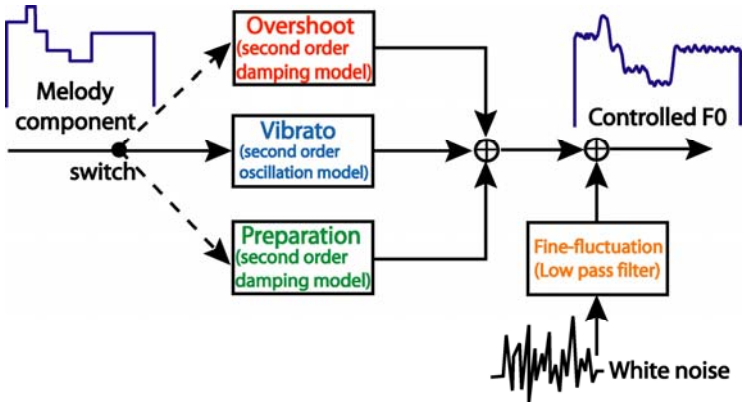
1. The first 10 semantic primitives that were shared by both Mandarin and Japanese listeners have the same valence (i.e., positive or negative correlation).
2. 6 of the 10 semantic primitives are associated with the same two acoustic features that have the highest correlations.
3. *bright*, *dark*, *low*, *heavy*, and *clear* are associated with average pitch (AP) and highest pitch (HP), and *strong* is associated with power range (PWR) and mean value of power range in accentual phrase (PRAP).

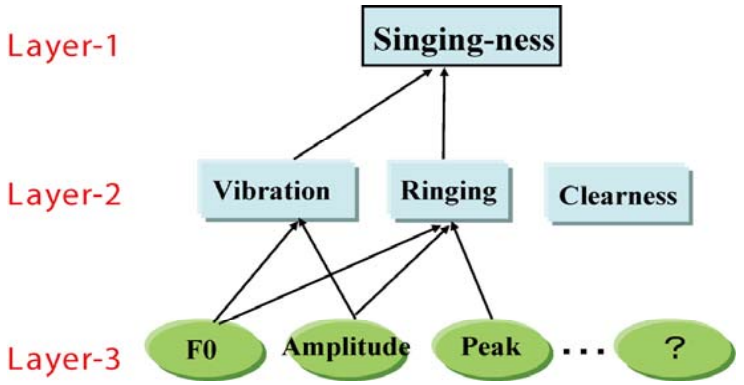
三層知覚モデルの応用：歌声





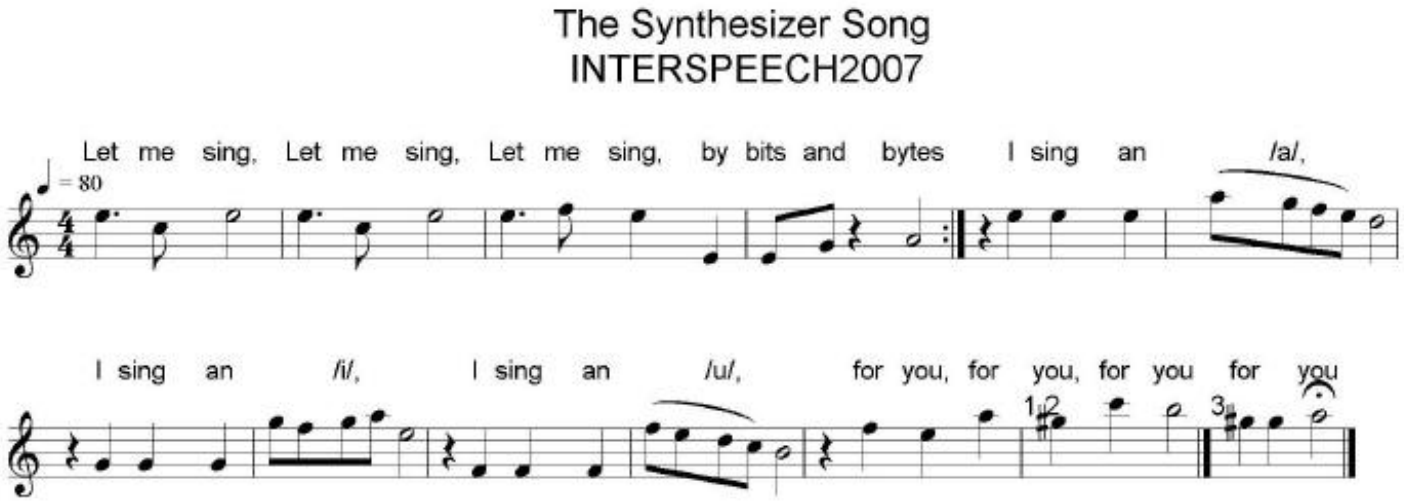
$$S(Z) = P \cdot G(Z) \cdot V(Z) \cdot L(Z)$$





	SPEAK	+ Melody	+ Ringing	+ Vibration	ALL
Classical singing (male)					

Demonstration



♪ Speaking voice (input): (male) (female)

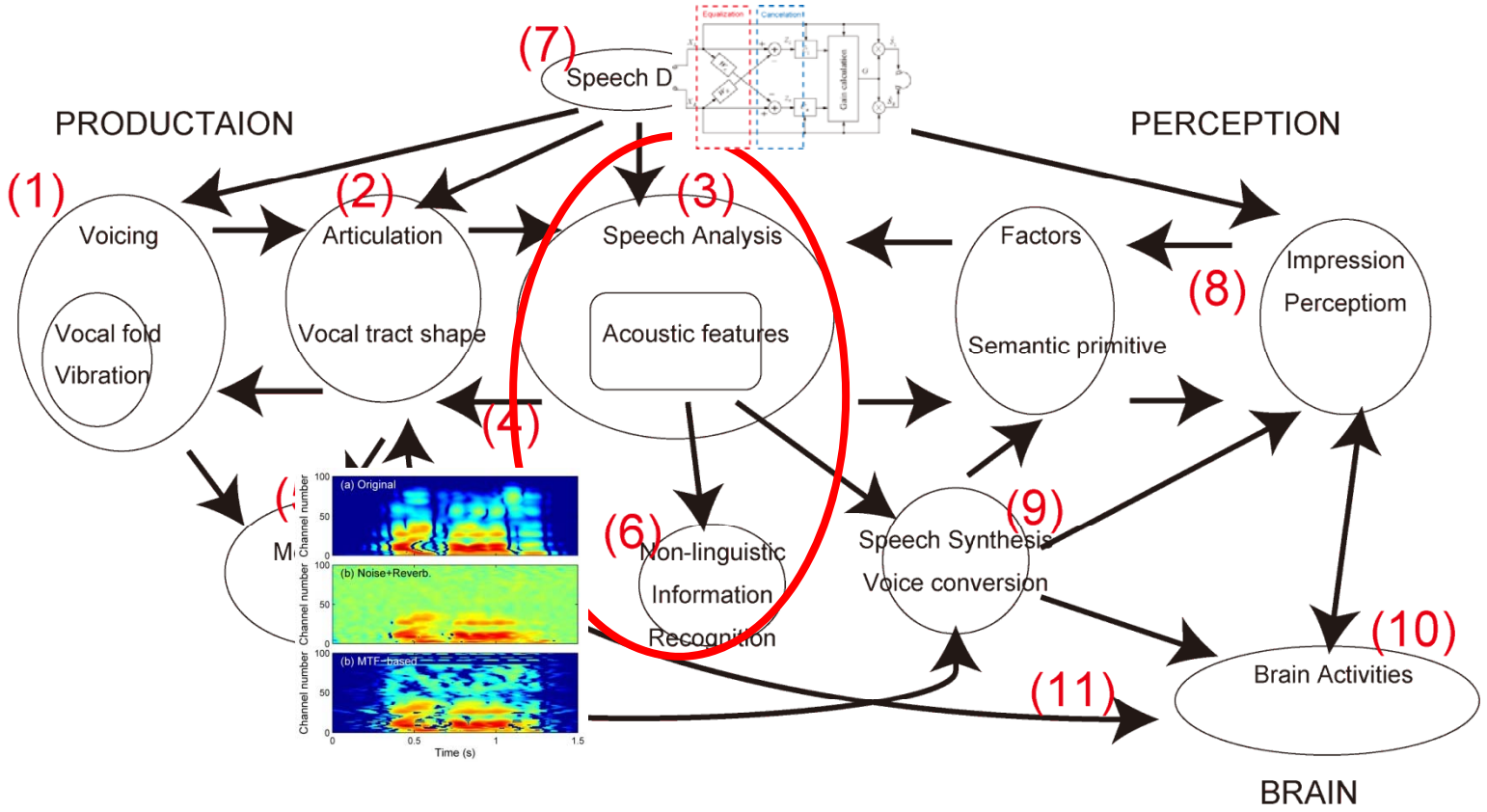
♪ Synthesized singing voice: (male) (female) (chorus)

We took the first place in SINGING SYNTHESIS CHALENGE held in the InterSpeech2007.

Another register ···: Falsetto

	Speaking	Singing
Speaker-A		
Speaker-B		

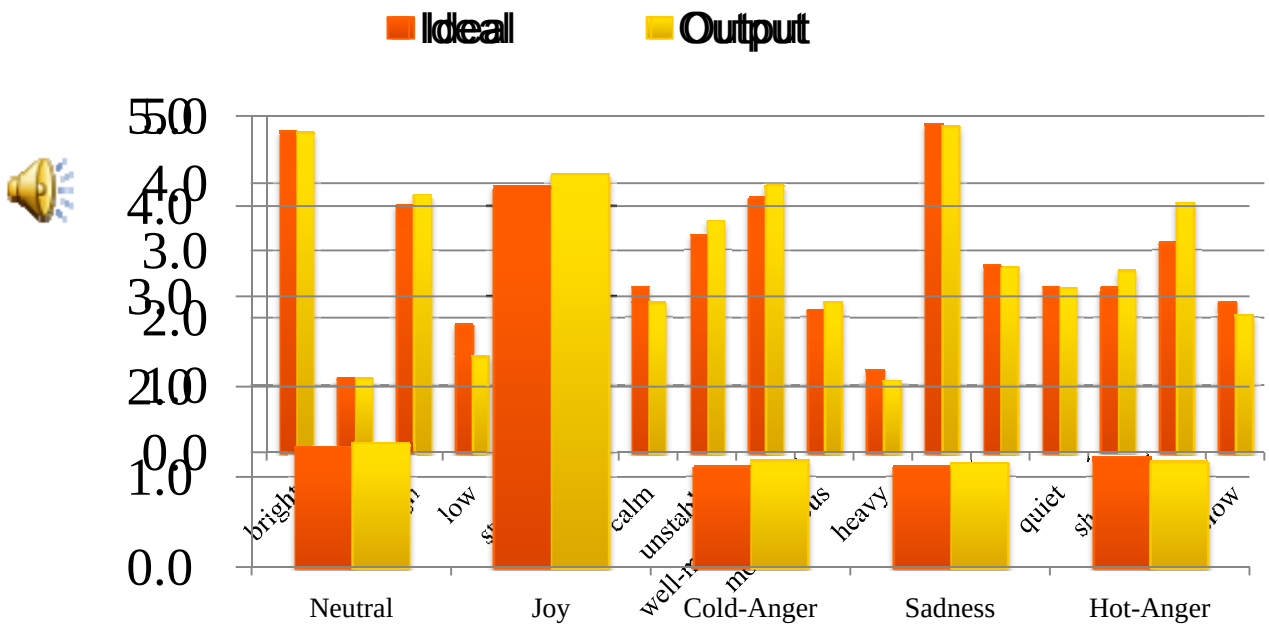
音声の分析 (Speech Analysis)



実環境での音声分析と感情認識

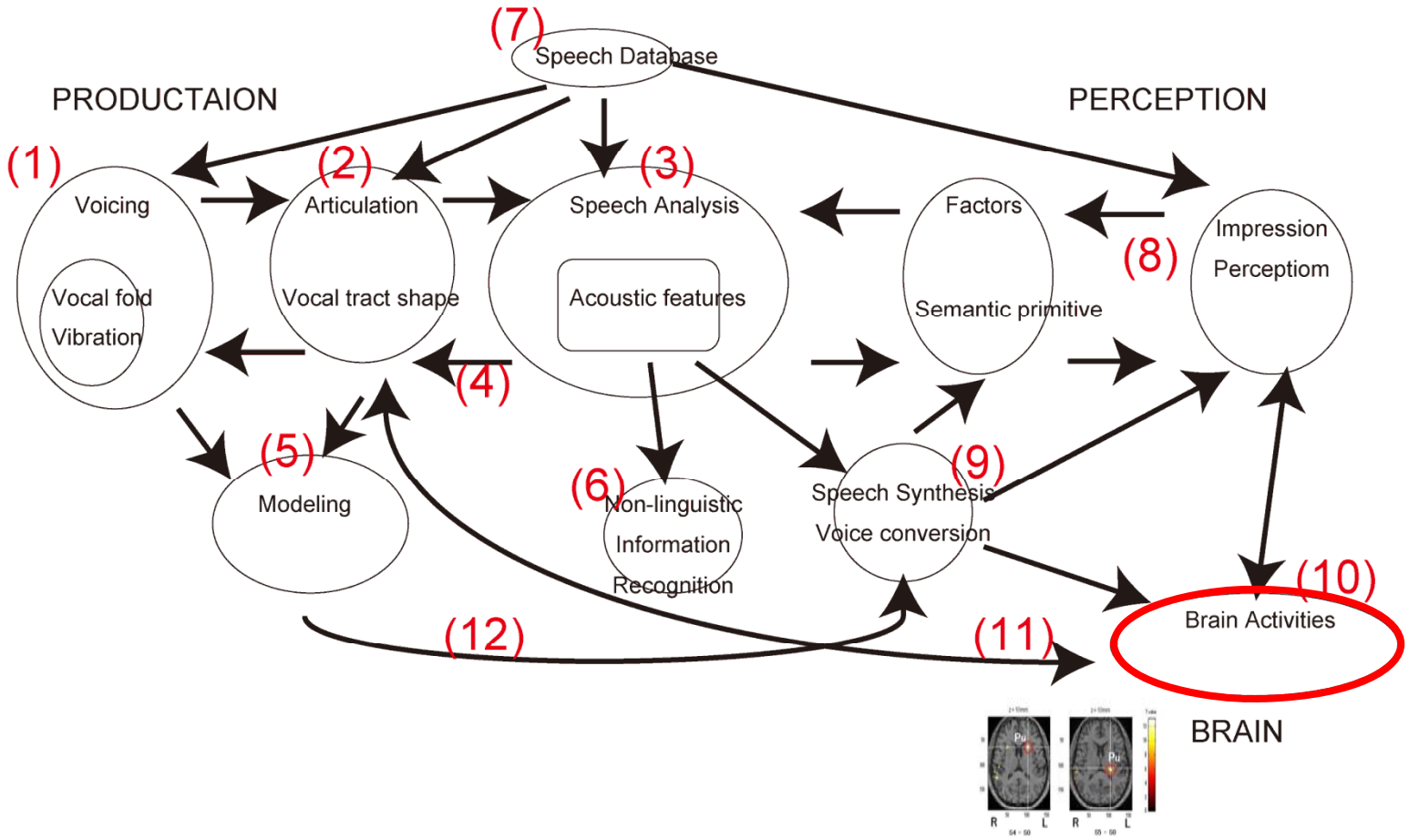
- 実環境（雑音・残響環境）中での高精度な音響特徴抽出
  - 雑音除去，残響回復，骨導音声ひずみ回復，等

- 音声中の感情認識
  - 三層感情知覚モデルの応用

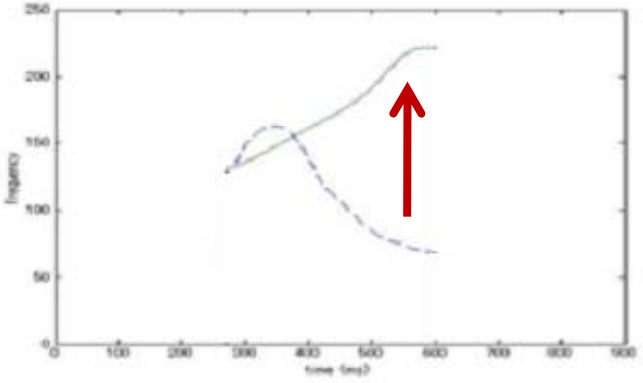


ドイツ語データベースへも適用中

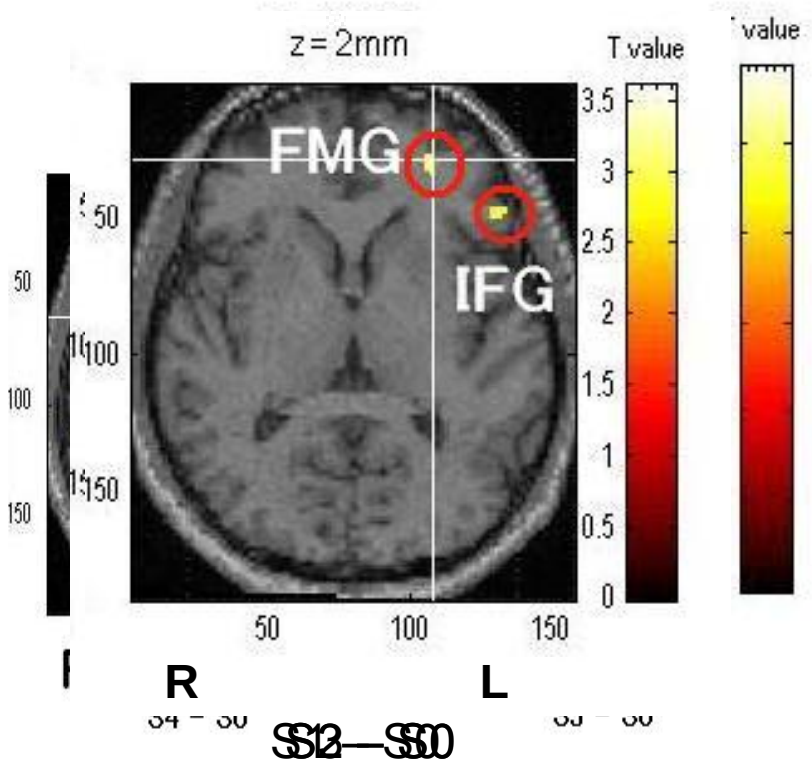
脳活動 (Brain Activity)



感情音声による脳活動測定



	基本周波数の時間変化	脳の賦活
S0	↘	
S1-S0	↘	大脳皮質 前頭葉
S2-S0	→	大脳皮質 上頭頂小葉、角回
S3-S0	↗	大脳基底核 尾状核
S4-S0	→	大脳基底核 被核
S5-S0	↗	大脳基底核 被核



感情音声による脳活動測定

	基本周波数の時間変化	脳の賦活	代表的な感情語	
S0			肯定、共感	
S1		大脳皮質 前頭葉	肯定、冷静	
S2		大脳皮質 上頭頂小葉、角回	落胆、悲しい	
S3		大脳基底核 尾状核	聞返し、驚き	
S4		大脳基底核 被核	疑い、否定	
S5		大脳基底核 被核	疑い、驚き	

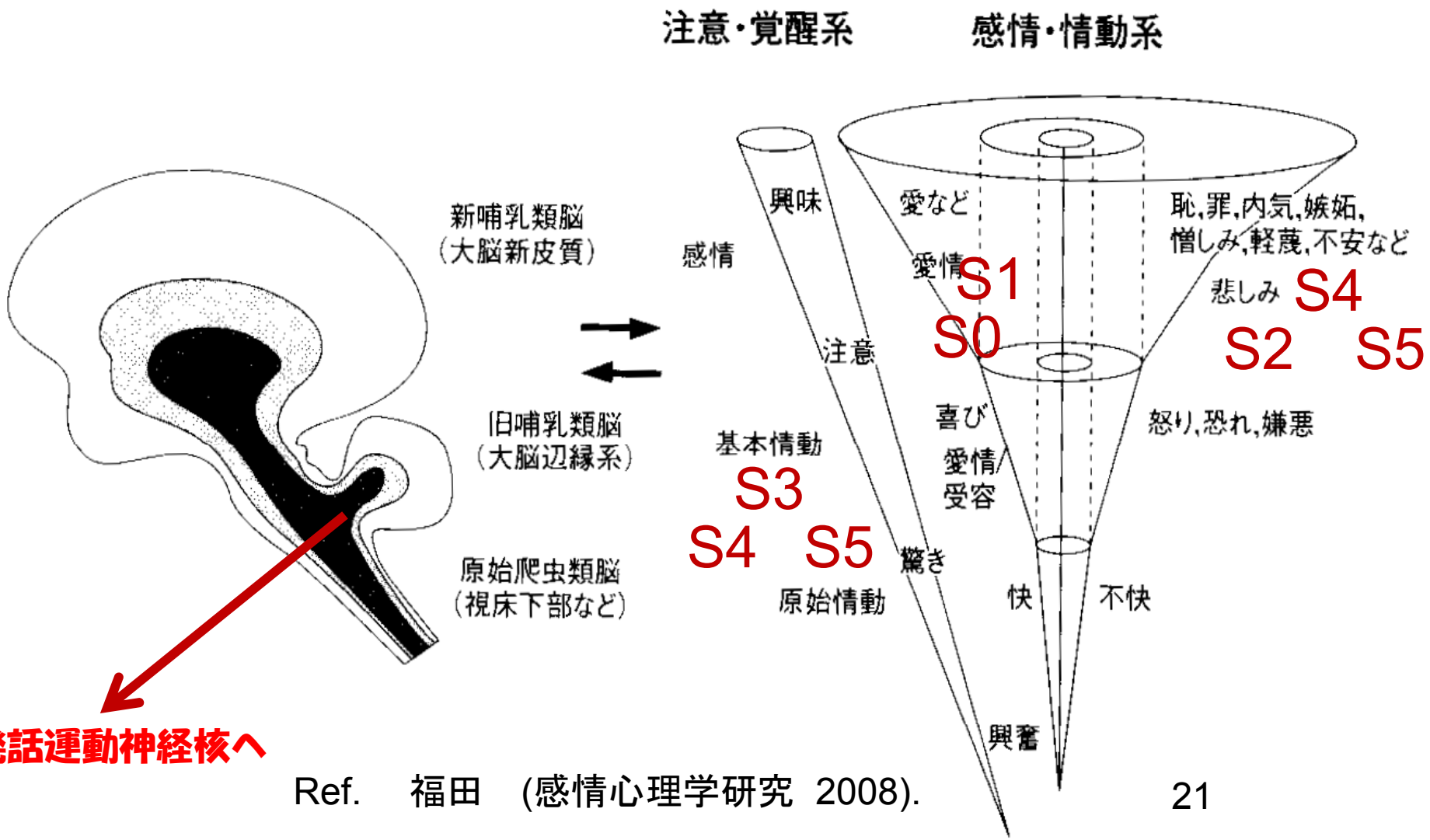
19

感情音声による脳活動測定

	基本周波数の時間変化	脳の賦活	感情階層説との対応	代表的な感情語
S0				肯定、共感
S1		大脳皮質 前頭葉		肯定、冷静
S2		大脳皮質 上頭頂小葉、角回		落胆、悲しい
S3		大脳基底核 尾状核		聞返し、驚き
S4		大脳基底核 被核		疑い、否定
S5		大脳基底核 被核		疑い、驚き

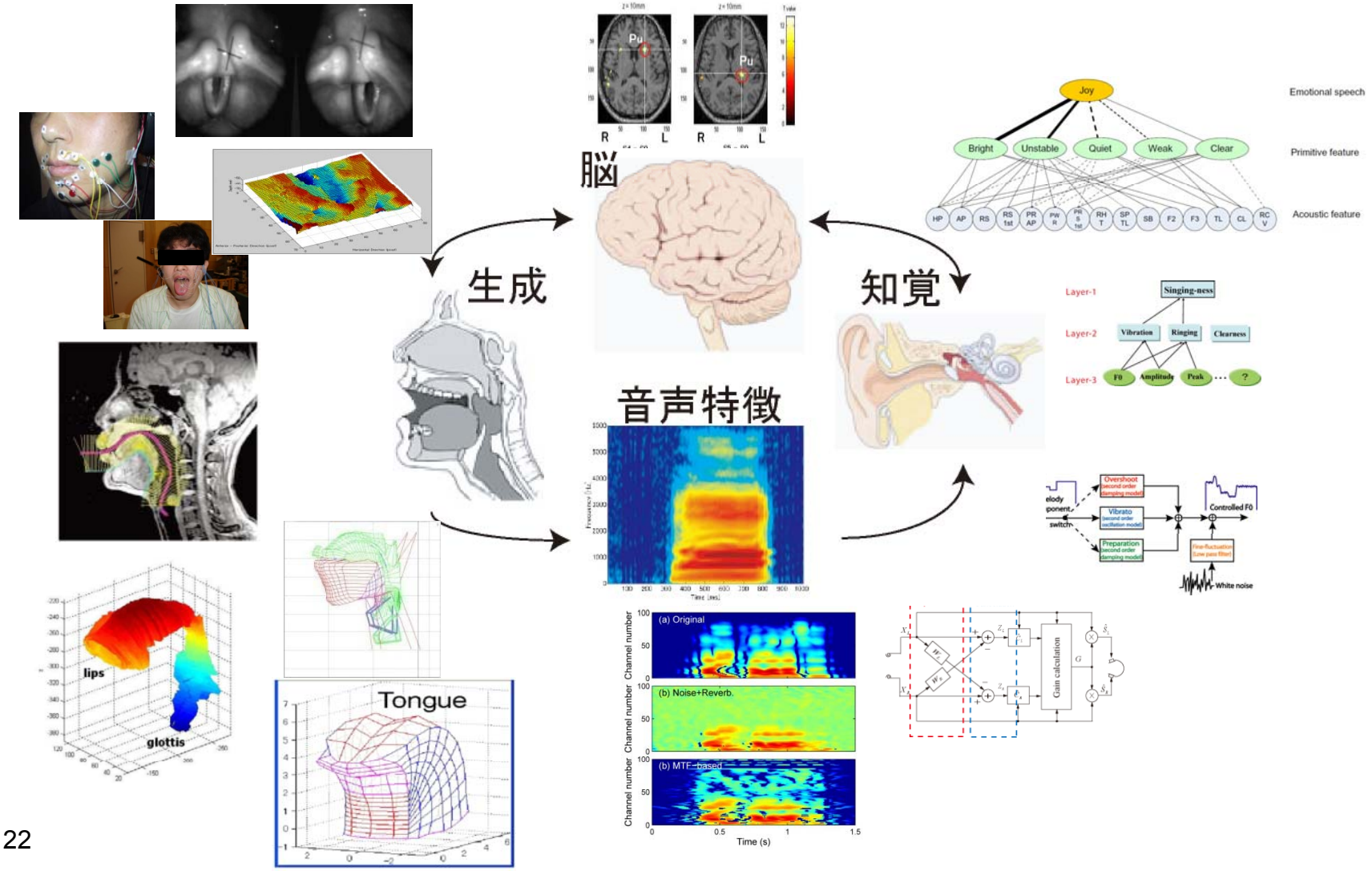
20

感情音声による脳活動測定



Ref.　福田　(感情心理学研究 2008).

ことばの鎖とヒトの観測，モデル化，そして工学システムへ



まとめ

- 音声生成・知覚の環構造の中で、
  - ヒトの観測、モデル化を通して、
  - 非言語情報の生成・知覚機構の解明を目指し、
  - 言語・民族・文化によらない共通要素とは何かについて

  検討した
- 共通要素を核として、
  - 非言語情報（感情音声）の合成・認識のための工学システムの実現

  を試みた

- その結果、
  - 音声中の非言語情報の生成と知覚に関する環構造を裏付ける結果が得られ、最終的な目標が達成された
  - 認識システム、および、合成システムのプロトタイプの構築を行った

波及効果

感情などの非言語情報を認識し、また、音声に付加し制御できるようになれば；

1.非言語情報によるコミュニケーションの解明が進み**グローバルでユニバーサルなコミュニケーション環境構築の一助**となる

2.工学的応用として、多言語への非言語情報の付加（音声合成への応用）、非言語情報の認識（音声認識・対話理解への応用）への道が開け、**多言語間の円滑なコミュニケーション、幼児や老人、音声・聴覚障害者も含めたユニバーサルなコミュニケーションへの応用**も可能となる

受賞・その他

- The 1st place in Singing Synthesis Challenge, InterSpeech2007（受賞日2007年8月31日）
- 信号処理学会　Best Paper Award　受賞（受賞日2009年3月2日）
- 情報処理学会　インタラクション2009　インタラクティブ発表賞　受賞（受賞日2009年3月6日）
- 日本音響学会　佐藤論文賞　受賞　（受賞日2010年3月9日）