Journal of Signal Processing, Vol. 14, No. 2, pp. 107-117, March 2010

研究室紹介

赤木研究室(北陸先端科学技術大学院大学)

Akagi Laboratory at JAIST

赤木　正人
Masato Akagi

http://www.ais.jaist.ac.jp/
http://www.jaist.ac.jp/is/2008ja/kenkyu/ichiran/akagi.html

Journal of Signal Processing
信号処理

研究室紹介

赤木研究室(北陸先端科学技術大学院大学)

Akagi Laboratory at JAIST

赤木　正人
Masato Akagi

http://www.ais.jaist.ac.jp/
http://www.jaist.ac.jp/is/2008ja/kenkyu/ichiran/akagi.html

1. はじめに

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・赤木研究室は，1992年のJAIST学生受け入れと同時に発足し，今年で18年となる。研究室では，同じ領域に属する党，鵜木，徳田の各研究室と共同して，音声信号処理に関する研究を行っている。以下，研究内容および成果の概要を説明する。

2. 研究の基本コンセプト

人間以外の動物にとって音を聞くこととは，多くは生存するため，すなわち，(1) 敵から身を守るために危険を察知する，あるいは，(2) 獲物のいる場所を特定して捕獲する，ための重要な手段である。これらは音による方向知覚，距離知覚の一例である。一方，人間にとって音を扱うとは，音による方向知覚，距離知覚に留まらず，音（声）によりコミュニケーションすること，すなわち，音声の生成・知覚（ことばを発すること／ことばを聞き理解すること）が生きていく上で重要な要素となっている。

人間が相互に円滑にコミュニケーションを行う場合，言葉を発して相手に自分の考え，感情などを伝えようとする一方で，相手が伝えてきた考え，感情などの情報を受け取って理解し，そして，適切な応答を行うことが必要である。図1に示すように，我々が話し相手にある考えを伝えよとする場合

(1) 何を伝えるかを意識する。
(2) その考えを日本語で伝えようとするならば，日本語に沿った単語を選びだし，日本語の文法にあった語順で並べる。
(3) 個々の音節とか音韻を発するために音声生成機構をどのように動かすかをプランニングする。
(4) 音声生成に関係する器官に運動神経を通して指令を与える。
(5) 様々な器官を関連して動かすことにより音声を発話する。

一旦発話された音声は，空気の疎密波となり空気中を伝搬する。このとき，環境による様々な外乱

北陸先端科学技術大学院大学情報科学研究科人間情報処理領域
〒923-1292 石川県能美市旭台1-1
School of Information Science, Japan Advanced Institute of Science and Technology (JAIST)
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
E-mail: akagi@jaist.ac.jp

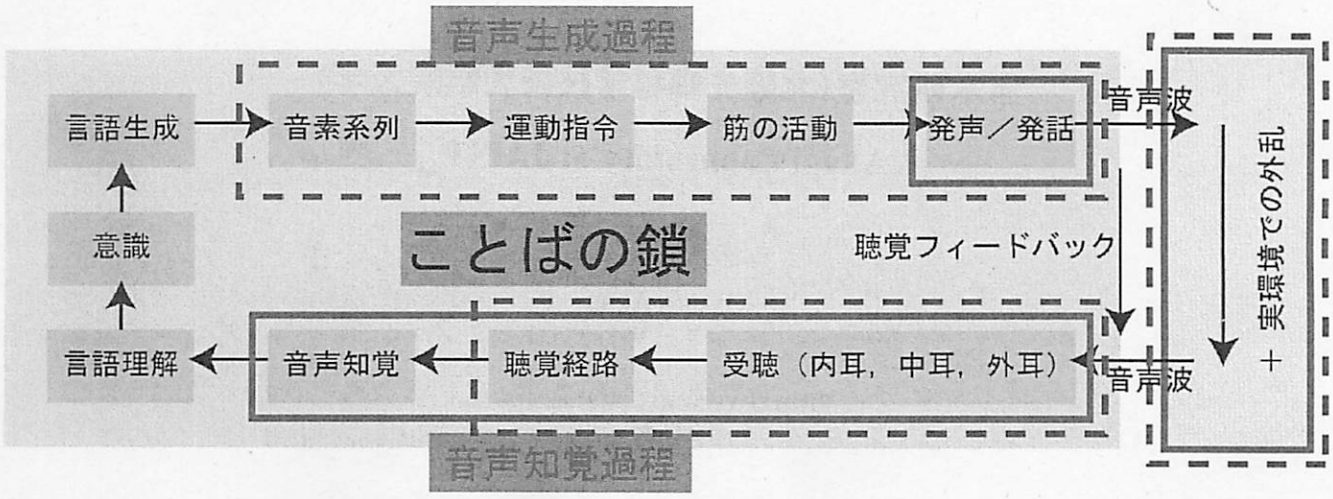

図1　音声コミュニケーションの基本（ことばの鎖）：実線枠の部分を中心に研究を行ってきた。点線枠は鵜木，党，徳田研究室で実施されている研究内容，重複部分は共同で実施している研究である。

（エアコンのファンの音，他人の話声などの雑音，部屋の状態によっては残響など）が加わり，聞き手に到達する。聞き手は

(1) 聴覚を使って音を受け取る（受聴）。
(2) 聴覚末梢系で神経発火パルスとして符号化したのち，脳中枢へ伝える。
(3) 様々な外乱に埋もれた音声を取り出し，知覚する。
(4) 音声を理解し相手の考えを汲み取る。

そして最後に

(5) 相手の考えに対する答えを意識し，話し手がたどったのと同じ道を通って相手に考えを伝える。

コミュニケーションを行おうとする人間相互でこのサイクルが旨く回ることによって，円滑なコミュニケーションが保たれるわけである。このサイクルのことを，”ことばの鎖”と読んでいる。

音声によるコミュニケーション（聞く・話す）は，上述のように，人間の基本的な営みである。また，音声によるコミュニケーションを議論する場合，”ことばの鎖”は一つの有用なモデルである。そこで，当研究室では，音声コミュニケーションを議論する基本的枠組みとして”ことばの鎖”を採用することとし，まず，「聞く・話す」を行う主体である人間を知り，そして，営みを記述・構成しこれを模擬して計算機上に実現することで，「聞く・話す」に関連する重要な要件を獲得する。さらに，これらの結果をもとに，工学的に意味あるシステムの実現を目指す。すなわち，”ことばの鎖”にもとづいた観測→モデル化→応用の流れを通して音声コミュニケーションを探究することを研究の基本コンセプトとしている。

3．研究概要

赤木研究室では，これまでに，音声によるコミュニケーション（聞く・話す）の中で

A)　「話す」：機械の口がより賢くなるように，より自然な合成音をつくることを目的として，音声スペクトルと声道形状の関係，合成音への個人性・感情などの非言語情報の付与，歌声らしい歌声の合成などの研究

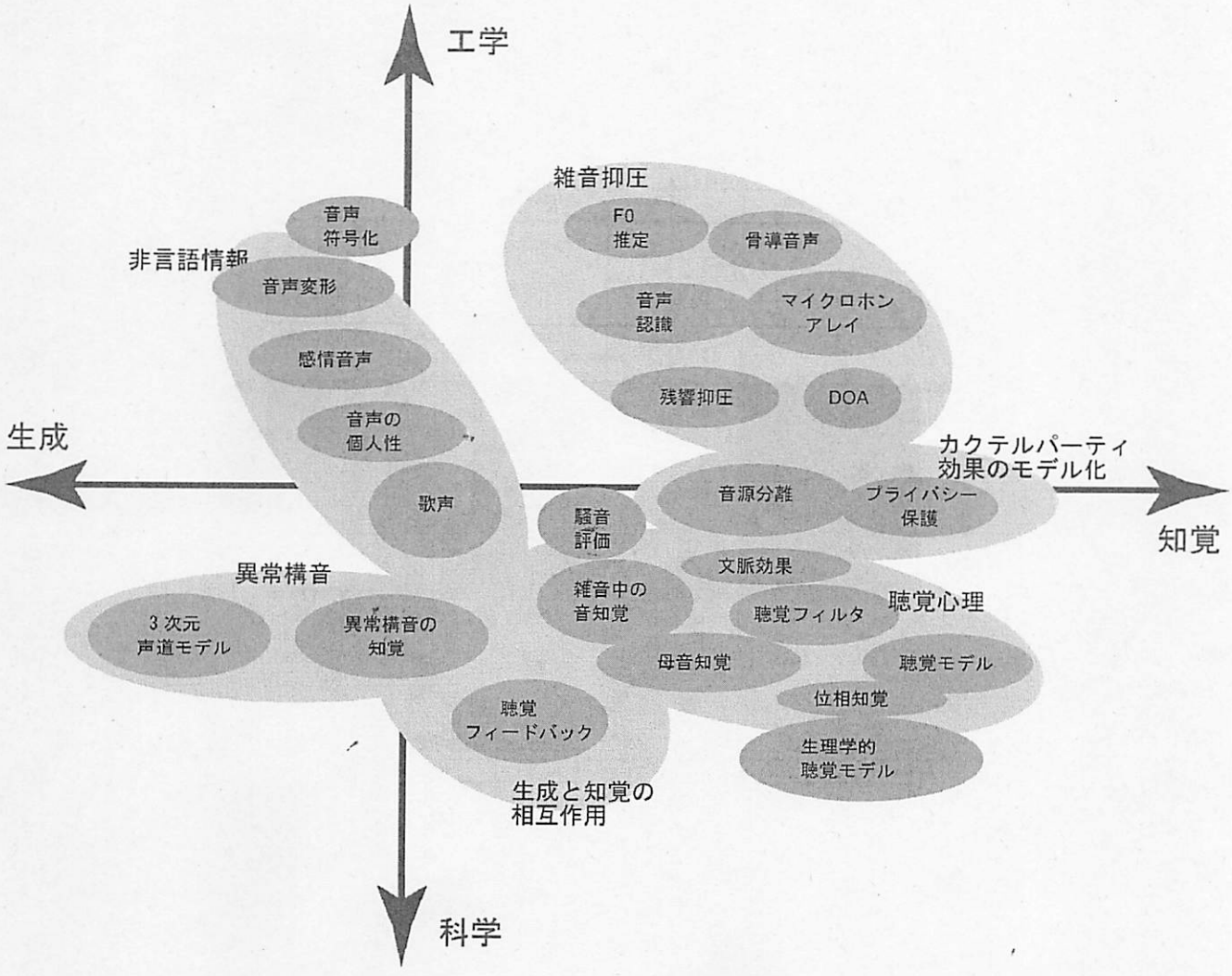

図 2　今までに JAIST で行った研究項目

B)　「聞く」: 雑音とか残響が存在する実環境でのヒトのすばらしい聴取能力を，少しでも機械の耳に与えて賢くするために，カクテルパーティ効果の実現，雑音中の音声強調などの研究

C)　「話す・聞く」の基礎検討：これらの基礎となる，心理データ，生理データにもとづいた音声生成・知覚機構のモデル化の研究

について，重点的に研究を行ってきた。現在までに行った（あるいは行っている）研究題目の関連図を図 2 に示す。以後，図 2 の各項目に従って，研究内容の説明を行う。

4．研究項目の説明

4.1 非言語情報

音声により送受される情報は，言語のみならず，非言語情報である感情，年齢・性別，話者の社会的ステータスまで様々である。当研究室では，音

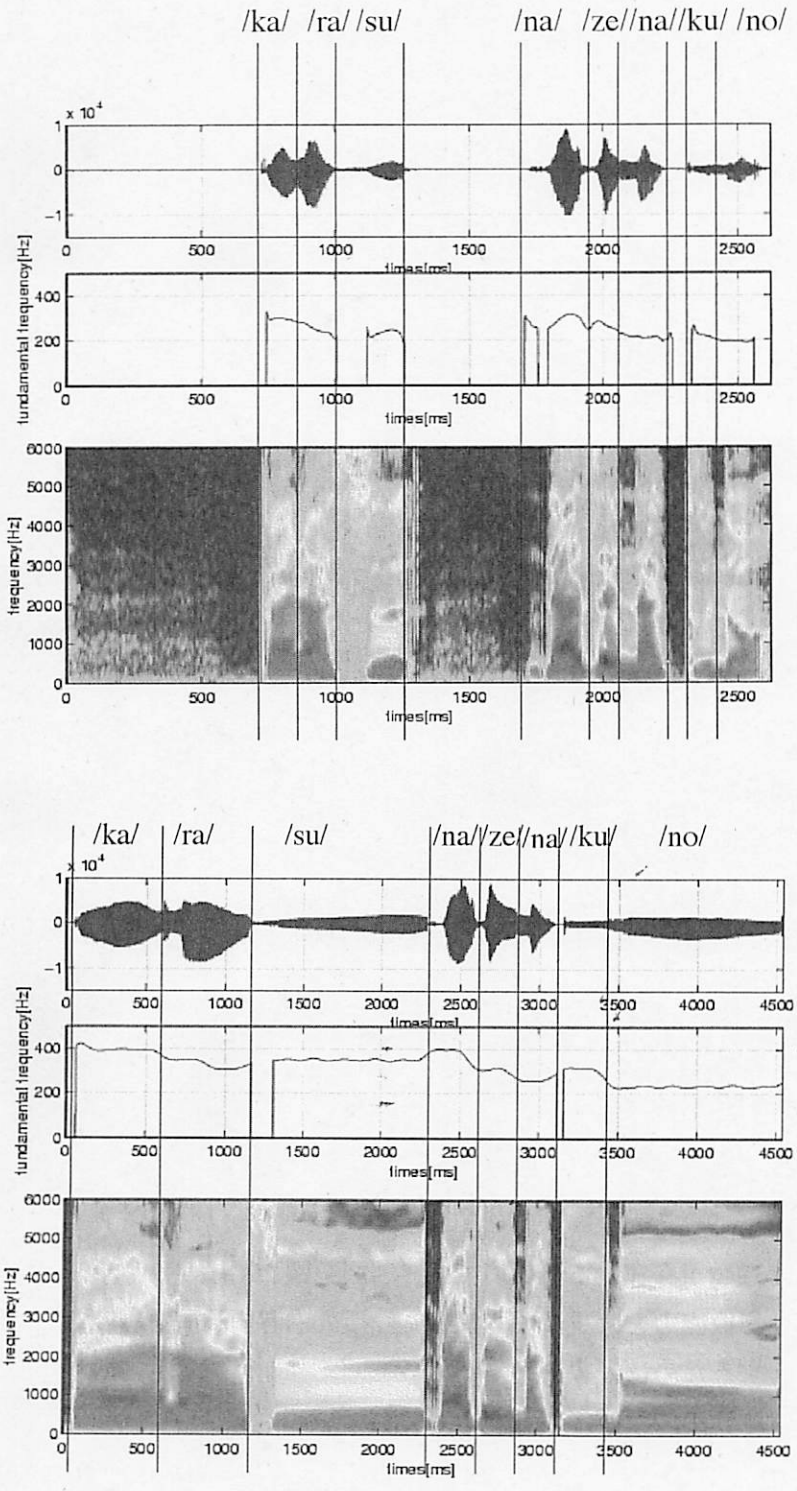

図3 （上）話声（“からすなぜなくの”），（下）話声から合成した歌声：話し手本人が歌っているように聞える。

声コミュニケーションにおける非言語情報の役割を明らかにすることを目的として，個人性，感情，歌声等をターゲットとして，非言語情報の生成・知覚，合成・認識について研究を行っている。現在までに，個人性知覚・感情知覚に関連する音響特徴量の同定[1]-[5]，非言語情報付加のための様々な手法の開発[6]-[9]，これらを用いた感情音声[5]，歌声の合成[10]，歌声知覚における脳活動計測[11]等を実施している。最近の成果として，歌声合成のコンテストである InterSpeech2007 Synthesis of Singing Challenge において第 1 位を獲得した[12]。また，この研究は，総務省戦略的情報通信研究開発推進制度（SCOPE）に採択されている。

4.2 音声回復

雑音・残響が存在する環境においては，人の音声了解度は著しく低下する。また，機械による音声認識システムにとっても，認識率の低下は免れない。そこで当研究室では，実環境に存在する雑音・

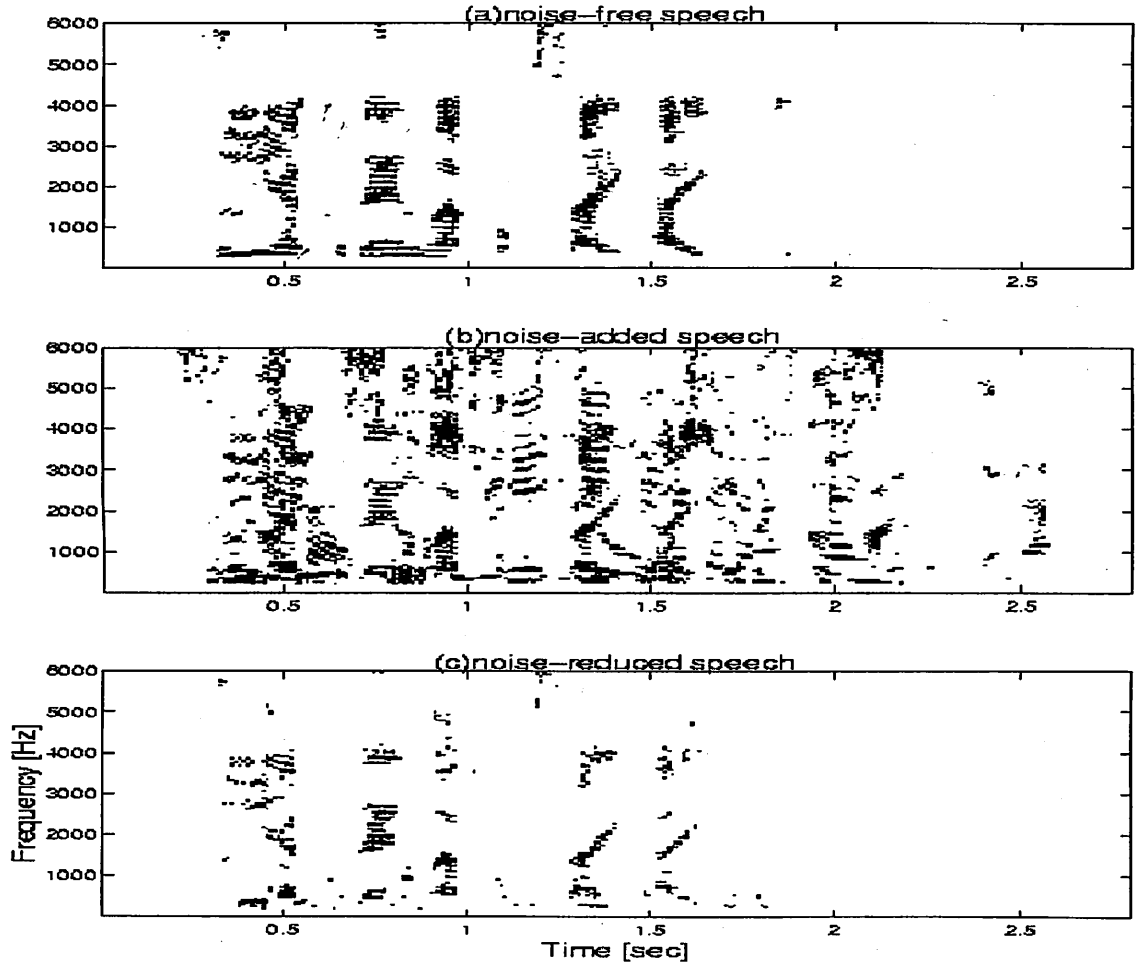

図 4　音声強調の例：マイクロホンアレイを用いて，4 人の同時発話（真中）から 1 人の音声を強調する（下）。上が取り出された音声の元音声。[14]

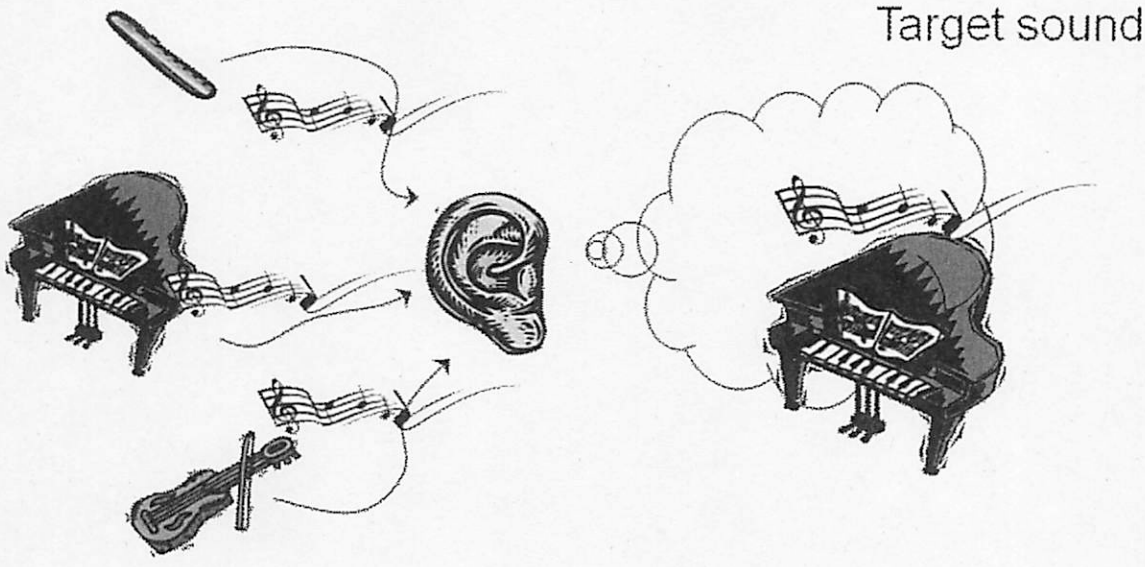

図5　カクテルパーティ効果における「聞き耳」(Attentive Hearing) のモデル化：複数の楽器が異なる旋律を演奏している状況で一つの楽器に注目して波形を強調する。[26]

残響に邪魔されないコミュニケーションの確立を目指して，マイクロホンアレイを用いた雑音抑圧および音声強調[13]-[18], 骨導音声によるコミュニケーション支援[19][20]，残響抑圧[21]について研究を行っている。また，これらを応用して，実環境での頑健な音声特徴抽出[22]および認識[23]についても研究を行っている。今後は，雑音・残響抑圧法の福祉機器（特に Hearing Aid）への応用，高雑音環境での音声通信などへの応用を試みる予定である。

4.3 カクテルパーティ効果のモデル化

雑音中での人の音声抽出過程（音源分離過程：カクテルパーティ効果）について調査を行い，これをモデル化することで，複数の音源の中から目的音を分離抽出する手法[24]-[26], これを応用して音声認識システムを構築する手法[27], 走行雑音が存在する車室内での効率的な報知音の呈示方法[28][29]，また，カクテルパーティ効果を逆手にとって会話におけるプライバシー保護を目的として音声了解度の低下を促進させる手法[30]について研究を行っている。今後は，複数の音源の中から目的音を知覚するメカニズムについて，心理物理学的手法を適用して，さらに深く基礎的検討を行い，応用システムの性能向上を目指す予定である。

4.4 聴覚心理モデル

ヒトの聴覚特性を調べ，これをモデル化するために，主に聴覚心理の立場から，モデル化の基礎となる様々な心理物理測定を行っている。研究内容は，位相知覚，音声知覚（母音知覚，文脈効果）等，多岐にわたる。そして，これらをもとに，聴覚マスキング特性のモデル化[31]，文脈効果のモデル化と音声認識への応用[32][33]，聴覚末梢系モデルの騒音評価への応用[34][35]を行ってきた。今後は，「非言語情報の生成と知覚」とからめた知覚

モデルの構築を推進していく予定である。

4.5 生理学的聴覚モデル

ヒトの聴覚特性を調べ，これをモデル化するために，主に聴覚生理の立場からモデル化を行っている。モデルを構築する場合，次の二種類のモデル化が考えられる。

（1．実態モデル）モデルによる真理追求のアプローチ：生理学，心理学において実体を用いて実験できない場合，精巧なモデルを用いて計算機上でシミュレーションを行い，様々な知見を得るためにモデル化。

（2．機能モデル）工学応用：人間は鳥を見て空に憧れ飛行機を作った。飛行機は鳥と同じように空を飛んでいるわけではないが揚力という物理学の基本原理は同じである。このように，基本原理を見つけだして工学的に応用することを試みるためのモデル化。

当研究室では，実態モデルとして音源方向定位をつかさどる蝸牛神経核および上オリーブ核のモデル化[36]，機能モデルとして聴覚有毛細胞→聴神経→蝸牛神経核→下丘にいたる初期聴覚系のモデル化[37][38]を行っている。

4.6 異常構音

口腔疾患，運動機能障害等のために構音が正常にできず，発話した音声にひずみを生じることがある。このひずみがどのような形態の構音から発せられるのか，また，ひずみと知覚される主原因は何か，を明確にすることは，発話訓練補助のみならず，人の音声生成・知覚機構を解明する上で有益である。当研究室では，昭和大学歯学部，東

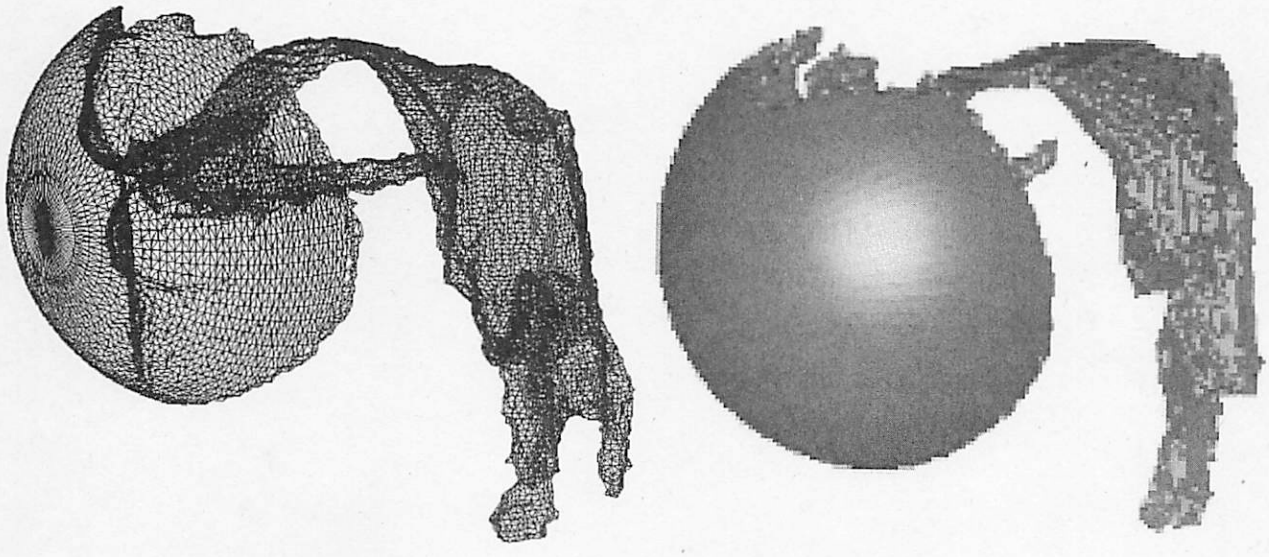

図6　/i/ 発話時のワイヤーフレームモデル（左）および3次元声道モデル（右）：MR画像より声道形状を抽出し，ワイヤーフレームモデルを作成する。その後ソリッドモデルである3次元声道モデルを作成し，有限要素法を用いて伝達特性を推定する。口唇からの放射を表現するために放射球面を取り付けている。[42]

京医科歯科大学と共同して，側音化構音音声[39]，癌による舌除去後の音声[40]，口蓋裂音声[41]等の分析を行うとともに，聴取実験により異常構音と知覚されるための音響物理関連量を明らかにしてきた。また，MRI によって取得した声道形状から有限要素法を用いて伝達特性をシミュレートすることで異常構音の生成機構の解明[42]を試みてきた。

4.7 生成と知覚の相互作用

音声知覚・生成は，音声による人ー人コミュニケーションの根幹を成すものである。また，人ー機械コミュニケーションにおいても，ヒトの音声生成・知覚機構を基礎として，これを工学的に実現した音声合成・認識が重要な役割を果たそうとしている。本来，音声知覚・生成は，音声コミュニケーションにおいて表裏一体を成すものであり，コミュニケーションを円滑に保つためには双方が一体となって働く必要がある。当研究室では，音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ，知覚・生成の相互作用の解明を図ることを目的として，様々な生理指標の測定を試みている[44][45]。

現在までに

1. 被験者のフィードバック音の変形への反応は，変形の方向と反対方向であり，発話において変形に対する補正がリアルタイムで行われていることが確認できた。
2. ホルマント周波数分析から，第 1，第 2 ホルマントにおいて明確な補償動作が観測された。これは，スペクトルに関する聴覚フィードバックにおいて短時間での補正反応を捕らえた初めての結果である。
3. 筋電（EMG）および舌運動（EMA）の分析結果から，摂動に対してこれを補償するような筋肉および舌の動きが観測された。
4. 補償動作は，変形開始から約 150 ms で始まり，290 ms で最大値に到達した。

ことが明らかとなっている。今後，fMRI および MEG 等を用いて脳活動の測定を行い，より詳細な結果を得る予定である。

5. おわりに

北陸先端科学技術大学院大学（JAIST）・赤木研究室で行っている（行われた）研究内容および成果の概要を説明した。詳細には説明できなかったが，もし興味がおありならば，大学の教員一覧[46]あるいは赤木個人のホームページ[47]からより詳細な情報が得られるはずである。また，参考文献に挙げた論文の一部は，JAIST リポジトリ[48]からコピー可能である。

参考文献

1. Akagi, M. and Ienaga, T. (1997). "Speaker individuality in fundamental frequency contours and its control", J. Acoust. Soc. Jpn. (E), 18, 273-80.（日本音響学会佐藤論文賞）
2. Kitamura, T. and Akagi, M. (1995). "Speaker individualities in speech spectral envelopes", J. Acoust. Soc. Jpn. (E), 16, 5, 283-289.
3. 赤木正人(2005). "表現豊かな音声 —その生成・知覚と音声合成への応用—", 日本音響学会誌, 61, 6, 346-351.
4. Sawamura K., Dang J., Akagi M., Erickson D., Li, A., Sakuraba, K., Minematsu, N., and Hirose, K. (2007). "Common factors in emotion perception among different cultures," Proc. ICPhS2007, 2113-2116.
5. Huang, C-F. and Akagi, M. (2008) "A

three-layered model for expressive speech perception," Speech Communication 50, 810-828.

6. Nguyen, P. C., Ochi, T., and Akagi, M. (2003). "Modified Restricted Temporal Decomposition and its Application of Low Rate Speech Coding," IEICE Trans. Inf. & Syst., E86-D, 3, 397-405.
7. Nguyen, P. C., Akagi, M., and Nguyen, P. B. (2007). "Limited error based event localizing temporal decomposition and its application to variable-rate speech coding," Speech Communication, 49, 292-304.
8. Nguyen B. P. and Akagi M. (2007). "A flexible spectral modification method based on temporal decomposition and Gaussian mixture model," Proc. Interspeech2007, 538-541.
9. Tomoike, S. and Akagi, M. (2008). "Estimation of local peaks based on particle filter in adverse environments," Journal of Signal Processing, 12, 4, 303-306.
10. Saitou, T., Unoki, M. and Akagi, M. (2005). "Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis," Speech Communication 46, 405-417.
11. Nakamura, T., Kitamura, T. and Akagi, M. (2009). "A study on nonlinguistic feature in singing and speaking voices by brain activity measurement," Proc. NCSP'09, 217-220.
12. Saitou, T., Goto, M., Unoki, M., and Akagi, M. (2007). "Vocal conversion from speaking voice to singing voice using STRAIGHT," Proc. Interspeech2007, Singing Challenge. (音声関連の国際会議において歌声合成コンテスト，第1位)
13. Mizumachi, M. and Akagi, M. (1998). "Noise reduction by paired-microphones using spectral subtraction," Proc. ICASSP98, II, 1001-1004
14. Akagi, M. and Kago, T. (2002). "Noise reduction using a small-scale microphone array in multi noise source environment," Proc. ICASSP2002, Orlando, I-909-912.
15. Li, J. and Akagi, M. (2006). "A noise reduction system based on hybrid noise estimation technique and post-filtering in arbitrary noise environments," Speech Communication, 48, 111-126.
16. Li, J. and Akagi, M. (2008). "A hybrid microphone array post-filter in a diffuse noise field," Applied Acoustics 69, 546-557.
17. Li, J., Sakamoto, S., Hongo, S., Akagi, M., and Suzuki, Y. (2008). "Adaptive l.l-order generalized spectral subtraction for speech enhancement," Signal Processing, vol. 88, no. 11, pp. 2764-2776, 2008.
18. Li, J., Sakamoto, S., Hongo, S., Akagi, M., and Suzuki, Y. (2009). "Two-stage binaural speech enhancement with Wiener filter based on equalization-cancellation model," Proc. WASPAA, New Palts, NY, 133-136.
19. Vu, T., Unoki, M., and Akagi, M. (2006). "A Study on Restoration of Bone-Conducted Speech with MTF-Based and LP-based Models," Journal of Signal Processing, 10, 6, 407-417.
20. Kinugasa, K., Unoki, M., and Akagi, M. (2009). "An MTF-based method for Blind Restoration for Improving Intelligibility of Bone-conducted Speech," Journal of Signal Processing, 13, 4, 339-342.
21. Unoki, M., Yamasaki, Y., and Akagi, M. (2009/08/25). "MTF-based power envelope restoration in noisy reverberant environments," Proc. EUSIPCO2009, Glasgow, Scotland, 228-232.

22. Ishimoto, Y. and Akagi, M. (2004). "Fundamental frequency estimation for noisy speech using entropy-weighted periodic and harmonic features," IEICE Trans. Inf. & Syst., E87-D, 1, 205-214.
23. Lu, X., Unoki, M., and Akagi, M. (2008/11/1). "Comparative evaluation of modulation-transfer-function-based blind restoration of sub-band power envelopes of speech as a front-end processor for automatic speech recognition systems," Acoustical Science and Technology, 29, 6, 351-361.
24. Unoki, M. and Akagi, M. (1998). "A method of signal extraction from noisy signal based on auditory scene analysis," Speech Communication, 27, 3-4, 261-279.
25. Akagi, M., Mizumachi, M.,Ishimoto, Y., and Unoki, M. (2000). "Speech enhancement and segregation based on human auditory mechanisms", Proc. IS2000, Aizu, 246-253.
26. Unoki, M., Kubo, M., Haniu, A., and Akagi, M. (2006). "A Model-Concept of the Selective Sound Segregation: — A Prototype Model for Selective Segregation of Target Instrument Sound from the Mixed Sound of Various Instruments —," Journal of Signal Processing, 10, 6, 419-431.
27. Haniu, A., Unoki, M., and Akagi, M. (2009). "A psychoacoustically-motivated conceptual model for automatic speech recognition," Proc. WESPAC2009, Beijing, CD-ROM.
28. Nakanishi, J., Unoki, M., and Akagi, M. (2006). "Effect of ITD and component frequencies on perception of alarm signals in noisy environments," Journal of Signal Processing, 10, 4, 231-234.
29. Uchiyama, H., Unoku, M., and Akagi, M. (2007). "Improvement in detectability of alarm signals in noisy environments by utilizing spatial cues," Proc. WASPAA2007, New Paltz, NY, pp.74-77.
30. 太長根，赤木，入江(2005). "音源の知覚的融合にもとづいた会話のプライバシー保護の検討", 平成17年春季音響学会講演論文、1-2-3.
31. Unoki, M. and Akagi, M. (2001). "A computational model of co-modulation masking release," in Computational Models of Auditory Function, (Eds. Greenberg, S. and Slaney, M.), NATO ASI Series, IOS Press, Amsterdam, 221-232.
32. Akagi, M., van Wieringen, A. and Pols, L. C. W. (1994). "Perception of central vowel with pre- and post-anchors", Proc. Int. Conf. Spoken Lang. Process. 94, 503-506.
33. 米沢、赤木(1997). "文脈効果のモデル化とそれを用いたワードスポッティング"、電子情報通信学会論文誌、J80-D-II, 1, 36-43.
34. Mizumachi, M. and Akagi, M. (2000). "The auditory-oriented spectral distortion for evaluating speech signals distorted by additive noises," J. Acoust. Soc. Jpn. (E), 21, 5 251-258.
35. Akagi, M., Kakehi, M., Kawaguchi, M., Nishinuma, M., and Ishigami, A. (2001). "Noisiness estimation of machine working noise using human auditory model", Proc. Internoise2001, 2451-2454.
36. Ito, K. and Akagi, M. (2005). "Study on improving regularity of neural phase locking in single neurons of AVCN via a computational model," In Auditory Signal Processing, Springer, 91-99.
37. 牧，赤木，廣田(2004). "蝸牛神経核背側核細胞の周波数応答特性に関する神経回路モデルの提案—トーンバースト刺激に対する応答

—”, 日本音響学会誌, 60, 1, 3-11.
38. Maki, K. and Akagi, M. (2005). "A computational model of cochlear nucleus neurons," In Auditory Signal Processing, Springer, 84-90.
39. Akagi, M., Suzuki, N., Hayashi, K., Saito, H., and Michi, K. (2001). " Perception of Lateral Misarticulation and Its Physical Correlates", Folia Phoniatrica et Logopaedica, 53, 6, 291-307
40. 斉藤、鈴木、藤田、道、高橋、赤木、和久本 (2000). ”MR 撮像法を用いた 3 次元声道形状の計測 −舌・口底切除症例の検討−”、昭和歯学会雑誌、20, 2, 198-214.
41. Kozaki-Yamaguchi, Y., Suzuki, N., Fujita, Y., Yoshimasu, H., Akagi, M., and Amagasa, T. (2005). "Perception of hypernasality and its physical correlates," Oral Science International, 2, 1, 21-35.
42. 西本, 赤木, 北村, 鈴木(2006). “有限要素法による声道伝達特性推定の有効性に関する検討”, 日本音響学会誌, 62, 4, 306-315.
43. Dang, J., Akagi, M., and Honda, K. (2006). "Communication between speech production and perception within the brain - Observation and simulation," J. Comp. Sci. & Tech., 21, 1, 95-105.
44. Matsuoka, R., Lu, X., Dang, J., and Akagi, M. (2004). “Investigation of interaction between speech perception and speech production,” Proc. KIT Int. Sympo. Brain and Language 2004, 27-28.
45. Dang, J., Akagi, M., and Honda, K. (2006). "Communication between speech production and perception within the brain - Observation and simulation," J. Comp. Sci. & Tech., 21, 1, 95-105.
46. http://www.jaist.ac.jp/is/2008ja/kenkyu/ichiran/
47. http://www.jaist.ac.jp/~akagi/
48. https://dspace.jaist.ac.jp/dspace/

赤木 正人 昭 54 名工大・工・電子卒。昭 59 東工大大学院博士課程情報工学専攻了。工博。同年電電公社（現 NTT）研究所入社。以来，ATR 視聴覚機構研究所，NTT 基礎研究所を経て，現在，北陸先端科学技術大学院大学情報科学研究科教授。この間，米国 MIT，オランダアムステルダム大学，英国ケンブリッジ大学滞在研究員。音声信号処理，聴覚機構のモデル化の研究に従事。昭 62 電子情報通信学会論文賞，平 9 および 17 日本音響学会佐藤論文賞，平 21 信号処理学会 Best Paper Award を受賞。日本音響学会，電子情報通信学会，信号処理学会，IEEE，ASA，ISCA 各会員。