

音声コミュニケーションにおける Speech Chain を考える

荒井隆行^{†1}

概要: 私たちは相手に伝えたいメッセージを言語化し、音声器官を動かして音響信号にする。音響信号は相手に受け取られると、聴覚器官を介して再びもとのメッセージに変換される。このような音声コミュニケーションに見られる連鎖、すなわち「Speech Chain (ことばの鎖)」は、音声生成と音声知覚が表裏一体となっていることを示している。Speech Chain はその名が誕生してから久しいが、未だ説明されていない点も多い。Speech Chain を通して、音声科学や音声学を改めて考える。

キーワード: 音声コミュニケーション, ことばの鎖, 音声科学, 音声学

Speech Chain in Speech Communication

TAKAYUKI ARAI^{†1}

1. はじめに

音声コミュニケーションにおいて、話し手のメッセージが聞き手に伝わるまでの様子を考えてみる。まず話し手は伝えたいメッセージを頭に思い浮かべる。そのメッセージは、具体的な語の並びとして話し手の頭の中で表現され、そしてそれに基づき話し手は運動神経を介して筋肉を動かし音声を発する。一方、その音声信号は聞き手の聴覚器官に入り、情報が末梢系から中枢へと伝えられる。そして、もともとのメッセージが聞き手の頭の中で理解される。このように、音声コミュニケーションに見られる各段階での事象の連鎖は「Speech Chain (ことばの鎖)」と呼ばれ、これは基本的な音声伝達の様子を表現したものである [1]。

現代の社会はコンピュータや情報通信ネットワークの発展と共に進歩し続けており、「ことばの鎖」も複雑化している。例えば、話し手と聞き手の間に電話回線やインターネットを始め、情報通信ネットワークが介在するケースもあるし、また話し手や聞き手のどちらかがコンピュータやロボットであることもある。音声符号化や音響信号処理、自動音声認識、音声合成などの音声学分野は、近年の深層学習やビッグデータの利用などとも相まってさらなる発展をみせている。

一方、この「ことばの鎖」自身、実際にはもっと複雑な過程からなっている。例えば、話し手の音声には話し手が発信しようとしたメッセージ以外に、話し手の個性や発話時の感情に関わる情報も含まれる。聴覚情報以外に話し手の表情やジェスチャーなどの視覚情報も聞き手が同時に受け取るなどである。結果的に、そこには音声学・音韻論を含む言語学、医学・生理学、脳神経科学、心理学、音響

学など、多くの学問が関係しており、それらが相互に影響し合っている。さらに、もし「ことばの鎖」のどこかに何らかの障害が存在すれば、それはコミュニケーション障害につながる。そのような観点から、音声言語医学、言語障害学、社会福祉学などの分野とも直結している。

このように、音声コミュニケーション分野には音声学の側面と音声科学や聴覚科学の側面があり、それらはいわば両輪を成している。そのような中、音声コミュニケーションにおいてその応用研究のみならず、音声生成や音声知覚などの音声に関わる基礎研究も重要であることは言うまでもない[2]。

2. 具体的な研究例

そのような背景のもと、私が取り組んでいるものの中には次のようなものがある。

2.1 声道模型

近代における音声研究において、例えば千葉・梶山による The Vowel [3]では声道模型を用いた母音の生成実験の話が書かれている。我々は The Vowel の出版 60 周年を記念して声道模型を復元した[4]。そのことをきっかけにして、今では静的な模型だけではなく、声道形状を変化させることのできるような動的な模型も開発し、様々な応用にもその裾野を広げている。それらは、音声を専門的に学ぶ方々のための音響教育のみならず、音声研究そのものに応用したり、子どもから大人まで広く多くの方々に音声の仕組みを理解していただくための展示やアウトリーチ活動（博物館・科学館での展示や科学教室・工作教室などを含む）、音声教育などの言語教育応用、言語障害者のための臨床応用、歌声合成などである[5][6][7]。

^{†1} 上智大学
Sophia University

2.2 第2言語学習者による音声生成・音声知覚と言語教育

第2言語の習得において、音声生成や音声知覚に関する母語の干渉は重要なテーマである。声道模型を用いて発音や聞き取りの訓練を行った研究では、学習者が自ら手を動かして声道模型を操作しながら反復練習を行うことによって、生成される母音の質がより改善することを示唆する結果が得られている[8]。調音の様子を科学的に観測することも重要であり、また視覚的に学習者に見せることも大事である。その点、超音波診断装置を用いた調音運動の観測も近年、盛んである。それを応用し、超音波画像や声道模型の様子を人間の顔にプロジェクションマッピングする技術も試みている[9]。最近では小学校での英語教育が話題となっているが、筆者はNHK（Eテレ）の子ども向け英語番組において、発音のための実験に関する監修も行っている。

2.3 マルチモーダル

声道模型を使って手で「発音訓練」を行う行為は、いわば触覚や聴覚、視覚を含む、マルチモーダルなアプローチと言える。また、視覚情報が聴覚情報処理を助けたり影響を与えることもしばしばある。人間の知覚においては、McGurk 効果[10]が有名であるが、日本語促音に関する音声知覚の実験でも、視覚情報が聴覚情報処理に影響を与える[11]など、興味深い結果も得ている。

2.4 雑音・残響下を含む音声認識

自動音声認識では静寂下での認識率の高い一方、雑音や残響など劣悪な環境では認識率が下がることがしばしば問題となる。人間の認識率が機械を上回ることも少なくない[12]。そこで自動音声認識では様々な音環境を想定して認識実験が行われることも多い。過去には、人間および機械による音声認識における重要な変調周波数帯域を比較し、両者の類似性を評価した[13][14]。一方、人間の音声知覚の限界はどうなのだろうか。その追求はすべてがすでに調べつくされているというわけではない。人間の音声情報処理を含め、未解明の点も多い。

2.5 医学や福祉に関して

音声に関わる病気の診断や、音声を利用した病気の診断もある。また、音声や音情報のバリアに対する支援、そして音声や音を利用した支援なども重要である。その一例として、筋委縮性側索硬化症（ALS）の患者などが声を失う前に自分の音声を録音し、いずれ声を失った後も音声合成技術を活用して自分自身の声でコミュニケーションを行うことを可能にする「マイボイス」プロジェクト[15]などにも参画している。

3. おわりに

以上、ご紹介した具体的な例はあくまでもその一例にしかならず、この他、多くの研究に日々関わっている。しかしそのほとんどが「Speech Chain（ことばの鎖）」そのものの研究であり、人間の音声コミュニケーション活動に関わるものである。人間が行う高度な音声言語処理を探究するということは結局、人間を知るということに他ならない。

謝辞 内容の一部は日本学術振興会の科学研究費（15K00930）の助成を得た。

参考文献

- [1] Denes, P. B. and Pinson, E. N., *The Speech Chain: The Physics and Biology of Spoken Language*, 2nd ed., W. H. Freeman, New York, 1993.
- [2] 荒井隆行, “音声生成と音声知覚を中心に人間による音声コミュニケーションに関する研究を考える,” 日本音響学会春季研究発表会講演論文集, pp. 1375-1376, 2017.
- [3] Chiba, T. and Kajiyama, M., *The Vowel, Its Nature and Structure*, Tokyo-Kaiseikan, 1941-42.
- [4] Arai, T., “The replication of Chiba and Kajiyama’s mechanical models of the human vocal cavity,” *J. Phonetic Soc. Jpn.*, vol. 5, no. 2, pp. 31-38, 2001.
- [5] Arai, T., “Education system in acoustics of speech production using physical models of the human vocal tract,” *Acoust. Sci. Tech.*, vol. 28, no. 3, pp. 190-201, 2007.
- [6] Arai, T., “Education in acoustics and speech science using vocal-tract models,” *J. Acoust. Soc. Am.*, vol. 131, no. 3, Pt. 2, pp. 2444-2454, 2012.
- [7] Arai, T., “Vocal-tract models and their applications in education for intuitive understanding of speech production,” *Acoust. Sci. Tech.*, vol. 37, no. 4, pp. 148-156, 2016.
- [8] Arai, T., “Hands-On Tool Producing Front Vowels for Phonetic Education: Aiming for Pronunciation Training with Tactile,” *Proc. of INTERSPEECH*, pp.1695-1699, Dresden, 2015.
- [9] Arai, T. and Uchida, R., “Projection mapping of articulation for education in phonetic science,” *Acoust. Sci. Tech.*, vol. 37, no. 4, pp. 175-177, 2016.
- [10] McGurk, H. and MacDonald, J., “Hearing lips and seeing voices,” *Nature*, vol. 264, pp. 746-748, 1976.
- [11] Arai, T., Iwagami, E. and Yanagisawa, E., “Seeing closing gesture of articulators affects speech perception of geminate consonants,” *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. EL319-EL325, 2017.
- [12] Arai, T., “Time-reversed reverberation yields lower speech recognition rate by human and machine,” *Acoust. Sci. Tech.*, vol. 34, no.2, pp. 142-146, 2013.
- [13] Arai, T., Pavel, M., Hermansky, H. and Avendano, C., “Syllable intelligibility for temporally filtered LPC cepstral trajectories,” *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2783-2791, 1999.
- [14] Kanedera, N., Arai, T., Hermansky, H. and Pavel, M., “On the relative importance of various components of the modulation spectrum for automatic speech recognition,” *Speech Communication*, vol. 28, pp. 43-55, 1999.
- [15] Arai, T., Kawahara, S., Homma, M. and Yoshimura, T., “A speech synthesizer of Japanese saves patients’ voices,” *J. Acoust. Soc. Am.*, vol. 140, no. 4, Pt. 2, p. 3365, 2016.