

韻律類型論への音響的アプローチ —音源特徴の聴取による韻律類型の識別—

小松 雅彦(北海道医療大学) 荒井 隆行・菅原 勉(上智大学)
koma2@hoku-iryu-u.ac.jp, {arai, sugawara}@sophia.ac.jp

1. はじめに

本発表では、音源特徴(音源フィルタモデルの音源部)の聴取による韻律類型の識別が可能かどうか調べた実験について報告する。

アクセントの種類(トーン・アクセント/ピッチ・アクセント/ストレス・アクセント)やリズムの種類(ストレス・リズム/シラブル・リズム/モーラ・リズム)は、よく知られた概念であるが、このような類型が音響的に定義できるかどうかは、未だ不明確である。図1に、暫定的な韻律類型の配置を示すが、このような配置の根拠は、産出・知覚の両面から今後検討されなければならない。

筆者らは、音源フィルタモデルという音響的なモデルの音源部が言語学的な韻律的特徴にほぼ対応すること(図2 参照)を、一連の音響音声学的な実験によって示してきた(小松他, 2003; Komatsu, 2003)。音源フィルタモデルは、音声を「音源」と「フィルタ」に分割する音響的モデルで、音源には、高さ、強さ、高調波対雑音比(Harmonics-to-Noise Ratio; HNR)が含まれる。大まかに言えば、音源が言語学的な韻律的特徴(超分節素性)を表し、フィルタが分節素性を表す。そして、ソリティという言語学的特徴が音声の大まかな分類を表し、それは、超分節素性であると同時に分節素性でもあると考えられる。小松他(2003)では、音源信号(LPC 残差信号)による日本語子音のソリティの識別実験、韻律類型の異なる4言語の音源特徴の分析により、音源特徴が韻律類型をおおよそ形作っていることを示した。

一方で、言語識別の先行研究を検討すると、その実験手法が、上記の問題にも適用できると考えられる。過去に、被験者に音声を聴取させそれが何語であるか判断してもらう言語識別の実験が多く行われてきた。多くの研究が、通常の音声や合成音を刺激音として用い、識別の手がかりが言語の韻律的特徴であると論じている。また、日英語の識別については、音源フィルタモデルのフィルタ部(スペクトル情報)を削減した刺激音である程度の識別が可能であることが示されている(Komatsu et al., 2002)。従って、韻律類型の異なる言語の音源特徴を抽出した音声を刺激とした言語識別実験を行うことが可能で、それにより韻律類型の相違が音源特徴に現れているかどうかを、聴取実験により調べるのが可能であると考えられる。

2 節では、過去の言語識別の聴取実験を概観する(Komatsu, 2002 等を改訂)。Ramus and Mehler (1999)によって、識別の手掛かりをパラメタ化した実験が初めて行われた。さらに、Komatsu et al. (2002)によって、純粋に音響的操作によるパラメタ化を行っても聴取実験が可能であることが示されたことを述べる。

3 節では、韻律類型の異なる英語、スペイン語、日本語、中国語の音声から、高さ、強さ、HNRが異なる、いくつかの「音源」信号を作成し行った聴取実験(Komatsu et al., in press)について報告する。

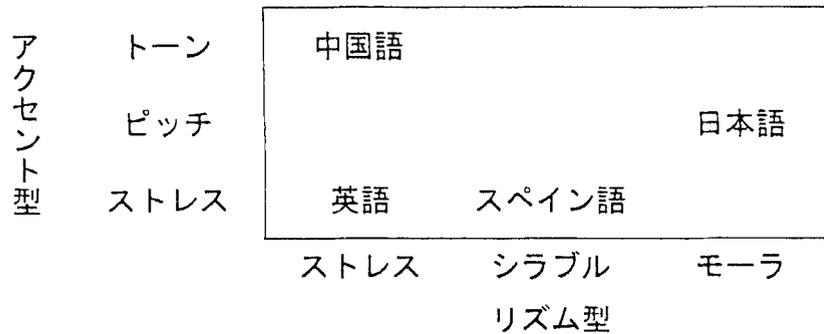


図 1: 韻律類型の暫定的配置

音響モデル	音源	フィルタ	音声
	・高さ ・強さ ・HNR	・スペクトル包絡	
言語モデル (近似)	韻律 ・アクセント ・リズム	ソノリティ ・主要音類	分節索性 音声言語

図 2: 音響的特徴(音源フィルタモデル)と言語学的特徴の対応(小松他, 2003)

2. 言語識別の先行研究

1960年代以降、多くの言語学者、心理学者、工学者によって、言語識別の聴取実験が行われてきた(表 1 参照)。聴取実験では、被験者に原音声あるいは加工された音声を聞いてもらい、言語または方言が識別できるかどうかを調べる。刺激音に用いられてきた加工音声は、原音声の「韻律」特徴を保持していると考えられたもので、ローパスフィルタ通過音声、EGG 波形、原音声の基本周波数・強さ・声のタイミングを模した三角波・正弦波、係数(スペクトル包絡)を一定にして再合成した LPC 再合成音、LPC 残差信号、原音声の強さを模した白色雑音駆動音声、原音声のリズムやイントネーションを保った再合成音と様々である(刺激音についての議論は、Komatsu, 2002 参照)。表 1 に示す研究は、Muthusamy et al. (1994)以外すべて、韻律が言語識別の何らかの手掛かりになっていると述べている。しかし、これらの研究は散発的に行われてきた感があり、実験方法、対象言語とも様々で、ここから韻律の通言語的差異を明確に論じることはできない。(表 1 に挙げた以外にも関連する研究として、非母語話者による発話の自然性の研究[三浦他, 1989; 大山・三浦, 1990]、言語識別と個人性の研究[Bond et al., 1998; Stockmal et al., 1998; Stockmal & Bond, 2002]などがある。)

このような聴取実験によって言語による韻律の差異を調べるには、識別の手掛かりをパラメタ化し、複数種類の刺激音を作成して聴取実験を行い、どの手掛かりが識別に有効であるかを確かめる必要がある。Ramus and Mehler (1999)は、4種類の刺激音を作成して実験を行い、子音・母音の時間的配置がストレス・リズム(英語)とモーラ・リズム(日本語)の識別の手掛かりであると結論付けている。例えば、彼らは、刺激音作成の際に、原音声を音素単位に区切り、子音区間を/s/に、母音区間を/a/に置き換えるような操作を行っている。彼らの方法は、音素という単位を認定した上でそ

れを操作しているため、音韻の概念が必要で、純粋に音響的なアプローチとは言えない。

Komatsu et al. (2002)は、純粋に音響的な手法を採り、LPC 残差信号や強さ曲線を白色雑音駆動したものなどを刺激音として用いている。この実験で、音源フィルタモデルのフィルタ部(スペクトル情報)を削減した刺激音で、日英語のある程度の識別が可能であることが示されている。この実験では、韻律を構成すると考えられる音響特徴のパラメタ化は十分でなかったが、このような音響的アプローチによる聴取実験が可能であることが示された。

表 1: 言語識別の先行研究

刺激音	対象言語	参考文献
通常の音声	10 言語 (英語、ファルシ語、フランス語、ドイツ語、日本語、韓国語、中国語、スペイン語、タミル語、ベトナム語)	Muthusamy et al. (1994)
	アラビア語 (6 国の方言)	Barkat & Vasilescu (2001)
	ロマンス語 (フランス語、イタリア語、スペイン語、ポルトガル語、ルーマニア語)	Barkat & Vasilescu (2001)
ローパスフィルタ通過音声	英語、スペイン語	Atkinson (1968)
	アラビア語、英語	Moftah & Roach (1988)
	日本語 (関東方言、関西方言)	麦谷他 (2000)
EGG 波形	英語、フランス語	Maidment (1976, 1983)
	アラビア語、英語	Moftah & Roach (1988)
三角波	英語、日本語、中国語	Ohala & Gilbert (1979)
正弦波	アラビア語 (西部方言、東部方言)	Barkat et al. (1999)
LPC 再合成音	不明	Foil (1986)
LPC 残差信号	英語、日本語	Komatsu et al. (2002)
白色雑音駆動音声	英語、日本語	Komatsu et al. (2002)
再合成音	英語、日本語	Ramus & Mehler (1999)

3. 聴取実験

3.1 実験目的・手順

聴取によるアクセント型とリズム型の識別が可能かどうかを調べ、それがどのように音響特徴と関連しているのかを調べるための聴取実験 (Komatsu et al., in press) について報告する。

対象言語は、韻律類型の異なる 4 言語 (中国語、英語、日本語、スペイン語) で、データは MULTEXT を用いた。MULTEXT は、40 個の文章をそれぞれの言語に翻訳し、それぞれの母語話者がそれを読んだ音声を集めたコーパスである。英語とスペイン語のサンプルは MULTEXT Prosodic Database (Campione, 1998) を、日本語は「日本語 MULTEXT」(β 版) (北澤, 2002) を利用した。中国語サンプルは、筆者らが録音したものを用いた。ただし、実験時間の制約から、40 個の文章のうち 9 個のみを用いた (各言語、3 話者×3 文章)。

上記の原音声から、表 2 に示す 6 種類の音声(長さは 5 秒間)を作成して、刺激音として用いた。Set 1~3 は振幅情報を保持する音声、Set 4 は高さ情報を保持する音声、Set 5~6 はその両方を保持する音声である。

これらの音声は、白色雑音・パルス列から、原音声の音源情報(強さ、HNR、高さ)の一部あるいは全部を保持するように作成されている。ただし、高さは、基本周波数の測定値ではなく、MOMEL アルゴリズム(Hirst et al., 2000)によって平滑化したものを用いた(中国語についてはアルゴリズムを一部修正した)。作成された音声は、原音声のスペクトル情報は保持していない。また、人間の声らしく聞こえるようにするため、約 -6dB/oct のスペクトル傾斜を持たせた。

聴取実験には、被験者として、言語学、言語障害学、音声工学を専門とする大学院生・研究者 20 名の協力を得ることができた。実験では、2 つの異なる言語の音声を続けて聞き、言語の順序を判断してもらった。例えば、Set 1 の中国語音声と Set 1 の英語音声がかえり、選択肢「(1) Chinese – English」「(2) English – Chinese」のどちらの順序であったかを選択してもらうという方法である。ここで、正答が得られれば、中国語と英語の識別ができたとみなす。実験は、Set 1 から Set 6 まで順に進めた。各セットは 6 個のサブセット(言語の組み合わせ:中国語・英語、日本語・スペイン語、中国語・日本語、英語・スペイン語、中国語・スペイン語、英語・日本語)で構成し、各サブセットは 6 試行で構成した。全部で 216 試行(6 試行×6 サブセット×6 セット)で、各被験者について約 1 時間かかった。

表 2: 聴取実験の刺激音

	保持する音響特徴	作成方法
Set 1	強さ	白色雑音
Set 2	強さ	パルス列
Set 3	強さ、HNR	白色雑音+パルス列
Set 4	高さ	パルス列
Set 5	強さ、高さ	パルス列
Set 6	強さ、HNR、高さ	白色雑音+パルス列

3.2 結果

全体として、保持される情報が増えるほど、正答率が向上する、つまり識別が容易になるという結果が得られた。全言語対の平均正答率は、Set 1 から Set 6 がそれぞれ、61.3%、61.1%、63.1%、62.8%、74.7%、79.3%であった。振幅情報のみを保持する Set 1~3 と高さ情報のみを保持する Set 4 の正答率は比較的 low、その両方の情報を保持する Set 5~6 の正答率は高かった。

各言語対ごとの正答率は、図 3 に示すとおりである。図 3 中で、例えば、“CE”は、中国語と英語の対を聞いた時の正答率、つまり中国語と英語とが区別できる程度を表している。図 3 の結果は、リズム型の近い言語の識別(中国語・英語、英語・スペイン語、スペイン語・日本語)が困難であることを示している。また、アクセント型については、音調言語(中国語)とそれ以外の識別は容易であると言える。ただし、中国語・英語の正答率がさほど高くないのは、両言語ともストレスを持っている

ことが影響しているのかもしれない。

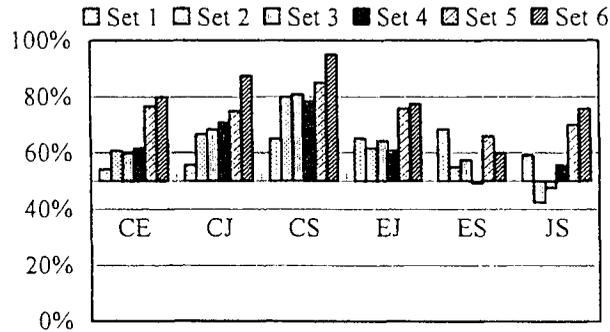


図 3: 各言語対の正答率。C: 中国語、E: 英語、J: 日本語、S: スペイン語。

3.3 刺激音の音響特徴との対応

刺激音の音響分析をいくつかの方法で行った結果、刺激音の音響特徴と聴取実験の正答率との関連が部分的に見られた。高さについては、中国語の基本周波数が目立って速い変動を示しており、それが音調言語(中国語)とそれ以外の識別を可能にしたと推測される。振幅情報の分析では、周期成分・非周期成分の瞬時振幅の分布、および強さ曲線の平均パターンにおいて、日本語が他の言語と異なった特徴を示したが、Set 1~3の正答率では日本語とそれ以外の識別が容易であるという傾向は見られない。実際には、単独の音響特徴ではなく、音響特徴の組み合わせが識別の手掛かりとなっているのであろう。

4. まとめ

筆者らは、韻律類型の異なる英語、スペイン語、日本語、中国語の音声から、高さ、強さ、HNRが異なる、いくつかの「音源」信号を作成し、聴取実験を行った。結果は、全体として韻律類型の識別が可能で、多くの情報が与えられるほど識別が容易であることを示した。さらに、似たリズム型(中国語・英語、英語・スペイン語、スペイン語・日本語)は識別が困難であった。アクセント型については、音調言語(中国語)と非音調言語(中国語以外)の識別が容易であった。また、刺激音の音響分析の結果、中国語の高さの変動がこの識別に関連していると考えられた。

本実験により、音源特徴が韻律類型をおおよそ表すという仮説が、聴取の面からも支持された。従来言われている韻律類型は音響的に定義されたものではなく、今後さらに詳細な実験を行うことによって、韻律類型そのものの再検討も可能になると考えられる。

参考文献

- 大山玄・三浦一郎. (1990). 外国人発話の日本語のプロソディーに関する研究. *日本音響学会講演論文集(平成2年9月)*, 263-264.
- 北澤茂良(編). (2002). *日本語 MULTEXT(β版)*. [CD-ROM]. 静岡大学.
- 小松雅彦・荒井隆行・菅原勉. (2003). 韻律類型論への音響的アプローチ. *第17回日本音声学全国大会予稿集*, 75-80.
- 三浦一郎・大山玄・鈴木博. (1989). 合成音声を用いた日本人英語のプロソディーに関する一検討. *日本音響学会講演論文集(平成元年10月)*, 239-240.

- 麦谷綾子・林安紀子・桐谷滋. (2000). 養育環境にある方言への選好反応の発達: 5~8ヵ月齢乳児を対象に. *音声研究*, 4(2), 62-71.
- Atkinson, K. (1968). Language identification from nonsegmental cues [Abstract]. *Journal of the Acoustical Society of America*, 44, 378.
- Barkat, M., Ohala, J., & Pellegrino, F. (1999). Prosody as a distinctive feature for the discrimination of Arabic dialects. *Proceedings of Eurospeech '99*, 395-398.
- Barkat, M., & Vasilescu, I. (2001). From perceptual designs to linguistic typology and automatic language identification: Overview and perspectives. *Proceeding of Eurospeech 2001*, 1065-1068.
- Bond, Z. S., Fucci, D., Stockmal, V., & McColl, D. (1998). Multi-dimensional scaling of listener responses to complex auditory stimuli. *Proceedings of International Conference on Spoken Language Processing '98*, Vol. 2, pp. 93-95.
- Campione, E. (Ed.). (1998). *MULTEXT prosodic database* [CD-ROM]. Paris: European Language Resources Association.
- Foil, J. T. (1986). Language identification using noisy speech. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing '86*, 861-864.
- Hirst, D., Di Cristo, A., & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 51-87). Dordrecht, The Netherlands: Kluwer Academic.
- Komatsu, M. (2002). What constitutes acoustic evidence of prosody? The use of LPC residual signal in perceptual language identification. *LACUS Forum*, 28, 277-286.
- Komatsu, M. (2003). Essay on acoustic correlates of prosodic typology. In T. Honma, M. Okazaki, T. Tabata, & S. Tanaka (Eds.), *A new century of phonology and phonological theory: A festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday* (pp. 492-507). Tokyo: Kaitakusha.
- Komatsu, M., Arai, T., & Sugawara, T. (in press). *Perceptual discrimination of prosodic types and their preliminary acoustic analysis*. Poster presented at Interspeech 2004, Jeju Island, Korea, October 2004.
- Komatsu, M., Mori, K., Arai, T., Aoyagi, M., & Murahara, Y. (2002). Human language identification with reduced segmental information. *Acoustical Science and Technology*, 23, 143-153.
- Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and Hearing: Work in Progress*, 2, 74-93. University College London.
- Maidment, J. A. (1983). Language recognition and prosody: Further evidence. *Speech, Hearing and Language: Work in Progress*, 1, 133-141. University College London.
- Moftah, A., & Roach, P. (1988). Language recognition from distorted speech: Comparison of techniques. *Journal of the International Phonetic Association*, 18, 50-52.
- Muthusamy, Y. K., Jain, N., & Cole, R. A. (1994). Perceptual benchmarks for automatic language identification. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing '94*, 333-336.
- Ohala, J. J., & Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody. In P. Léon & M. Rossi (Eds.), *Problèmes de prosodie: Vol. 2, Expérimentations, modèles et fonctions* (pp. 123-131). Paris: Didier.
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, 105, 512-521.
- Stockmal, V., & Bond, Z. S. (2002). Same talker, different language: A replication. *Proceedings of International Conference on Spoken Language Processing 2002*, 77-80.
- Stockmal, V., Moates, D. R., & Bond, Z. S. (1998). Same talker, different language. *Proceedings of International Conference on Spoken Language Processing '98*, Vol. 2, 97-100.