

韻律類型論への音響的アプローチ

小松雅彦・荒井隆行（上智大学理工学部）

菅原 勉（上智大学外国語学部）

koma2@splab.ee.sophia.ac.jp

1. はじめに

本発表では、韻律の種類の音響的特徴に関する実験結果の一部を報告し、音源フィルタモデルの音源部が言語学的な韻律的特徴の近似となりうることを示す。

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。従来のアクセント研究では、機能（音韻）の研究や、機能と形式（音響）の関係の研究が主なように思われ、純粋な形式の通言語的研究は少ない。Eady (1982)は、トーン・アクセント言語である中国語とストレス・アクセント言語である英語の高さの比較を行っているが、ピッチ・アクセント言語は扱っていない。また、Eadyは、高さの全体的特徴は比較しているが、局所的特徴は考察していない。リズムについては、Ramus et al. (1999)や Grabe and Low (2002)のように、音響的特徴を定量的にとらえようとする試みが見られる。しかし、彼らの研究は、母音や子音という音韻的単位の存在が前提になっているという点で、純粋に音響的にリズムを定義しているわけではない。本研究は、韻律の種類を、純粋に音響的かつ定量的にとらえることを目的としている。

本発表では、音源フィルタモデルという音響的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というを示す。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（図1上段を参照）。本研究では、アクセントの種類は、Eady (1982)に従い、おもに高さによって特徴づけられると考える。リズムの種類は、Ramus et al. (1999)、Grabe and Low (2002)いずれでも、音声中の子音区間と母音区間の長さによって特徴づけられている。本研究では、強さとHNRを「子音性」「母音性」を示す音響的要素として用いる。強さは、母音性の区間では大きくなり、子音性の区間では小さくなる。HNRは、周期成分の振幅と非周期成分の振幅の比で、母音性の区間では高くなり、子音性の区間では低くなる。

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とみなせるが、厳密には、そのように単純には対応しない（図1参照）。先にも述べたように、リズムを特徴づけるためには、音声の大まかな分類（子音／母音）のようなものが必要である。しかし、子音／母音の別は、音声の分節素性的特徴でもある。本研究では、ソノ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ている。

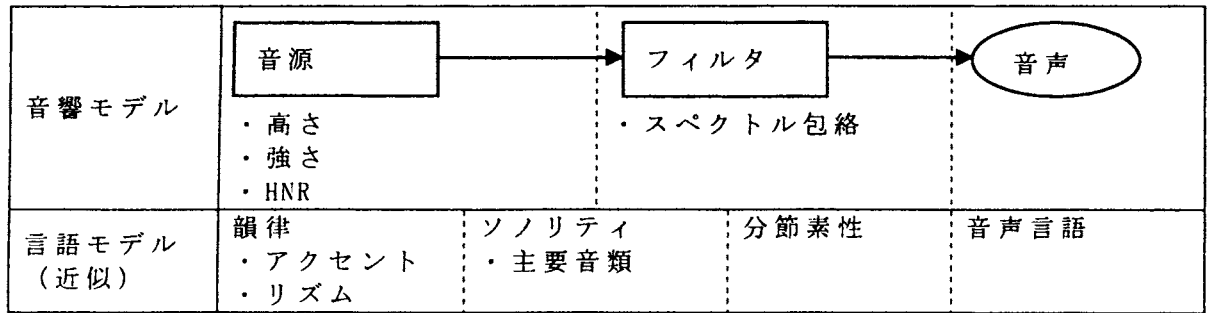


図 1. 音響的特徴（音源フィルタモデル）と言語学的特徴の対応

2. 音源とソノリティ：音源による子音の知覚実験

音源フィルタモデルの音源部に相当する LPC 残差信号を用いた日本語子音の知覚実験 (Komatsu et al., 2000) では、音源にはソノリティ情報が多く含まれている一方で、分節素性的情報が抑制されているという結果が得られた。

この実験では、17 個の日本語 /C+/a/ 音節 ([ka], [ga], [sa], [dza], [ja], [dza], [ta], [da], [tja], [na], [ha], [ba], [pa], [ma], [ja], [ra], [wa]) から、LPC (サンプリング周波数 16 kHz、分析次数 22 次) によって、残差信号を作成した。残差信号の強さは、原音声と同じになるように調整した。また、自然音声に近く聞こえるように、残差信号のスペクトルを約 -6 dB/oct 傾斜させた。この音声信号を刺激音として、15 名の日本語母語話者に対して知覚実験を行った。

知覚実験の結果から、音素と主要音類 (阻害音 / 鼻音 / 流音 / 渡り音; Clements, 1990) の正答率をそれぞれ計算した。例えば、[ka] が [ta] と聞き取られた場合は、音素の正答率の計算では誤答とカウントするが、主要音類の正答率の計算では両者とも阻害音なので正答とカウントする。

正答率は、主要音類については 66.4% と高く、音素については 20.0% と低かった。この実験では、ソノリティが調音方法の知覚の手掛かりになってしまうので、音素の正答率はチャンス・レベル ($1 \div 17 = 5.9\%$) ほど低くはなり得ない。そのことを考慮すると、音素の知覚に必要な情報は、刺激音の中にほとんど含まれていないと考えられる。

さらに、音源部でどのようにソノリティが表されているか調べるため、この知覚結果を、多次元尺度法により分析した (Komatsu et al., 2002)。この分析では、まず、知覚実験の結果から、子音の異聴表を作成した。刺激である LPC 残差信号は音源フィルタモデルの音源に相当するので、子音知覚の混同パターンは、音源における子音の類似性を表すと考えられる。この異聴表を多次元尺度法により分析した。

データの適合度は十分ではなかったが、各子音は、3 次元の知覚空間中に、ソノリティ順位ごとに固まって分布した。図 2 に示すように、無声破裂音、無声摩擦音、有声阻害音、鼻音 / 渡り音は、それぞれ固まっている。

また、知覚空間の各次元は、音響的特徴・音韻素性との対応が見られた (表 1 参照)。各次元は、刺激音の音響的特徴の測定結果と有意な相関があり、また Blevins (1995) でソノリティと関連するとされた音韻素性のいくつかとの対

応が見られた。

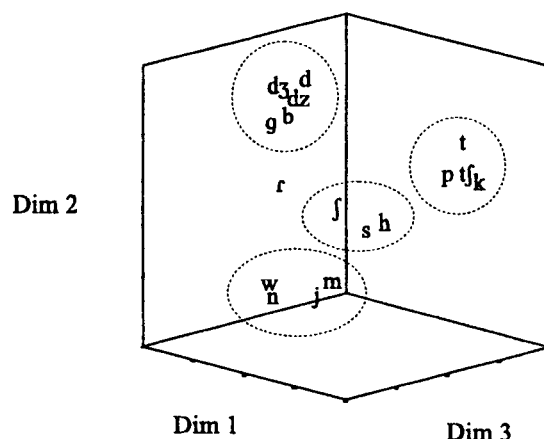


図 2. LPC 残差信号による子音知覚の三次元分析

表 1. 知覚空間の各次元と音響的特徴・音韻素性との対応

| | | ソノリティ 低 | | ソノリティ 高 | |
|------|----|---------------|--|---------------|--|
| 次元 1 | 音響 | HNR 低 | | HNR 高 | |
| | 音韻 | [-voice] | | [+voice] | |
| 次元 2 | 音響 | 振幅 小 | | 振幅 大 | |
| | 音韻 | 高さ 低 | | 高さ 高 | |
| 次元 3 | 音響 | 長さ 短 | | 長さ 長 | |
| | 音韻 | [-continuant] | | [+continuant] | |

3. 韻律類型と音響的特徴

本節では、韻律類型の異なる中国語、英語、日本語、スペイン語の分析 (Komatsu and Arai, in press) を示し、アクセントおよびリズムの類型を定量的に表す音響的特徴の候補について論じる。本節では、特に、音源を構成する音響的要素（高さ、強さ、HNR）の局所的特徴について述べる。データは、1 話者につき、約 30 秒程度の朗読音声である。

3.1 高さ曲線をスタイライズした線分の分布

高さの変化の局所的な形を調べた。まず、図 3 に示すように、もとの高さ曲線を線分でスタイライズした（分解能：2 semitones）。次に、各線分を、図 4 のようにプロットし、線分の特徴の分布を調べた（各言語とも、男性 3 名、女性 3 名の音声データをプロット）。図 4 の各点は、1 つの線分を表す。x 軸方向の値が大きな点は、時間的に長い線分、すなわち、高さがゆっくりと変化する、あるいは長い時間変化しないものを表す。y 軸方向（正または負）の値が大きな点は、高さの速い変化（上昇または下降）を表す。ただし、y 軸方向に極端に大きな値（例えば、1 以上または -1 以下）は、高さ推定の誤りによるものである可能性が高い。

図 4 を見ると、中国語は、速い変化が多いことが分かる。それと対照的に、英語はゆっくりとした変化が多い。日本語は、中国語と英語の間のように見える。スペイン語は、英語に近いが、速い小さな変化がもっと多く見られる。

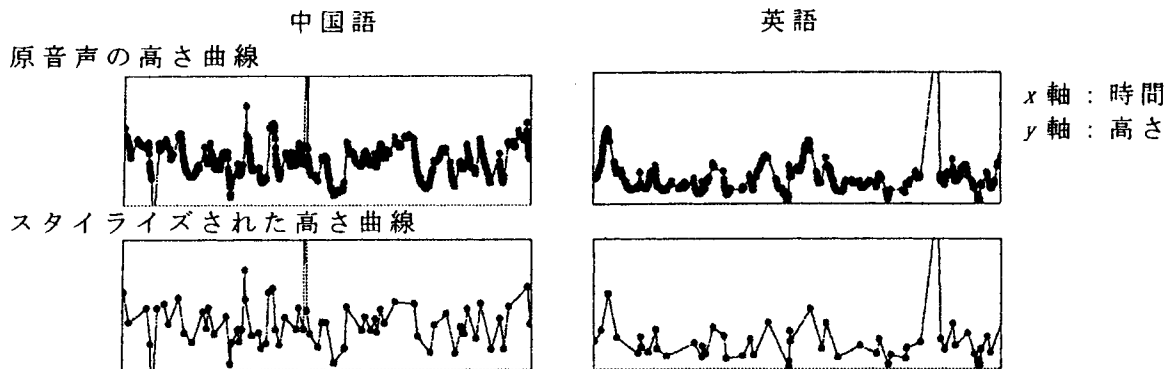


図 3. 高さ曲線のスタイライズ

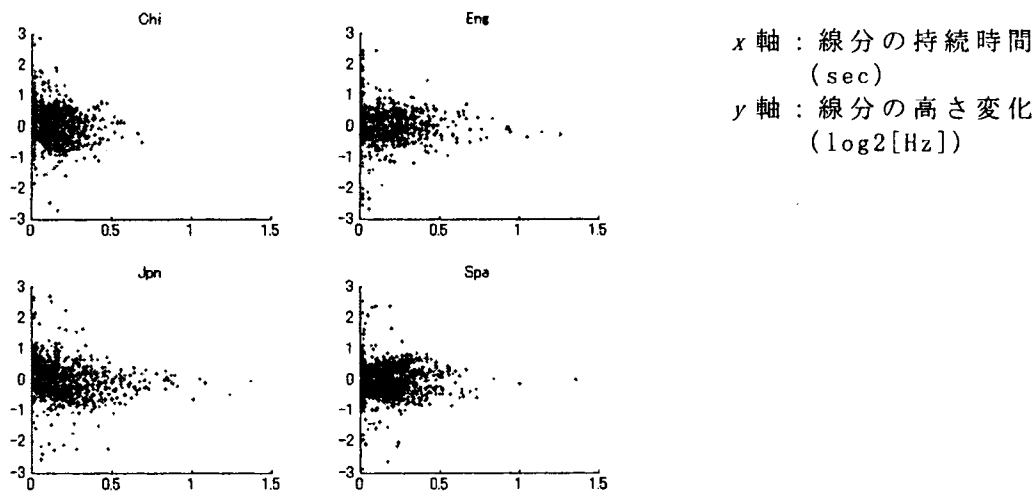


図 4. スタイライズされた高さ線分の分布

3.2 周期成分・非周期成分の瞬時振幅

Ramus et al. (1999)は、リズム類型の判別の指標の一部として、音声の子音区間と母音区間の長さの割合を挙げている。そこで、周期成分と非周期成分の瞬時振幅を 10 ms の分析フレームごとにプロットし、母音性の強いフレームと子音性の強いフレームの分布を調べた(各言語とも、男性 3 名、女性 3 名)。図 5 の各グラフで、右上の方の点は、音声信号の強さが大きいことを示す。左上の方の点は HNR が低く(子音性が強い)、右下の方の点は HNR が高い(母音性が強い)。

図 5 は、英語はもっとも母音区間長の割合が小さく、スペイン語は次に小さく、日本語はもっとも大きいという Ramus et al. (1999)の主張を裏付けている。英語のグラフの逆三角形型の分布は、母音性の強いフレームより子音性の強いフレームが多いことを示している。日本語のグラフは、子音性の強いフレームと母音性の強いフレームの数が均衡した分布を示している。スペイン語のグラフの分布は、英語と日本語の間のような形をしている。中国語の分布は、日本語やスペイン語に似ているようにも見えるが、はっきりしない。

3.3 平均音節

図 6 に、音声信号の強さ曲線、周期成分の振幅曲線、非周期成分の振幅曲線をそれぞれ、各音節の中心で同期させて平均したもの(それぞれ、濃さの異なる

る3種類の線)を示す。各グラフは、3名の話者のデータを示している。(1本の曲線が、1人の話者の平均音節を示す。)

これらの曲線には、普遍的な音節の特徴が見られる。中国語のグラフを例に取ってみる。x軸の20が音節の中心であるが、10~20の範囲の方が、20~30の範囲よりも音声信号の強さが小さく、非周期成分の振幅が大きい。これは、音節の頭部に、より子音性の強い要素が来る傾向があり、尾部には、二重母音の後半や鼻音のようなより母音性の強い要素が来る傾向があることを示している。このような傾向は、中国語以外の言語のグラフにも観察される。

平均音節のグラフには、言語による違いも現れている。日本語では、中心のピークの前と後(約5フレーム目と35フレーム目)にはっきりとしたこぶが見られる。このことは、日本語では音節がかなり等時的に現れていることを示している。このようなこぶは、スペイン語でも、日本語ほどははっきりとは見られない(ここには掲載していないが、スペイン語男性のグラフでは女性よりははっきりと見られる)。

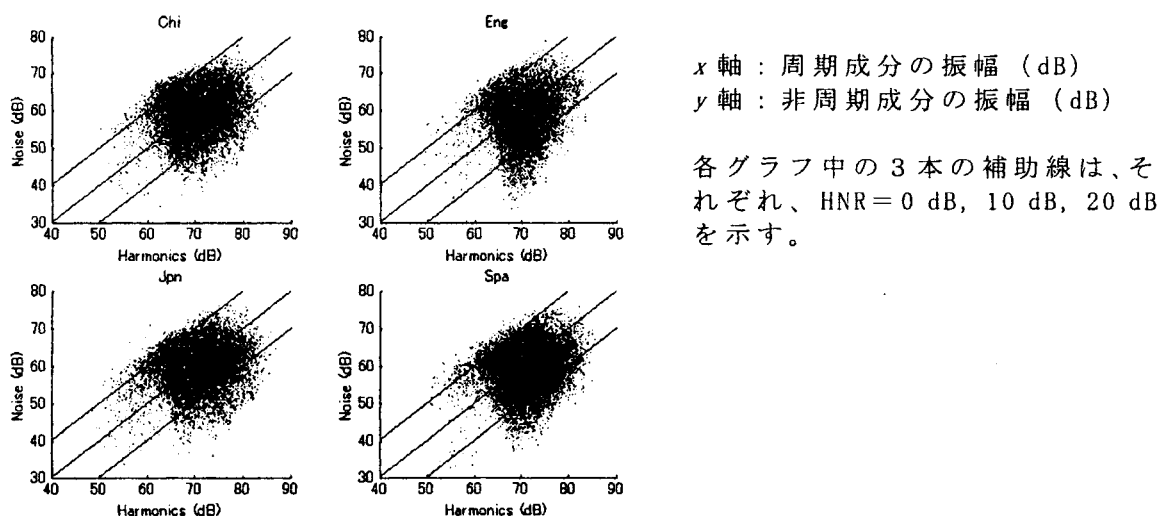


図 5. 周期成分・非周期成分の瞬時振幅

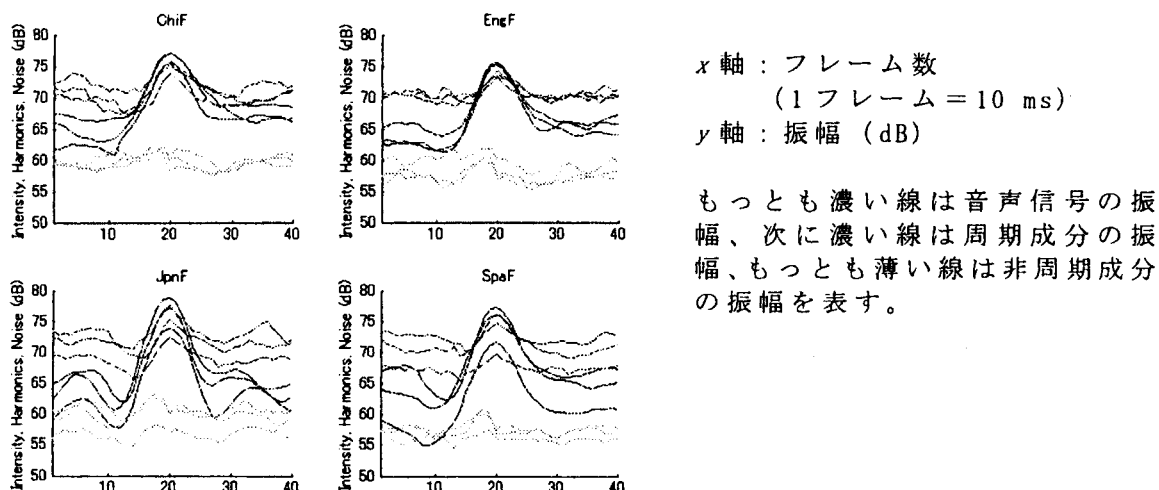


図 6. 平均音節 (女性話者)

4. まとめ

LPC 残差信号による日本語子音の知覚実験では、ソノリティの別に基づく主要音類の判別が 66.4% できたのに対し、個々の子音の同定は 20.0% しかできなかった。また、知覚結果に多次元尺度法を適用してみると、三次元空間上で、子音がソノリティ順位ごとにおおよそ分かれて表された。これらのことから、音源（残差信号）は、ソノリティについての情報を含むが、分節素性的情報はほとんど含んでいないことが分かる。（先行研究の結果などから、ソノリティはリズムにとって重要であると考えられる。）

韻律類型が異なる中国語、英語、日本語、スペイン語の分析では、音源を構成する音響的要素（高さ、強さ、HNR）の局所的な特徴が言語によって異なることが示唆された。

上記の結果により、音源フィルタモデルの音源部が、韻律類型を形作っていると言える。本発表の範囲内では、どのような音響的特徴が韻律類型を表しているのかを明確にするまでには至っていないが、このようなアプローチは妥当であると考えられる。

参考文献

- Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206-244). Cambridge, MA: Basil Blackwell.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In M. E. Beckman & J. Kingston (Eds.), *Papers in laboratory phonology 1* (pp. 283-333). Cambridge: Cambridge University Press.
- Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech*, 25, 29-42.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.
- Komatsu, M. (in press). Essay on acoustic correlates of prosodic typology. In T. Honma, M. Okazaki, T. Tabata, & S. Tanaka (Eds.), *Festschrift for Dr Haraguchi on the occasion of his sixtieth birthday*. Tokyo: Kaitakusha.
- Komatsu, M., & Arai, T. (in press). Acoustic realization of prosodic types: Constructing average syllables. *LACUS Forum*, 29.
- Komatsu, M., Tokuma, S., Tokuma, W., & Arai, T. (2002). Multi-dimensional analysis of sonority: Perception, acoustics, and phonology. *Proceedings of International Conference on Spoken Language Processing 2002*, 2293-2296.
- Komatsu, M., Tokuma, W., Tokuma, S., & Arai, T. (2000). The effect of reduced spectral information on Japanese consonant perception: Comparison between L1 and L2 listeners. *Proceedings of International Conference on Spoken Language Processing 2000*, Vol. 3, pp. 750-753.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

* 本稿は、Komatsu (in press)の一部を改変したものである。本研究は、初期において他の研究者との共同研究として行われてきた。徳間伸一氏、Won Tokuma 氏、その他の共同研究者に感謝します。