

- [20] Fujisaki, H.: Information, prosody, and modeling—with emphasis on tonal features of speech, *Proceedings of Speech Prosody 2004* (2004) pp. 1-10.

## 2.3 パラ言語情報と音声の生成・知覚

### 2.3.1 研究の目的

音声に含まれる情報は、言語情報、パラ言語情報、非言語情報の3種に分類される<sup>[1]</sup>。このうちパラ言語情報は、話し手の意図や態度のように話者が意図的に表出する情報でありながら文字には転写されない情報である<sup>[2]</sup>。本研究では、パラ言語情報の伝達メカニズムを生成と知覚の両面から検討した<sup>[3-6]</sup>。

### 2.3.2 パラ言語情報の生成

#### a. 音響的分析

##### (1) 音響分析用資料

キャリア文「～ですか」の文頭に「そ<sub>1</sub>う」「や<sub>1</sub>まの(山野)さん」「あな<sub>1</sub>た」「え<sub>1</sub>(絵)」「え<sub>1</sub>ー(A)」(記号<sub>1</sub>はアクセント核)など15語を配した発話を、標準語話者3名に種々のパラ言語的意味を意図して繰り返し(最低10回以上)発話させた資料を分析した。パラ言語的意味としては、文献<sup>[7]</sup>を参考として「N 中立」「F 強調」「I 無関心」「A 感心」「D 落胆」「S 疑い」の6種類を採用した。話者3名は、男性2名(ST, YS), 女性1名(JH)である。

##### (2) 持続時間の特徴

まず、発話全体の持続時間を検討する。図2.8は、「山野さんですか」を構成する各モーラの持続時間平均値がパラ言語的情報によって変動する様子を、「N」の平均を1.0として表示している(右端は全体の持続時間の変動)。発話全体としても各モーラにおいても「N, F, I」は短く、「A, D, S」は長いことがわかるが、同時に発話冒頭と発話末尾のモーラにおいては変動がことに著しい。

図2.8では、「A, D, S」の平均持続時間長が「N」の2倍を超えている。このような変動下において、日本語の重要な音韻特徴である分節音の長短の対立が保存されるかどうかは興味深い問題である。「絵ですか」と「Aですか」

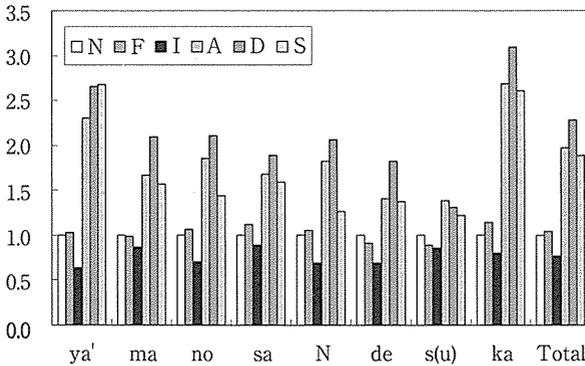


図 2.8 バラ言語情報による発話持続時間の変動 (話者 ST)

の発話冒頭に位置する短母音/e/と長母音/eH/の持続時間を検討したところ、「A, N, S」の変動下においても長短の差が明瞭に保存されていることが判明した。

### (3) 音声基本周波数 ( $F_0$ ) の特徴

図 2.9 に、発話「そうですか」における  $F_0$  曲線の典型例を示す (第 4 モーラ「す」では母音が無声化するため、 $F_0$  が観察されないことに注意)。図中の縦線はスペクトログラムから決定した音声境界、矢印は「そっう」のアクセント核に起因する  $F_0$  の下降開始時刻を示している。「N, F, I」の形状はこの発話に対して音韻論が予測する  $F_0$  形状に合致しているが、「A, D, S」ではいくつかの点において逸脱が生じている。

### (4) 句頭の上昇

図 2.9 の発話は重音節で始まるため、通常は句頭における  $F_0$  上昇が弱化するか消失する。実際に、図 2.9 のうち「N, F, I」はそうなっているが、「A」と「S」では明瞭な上昇とそれに先だつ低ピッチ区間が観察される。

### (5) 句末の上昇

図 2.9 では、「N, F, S」の句末に上昇イントネーションが生じているが、その上昇形態は「N, F」と「S」とで質的に異なっている。前者が単純な上昇調であるのに対し、後者ではまず低ピッチ区間が持続した後に上昇が始まる。また、上昇の幅も大きい。これは川上が指摘した 4 種の上昇調のうち、「反問の上昇」に該当する<sup>[8]</sup>。

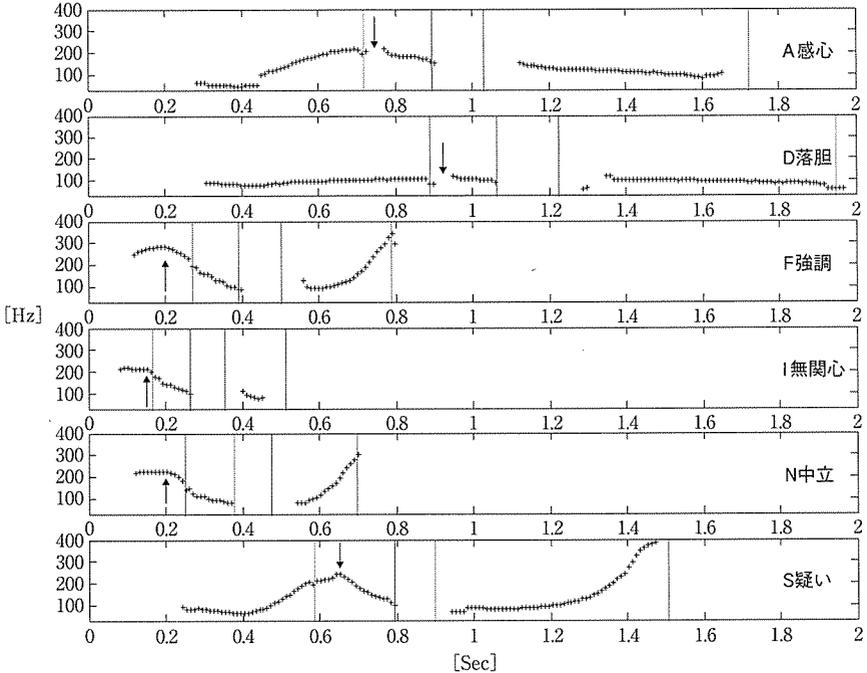


図 2.9 「そうですか」の典型的  $F_0$  形状 (話者 ST)

(6) アクセントによる下降のタイミング

最後に、アクセント核による  $F_0$  下降のタイミング (図 2.9 の矢印参照) にもパラ言語情報の影響が認められる。資料全体を通して、「A, S」の下降タイミングには「N, F, I」に比べて有意に遅れている<sup>[5]</sup>。

(7) フォルマント周波数

パラ言語情報の影響はいわゆる分節的特徴にも及ぶ。図 2.10 では「あなたですか」に含まれる 4 個の/a/について第 1 第 2 フォルマント周波数 ( $F_1$ ,  $F_2$ ) の散布図を示した。「S」の  $F_2$  は組織的に「A」よりも上昇している。

b. 調音運動の分析

前節で触れたフォルマント周波数の差が音響分析の誤差ではなく、実際に調音運動の差によって生み出されていることを確認するため、EMA 装置による

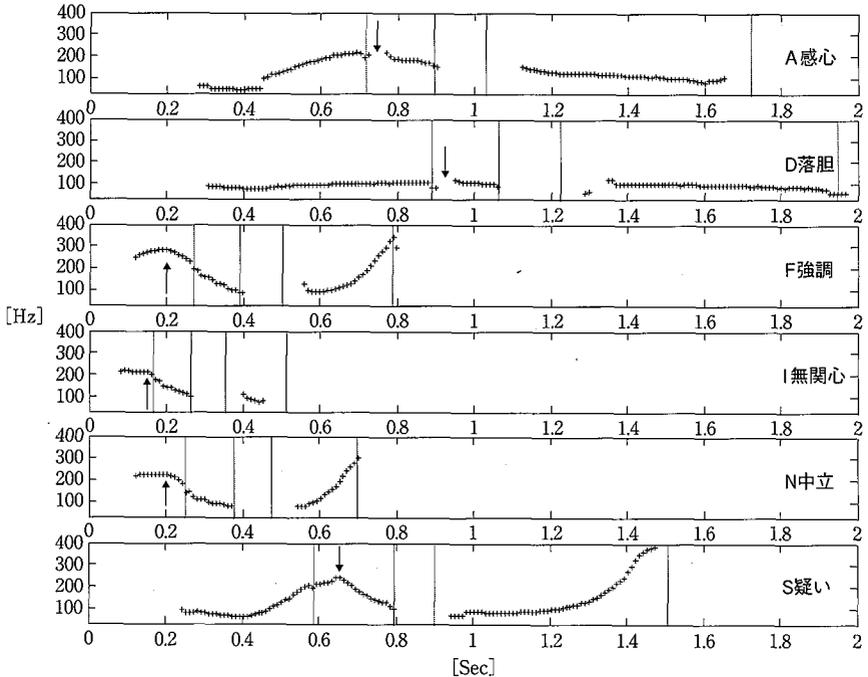


図 2.9 「そうですか」の典型的  $F_0$  形状 (話者 ST)

(6) アクセントによる下降のタイミング

最後に、アクセント核による  $F_0$  下降のタイミング (図 2.9 の矢印参照) にもパラ言語情報の影響が認められる。資料全体を通して、「A, S」の下降タイミングには「N, F, I」に比べて有意に遅れている<sup>6)</sup>。

(7) フォルマント周波数

パラ言語情報の影響はいわゆる分節的特徴にも及ぶ。図 2.10 では「あなたですか」に含まれる 4 個の/a/について第 1 第 2 フォルマント周波数 ( $F_1$ ,  $F_2$ ) の散布図を示した。「S」の  $F_2$  は組織的に「A」よりも上昇している。

b. 調音運動の分析

前節で触れたフォルマント周波数の差が音響分析の誤差ではなく、実際に調音運動の差によって生み出されていることを確認するため、EMA 装置による

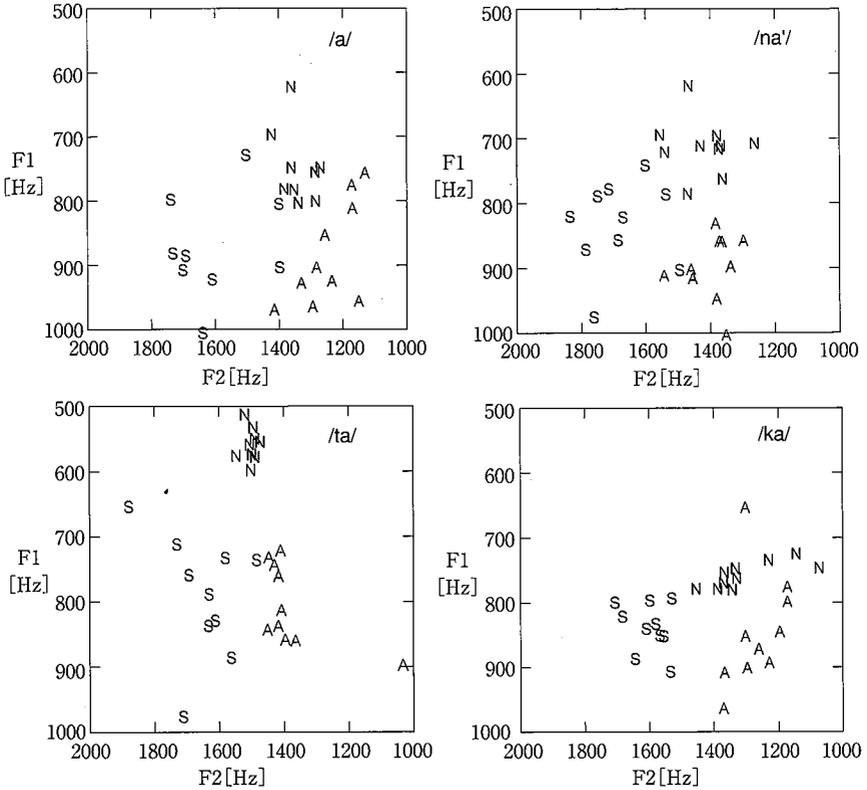


図 2.10 「あなたですか」に含まれるの/a/の F1~F2 平面での分布 (話者 ST)

調音運動の測定を実施した<sup>[3]</sup>. その結果, 「S」のサンプル群では, 舌全体が明らかに「A」よりも前寄りに位置していることが判明した. 図 2.11 は「笹田が」に含まれるすべての分節音に対して, 舌の前後 (水平方向) 位置の平均値と標準偏差を示したものである. 母音だけでなく, すべての子音 (/s, d, g/) も「S」の場合に前寄りに調音されていることがわかる.

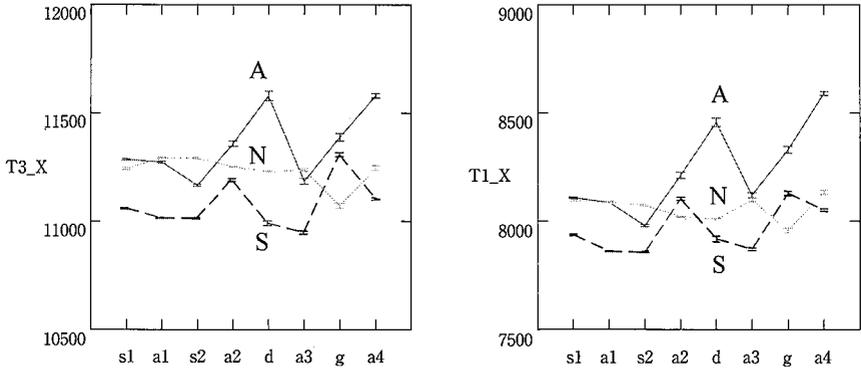


図 2.11 舌の水平方向の平均値と標準偏差。横軸は音韻。縦軸は舌背で代表させた舌の前後位置を表しており、値が小さいほど前寄り（唇より）である。測定の単位は  $10^{-2}$ mm。

### 2.3.3 パラ言語情報の知覚

#### a. パラ言語情報の知覚空間

多次元尺度法 (MDS) を用いて、パラ言語情報の知覚空間を構築することを試みた<sup>[5]</sup>。MDS の入力となる類似性行列は、任意の 2 刺激間が同一のパラ言語情報として同定された確率と定義し、先に紹介した同定実験データから計算した。この計算を 2.3.2 a. (1) で説明したサンプル (436 個) に対して実施し、436 行 × 436 列の類似性行列を得た。

統計パッケージ SAS の MDS プロシージャを用いて分析した結果は、3 次元解で STRESS=0.04 となった。図 2.12 に刺激の散布状態を示す。この図では個々のサンプルの分布は省略して、各被験者 (3 名) についてパラ言語情報ごとに 68% 確率楕円を表示している。図中のシンボルは、話者 ST による「N」の分布を “st-N” のように表示している。

図 2.12 の第 1 次元 (DIM 1) は、「A, D, S」と「N, F, I」を分離している。「A, D, S」は、持続時間が長く振幅が大きく、 $F_0$  と発声様式にも特徴が認められるなど、音声学的に顕著な性格を有するサンプルであることから、DIM 1 はパラ言語情報の「顕著さ (salience)」の軸と解釈できる。第 2 次元 (DIM 2) は、「S」と「A, I」を分離している。「S 疑問」が聞き手の反応

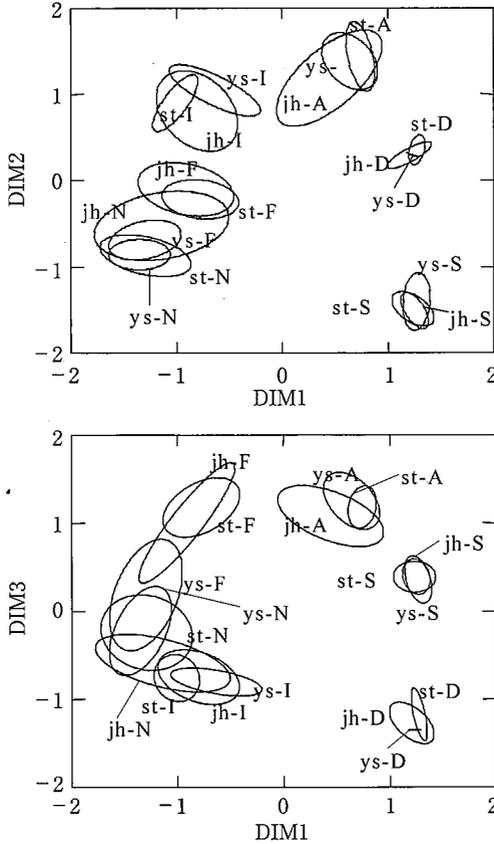


図 2.12 MDS により構成されたパラ言語情報の知覚空間

(回答)を要求する発話である一方、「A」や「I」が反応を要求しない発話であることから、DIM 2は「問い合わせ (query)」の軸であると解釈できる。第3次元 (DIM 3) は、「大きさ (loudness)」の軸と解釈できる。聞こえの大きなサンプル「A, F, S」と聞こえの小さなサンプル「N, I, D」を分離しているからである。

**b. 音響特徴との関係**

MDS 空間における布置とサンプルの音声学的特徴との関係を検討するた

め、各サンプルの DIM 1~3 上での座標値について、音響分析で得られた種々の特徴量との回帰分析を実施した。その結果、DIM 1 は発話全体の持続時間と、DIM 2 は発話全体のピッチレンジと、そして DIM 3 は発話全体の平均 (RMS) 振幅と強く相関していることがわかり各軸の解釈と一致した。また、発話全体の平均  $F_0$ 、ピッチレンジ、平均振幅、平均持続時間、第 2 フォルマント周波数のすべてに統計的な有意差が認められた。

### c. 非母語話者による知覚

次に、前節と同じ実験を日本語を母語としない被験者群に実施することによ

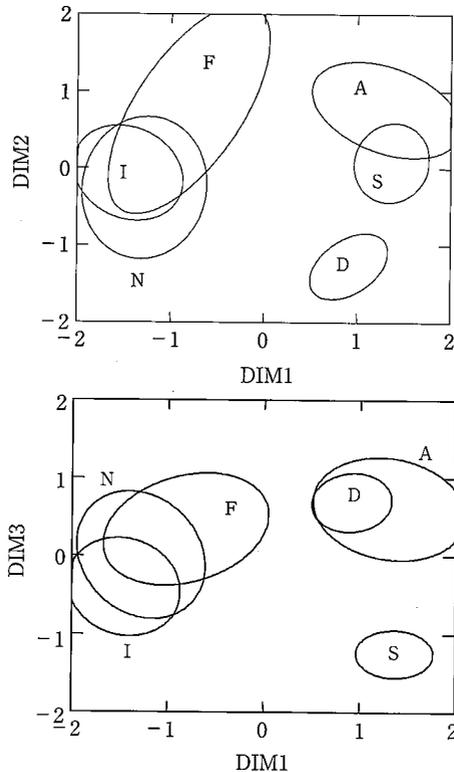


図 2.13 MDS により構成されたパラ言語情報の知覚空間 (米語を母語とする日本語学習者群)

って、パラ言語情報の言語依存性を検討した<sup>[11]</sup>。被験者としては、日本語の学習経験があり、実験時に日本に滞在していた米語母語話者 11 名と、日本語学習経験が皆無の米語話者 15 名の 2 群を用いた。前者を学習者群、後者を非学習者群とよぶ。知覚実験の方法は、日本語母語話者の場合とまったく同一である。

MDS における STRESS 値（3 次元解）は、学習者群で 0.07、非学習者群で 0.16 となった。図 2.13 に学習者群、図 2.14 に非学習者群における布置を示す。図 2.12 に表れたような発話者の明瞭な分離が認められなかったため、図 2.13, 2.14 の確率楕円ではパラ言語情報のタイプだけを表示している。

学習者群の知覚空間は、母語話者群の空間に類似している。DIM 1 は「顕著さ」の軸と解釈できる。また、DIM 2 と DIM 3 はそれぞれ母語話者群の DIM 3 と DIM 2 に対応しているとみなせる。

非学習者群の DIM 3 では、すべてのパラ言語情報の分布が大幅に重複しており、無意味な軸となっている（そのため、図 2.14 では DIM 3 を表示していない）。しかし非学習者群の DIM 1 は、母語話者群と同様（したがって、学習者群とも同じく）「A, D, S」と「N, F, I」を分離しており、やはり「顕著さ」の軸と解釈できる。DIM 2 は「D」とそれ以外のパラ言語情報とを分離す

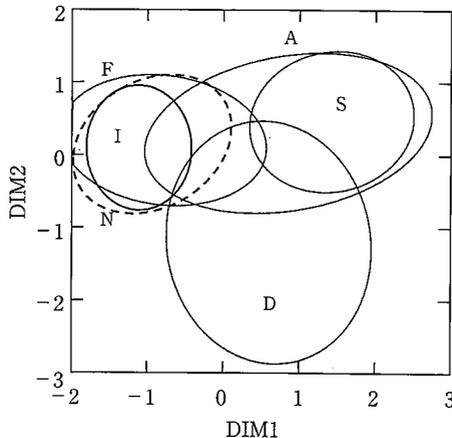


図 2.14 MDS により構成されたパラ言語情報の知覚空間  
(米語を母語とする日本語非学習者群)

る軸となっているが、この構造は実は母語話者群の DIM 3 および学習者群の DIM 2 が果たしている機能である。図 2.12 の DIM 1 と DIM 3 のパネルないし図 2.13 の DIM 1 と DIM 2 のパネルを時計回りに 45 度ほど回転させると、図 2.14 に類似した分布が得られる。非学習者は、母語話者および学習者とは少し異なった形で「大きさ」の軸を保有していると考えられる。

それでは、母語話者ないし学習者における「問い合わせ」の軸はどうかという、図 2.14 からこれに該当する軸を読み取ることができない。実際、図 2.14 では、DIM 1 においても DIM 2 においても「S」と「A」とがほぼ完全に重複してしまっている。この点が、非学習者群の知覚空間の大きな特徴である。

### 2.3.4 議 論

#### a. パラ言語情報の言語依存性

最初に 2.3.3 節の実験結果について議論する。非学習者群の知覚空間には、母語話者群の DIM 2 (学習者群の DIM 3) である「問い合わせ」の軸が欠落していた。母語話者にとって、この軸を認知するための重要な手がかりとなるのは発話末尾のイントネーション特徴である。

非学習者群が、日本語における 2 種類の上昇イントネーションを聞き分けることができない (あるいは聞き分けることができても、その機能を正しく推定することができない) ことは、パラ言語情報に一定範囲で言語依存性が認められることを示唆している。この発見には 2 つの意義がある。第一に、言語情報が完全に言語依存性の情報であること、また非言語情報である感情の知覚にはかなり高度な言語普遍性が認められること<sup>[13]</sup>と比較すると、パラ言語情報が両者の中間的性格をもつことが示唆される。第二に、この発見は今回のサンプルに含まれていた 2 種類の上昇調イントネーションの間に明瞭な記号的差異が存在していることの強い証拠と解釈することができる。

#### b. 音声生成におけるパラ言語情報の位置

本研究の成果は、パラ言語情報に関する発話のプランニングに 2 つの異なる経路が存在することを示唆している。すなわち、言語情報を参照するプランニングと参照なしのプランニングである。母音長の音韻対立を保存した発話持続

時間長の制御や2種類の上昇調イントネーションの選択は、前者に属する。また、アクセント核による  $F_0$  下降タイミングの制御も、それが知覚されるアクセント位置の変更を伴わない範囲で制御されているという点で、前者に属する制御であると考えられる。これら以外のすべての制御（発話全体の持続時間、発話全体の  $F_0$  レンジ、発話全体の振幅など）が後者に属する。これらは声質 (Voice-quality) と総称することができる<sup>[11]</sup>。

### 2.3.5 結 論

音響学、生理学、心理学、言語学の観点からパラ言語情報の伝達メカニズムを検討した。パラ言語情報の科学的研究は、本研究その他によって現在その端緒についたばかりであり、今後検討すべき多くの課題が残されている<sup>[12]</sup>。また本稿では紙幅の関係で大まかな解説しか行えなかった。より詳しくは参考文献に当たっていただきたい。

#### [参考文献]

- [ 1 ] H. Fujisaki, Prosody, Models, and Spontaneous Speech, In Y. Sagisaka, et al., (Eds.): *Computing Prosody, Computational Models for Processing Spontaneous Speech*, New York: Springer (1997) pp. 27-42.
- [ 2 ] 前川喜久雄: パラ言語的情報, 別冊国文学, 現代日本語必携, No. 53 (2000) pp. 172-175.
- [ 3 ] K. Maekawa and T. Kagomiya: Influence of paralinguistic information on segmental articulation, *Proc. 5th ICSLP 2000*, Beijing, Vol. 2 (2000) pp. 349-352.
- [ 4 ] 前川喜久雄, 北川智利: パラ言語情報の知覚, 日本行動計量学会第 28 回大会論文集 (2000) pp. 157-160.
- [ 5 ] 前川喜久雄, 北川智利: 音声はパラ言語情報をいかに伝えるか, *認知科学*, 9-1 (2002) pp. 46-66.
- [ 6 ] M. Fujimoto and K. Maekawa: Variation of phonation types due to paralinguistic information: An analysis of high-speed video images, *Proc. ICPHS 2004*, Vol. 4 (2004) pp. 2401-2404.
- [ 7 ] O. Mizutani and N. Mizutani: *Aural Comprehension Practice in Japanese*. The Japan Times (1979).
- [ 8 ] 川上 稔: 文末などの上昇調について, *国語研究*, No. 16 (1963) pp. 25-46.
- [ 9 ] 前川喜久雄: 声質とパラ言語情報, 韻律に着目した音声言語情報処理の高度化 (科研費特定研究 2003 年度研究成果報告書) (2004) pp. 15-18.
- [10] K. Maekawa, H. Kikuchi, Y. Igarashi and J. Venditti: X-JToBI, An extended J\_ToBI

- for spontaneous speech, *Proc. ICSLP 2002*, Denver, 3 (2002) pp. 1545-1548.
- [11] K. Maekawa: Production and Perception of "Paralinguistic" Information, *Proc. Speech Prosody 2004*, Nara (2004) pp. 367-374.
- [12] 前川喜久雄: パラ言語情報研究の課題, 音響学会秋季研究発表会講演論文集 (2002) pp. 247-250.
- [13] Scherer, K.R.: Cross-cultural investigation of emotion inferences from voice and speech, Implications for speech technology, *Proc. 5th ICSLP 2000*, Beijing, 2 (2000) pp. 379-382.

## 2.4 日本語のリズムと時間制御

### 2.4.1 2種類のリズム

リズムという言葉は、「流れる」という意味のギリシャ語に由来する。「流れる」とは「よどみなく流れる」ということである。音楽やスポーツの世界でも言葉の世界でも、リズムとは一定の構造が繰り返し起こること、ないしはそのような繰り返しから生じる心地よさを意味する。

リズムは人間の世界にだけ見られるわけではない。自然界の出来事の多くはリズムをもっている。地球が自転することによって日（朝—昼—夜）という単位が、月の満ち欠けによって1ヵ月という単位がそれぞれつくり出される。さらに、地球の公転によって1年（春夏秋冬）という単位がつくり出される。周期性をもったこのような単位はすべてリズムの単位であって、定期的に繰り返し起こることによって時間軸上に自然のリズムをつくり出している。

言葉のリズムも、基本的にこの自然界のリズムと同じである。一定の言語構造や特徴が周期的に起こることによって、言葉の中に快適さをつくり出すのである。言語のリズムは、何が繰り返し起こるかによって、音節拍リズム (syllable-timed rhythm) と強勢拍リズム (stress-timed rhythm) の2種類に大別され、自然言語はこのいずれかの構造を選択するものと仮定されている<sup>[1,2]</sup>。音節拍リズムは、各音節が同じ長さで繰り返されることによってつくられるリズムであり、一方、強勢拍リズムは文レベルの強勢形式（文強勢）が等時的に繰り返されることによってつくられる。フランス語は前者の、英語やドイツ語は後者の典型的な例とされている。この2つのリズム構造を図示すると、次のようになる（丸は音節を表し、丸の大きさは強勢の度合いを表す）。