

# 韻律コーパスの作成とその自動化<sup>#</sup>

## Development of Prosodic Corpora and Support Tools

静岡大学情報学部

Faculty of Information, Shizuoka University

北澤 茂良

Shigeyoshi Kitazawa

The aim of the research group is to establish an advanced scheme of developing prosodic corpora for speech synthesis as well as speech recognition and other basic researches. Scope of the study is not only limited to realizing speech processing applications, but also to para- and non- linguistic information in real-life speech. Realization of multi-modal speech corpus is also aimed at. Although both of heuristic manual and computational automatic frameworks are introduced for the research works, major focuses are placed on development of several corpora in the framework of J-ToBI coding incorporating statistical and syntactic methods. Roles of non-lexical utterances in dialogues are investigated through analysis of their prosodic features. The research also covers the control of prosodic features in the automatic annotation of prosodic contents, which includes linguistic and non-linguistic modalities, and in multi-modal recording and annotation. Through four years' of research work, a number of outstanding results have been accomplished; Japanese MULTEXT as a parallel corpus, a prosodically annotated speech database of ASJ, PASD, and MAP-task dialogue speech, several multi-modally recorded and annotated free conversations, development of tools to support phonemic segmentation and J-ToBI annotation, generation of F0 contours for Fujisaki model fitting, and so on.

Keywords: Prosodic Corpus, Prosody, F0 Model, MULTEXT, Speech Database, Spontaneous Speech, J-ToBI, Multi-modal, Emotional Speech

### 1. 研究の目的

音声をデータベースとして蓄積する場合、その特徴を記述するラベル付けを行うことによって、データベースの利便性が大きく向上する。文節については、国際音声記号をはじめ確立した表記法があり、それをもとに音素ラベル付けが行われ、音素HMMの学習等に利用されている。しかしながら、韻律についてはそのような表記法がなく、従って、韻律ラベル付けされたコーパスの開発も音素の場合と比較して遅れている。韻律の研究を行う上で、韻律ラベリングされた音声コーパスの整備は不可欠である。

韻律ラベリングに際し、まず問題となるのが、その

表記法である。英語の韻律を記述する指標として単語境界情報、抑揚情報等を表記する Tones and Break Indices (ToBI) システムが開発されているが、定量的な記述に問題がある。これに対し、ピッチアクセント言語である日本語の韻律構造に即した韻律表記方法を開発する。すなわち、まず、生成過程のモデルに基づいた韻律構造のレベルとして、韻律文：韻律節：韻律句：韻律語：を定義し、次に、それに基づいて朗読調音声のみならず会話音声・対話音声の韻律的特徴を適切に表現する韻律表記方式を開発して、韻律ラベリングされた音声コーパスを作成する。

音声コーパスは多量に及ぶため、そのラベリングの自動化が重要な課題となる。一般に、音素情報に関し

<sup>#</sup> 研究課題：韻律コーパスの作成とその自動化、研究課題番号：12132204

ては、発声内容が既知の条件下で認識の場合と似た操作をすることでラベリングの自動化が図れるが、韻律自動ラベリングについては精度の高い手法を新しく開発することが求められる。既に、本領域研究のメンバーにおいて、HMM を用いた強制切り出しによって得た音素境界情報と韻律的特徴とを併用したラベリング情報の推定、基本周波数パターンの時間微係数とフィルタリングを利用した生成過程モデルのパラメータ値の推定、言語情報を利用したラベリングの推定精度を向上、等の手法がいくつか提案されているが、これらを基礎として有効な韻律自動ラベリング手法を開発する。

韻律的特徴は言語情報のみと関連するものではない。この観点から、Spontaneous Speech においては、朗読音声の場合と比較して、さらに韻律の役割が重要となる。そこで、方言や感性の表出などのパラ言語・非言語情報（音響情報のみならず身振りなどの映像情報も含めて）に関する記述も韻律コーパスに含める。これに伴い、Spontaneous Speech として会話文、あるいは講演文 (Monologue) をもデータベース化する。各班の協力のもとに実際に韻律ラベリングされた朗読音声・対話音声コーパスを作成し、各班共用の研究資産とする。

## 2. 研究の体制と成果

Table 1. Members and their research topics.

| Name                | Affiliation         | Major research topic   |
|---------------------|---------------------|--|
| Shigeyoshi Kitazawa | Shizuoka University | Development of a prosodic corpus                                 |
| Shinya Kiriya       | Shizuoka University | Automatic annotation of prosodic corpus                          |
| Toshihiko Itoh      | Shizuoka University | Annotation of prosodic corpus                                    |
| Shuichi Itahashi    | Tsukuba University  | Development of prosodic corpus based on existing speech database |
| Akira Ichikawa      | Chiba University    | Development of a multimodal prosodic corpus                      |
| Nick Campbell       | ATR                 | Design of prosodic annotation                                    |

Table 2. Group meetings.

| Fiscal Year | Date        | Place               |
|-------------|-------------|---------------------|
| 2000        | October 26  | University of Tokyo |
| 2001        | January 19  | Chiba University    |
|             | July 5      | Doshisya University |
| 2002        | November 15 | ATR                 |
| 2003        | July 12     | Hamamatsu ACT Tower |

表1のように、本研究班は代表者北澤と分担者桐山（北村）伊藤、板橋、市川、ニック・キャンベルで構成され、研究目的を達成するため、主として北澤らは韻律コーパスの韻律ラベリングの観点から、桐山は韻律パラメータの自動抽出の観点から、伊藤は韻律コーパスの作成支援の観点から、板橋は既存音声データベースへの韻律情報の抽出とラベル付与の観点から、市川は対話音声の韻律情報の抽出とラベル付与の観点から、ニック・キャンベルは対話音声の非言語情報の記述方法と韻律記述方法の考案・評価の観点から、表2のように年2回程程度のペースで班会合を行う他、電子メールを活用して、それぞれ緊密に連絡を取りながら研究を進めた。

韻律コーパス作業グループでは、本特定領域内の韻律コーパスに関心を有する有志の会合を数回開催し、自由討議、報告、アンケートの実施、シンポジウムの開催などを通じて、韻律コーパスの仕様を確定していった。

以下に、各課題と成果の概略を述べる。

### 2.1 研究課題

韻律コーパスを構築し、本特定研究内で評価を受け改善することを本研究班の課題とする。そのために、韻律表記方式の検討、大量データ処理のために自動ラベリングの方式を開発する。

**韻律表記方式を開発：** ピッチアクセント言語である日本語の韻律構造に即した韻律表記方法を開発する。すなわち、生成過程のモデルに基づいた韻律構造のレベル（韻律文：韻律節：韻律句：韻律語：）を定義して、これらの韻律表記に基づいて朗読調音声のみならず会話音声・対話音声の韻律的特徴を適切に表現する韻律表記方式を開発する。

**ラベリングの自動化：** 韻律コーパスは多量に及ぶため、そのラベリングの自動化が必要となる。コーパス構築のためのソフトウェアを整備し、韻律コーパス作成の自動化を図る。

**パラ言語・非言語情報：** 韻律的特徴は言語情報のみと関連するものではない。韻律の役割が発揮されるのはいわゆる Spontaneous Speech である。音声の背景特性、発声的要素などのパラ言語情報（音響情報のみならず身振りなどの映像情報も含めて）に関する記述も韻律コーパスに含める。

**韻律コーパスの統合：** 国内外の研究動向韻律表記方法の調査比較と韻律コーパスの統合を図る（キャンベ

ル)

## 2.2 成果

### 新規の韻律の収録（静岡大学）

韻律コーパスの先行例としてEUのMULTEXT 韻律データベースの調査を行った。フランスのデータベース機構ELRAとMULTEXT プロジェクト実施機関のAix プロヴァンス大の音声科学研究所を訪問し意見を交換した。

MULTEXT の40パッセージを日本語に翻訳する過程を通じて様々な状況設定下の韻律を考慮した。EUROM1の音声収録仕様に準拠して、無響室で男女各3名のプロナレーターによる朗読音声に加えて自発発話音声を収録した。同時に、声帯振動の物理的観測値としてEGG 波形を収録した。

MULTEXT で開発されたMOMEL アルゴリズムによる様式化されたイントネーションパターンの抽出アルゴリズムを導入し、日本語に適用したときの問題点を明らかにした(Nick)。さらにMOMEL アルゴリズムで抽出したイントネーションパターンからTEMAX 法による韻律リズムパターンを言語間で対照比較し、朗読と自発性発話と比較観察し、日本語固有のリズムがあることを明らかにした。

EGG およびPraat によって基本周波数を正確に求めて、さらに『韻律句』・アクセント核のラベリング作業を研究支援者を雇用して行い、これらを添付した試作版の『韻律コーパス』をプロジェクトの内外に広く配布し評価を求めた。[1]

韻律コーパスとして日本語のMULTEXT 韻律データベースの40パッセージに、J-ToBI 韻律タグ付けを完了した。J-ToBI タグの多くは書起しテキストから生成可能であった。しかし、『韻律句』の結合度の指標BI=2,3の判断、強調されたアクセント句の判定、F0のピーク値と基準値については熟練者による聴取判断が不可欠であることがわかった。これらのラベリング作業について研究支援者を雇用して行った。対話音声の語順・文脈の相違が強調の判断に与える影響についてと、接続境界における音響的特徴の詳細について国際会議ICSLP2002にて発表した。

日本語のMULTEXT 韻律データベースのJ-ToBI 韻律タグ付けと同様の手法で、筑波大学と千葉大学と東京大学と東工大グループの既存音声コーパスの各種案内読上げと模擬対話と対話音声、マルチモーダル対話音声、天気予報、模擬感情音声へのJ-ToBI タグ付けを行った。これらのラベリング作業についても研究支援者を雇用

して行った。言語情報を利用した韻律ラベリング手法の開発と、音素ラベリング支援のための音素自動セグメンテーションと、接続境界における音響的特徴の詳細について研究成果を発表した。[2,3]

### 対話音声の韻律情報の統計的集計

文献検索に関する音声対話の韻律情報を統計的に集計した。基本周波数の分析結果から、間投詞を含む発話の正規化基本周波数の値は、負の値をとりやすい傾向を示し、短時間平均音声パワーは、間投詞を含まない発話よりも小さな値になりやすい傾向を示した。（板橋）

**既存の音声コーパスのラベリングと韻律記述モデル**  
文献検索をタスクとした音声対話データの基本周波数分析を行い、韻律分析および基本周波数パターンの記述モデルを検討した。このモデルの有効性を確認するために日本音響学会研究用連続音声データベースの「各種案内タスクコーパス」(1,000文、話者 36名、計12,500文)の韻律分析を大量に行い、モデルの精度を高める。「各種案内タスクコーパス」の基本周波数抽出と韻律ラベリングを外注で行った。

日本音響学会研究用連続音声データベースの「ASJ 案内タスクコーパス」と「模擬対話コーパス」(37対話、話者 37名、計 3,130文)の韻律分析を大量に行い、F0の開始値、最終値、最大値、最小値、最大値・最小値・終了値の位置の7パラメータから発話単位のクラスタリングを行ったところ、読上げと模擬対話の違いが明らかになった。また韻律モデルパラメータの自動抽出の研究を進めた。[4]

日本音響学会研究用連続音声データベースの各種案内読上げ文と模擬対話、重点領域研究「音声対話」の対話音声コーパス(92対話、話者 82名、計 1.5GB)の3種のコーパスに基本周波数分析と発話ラベルと付与したのに関して、200ms 以上の無音区間で区切られた音声区間を発話単位として、発話単位長を読上げ音声と模擬対話音声で比較した。模擬対話では間投詞や割込みによって発話単位が短くなる。音声パワーと基本周波数の標準偏差は対話に比べて読上げは狭い範囲に集中していることが分かった。（板橋）

### 音素セグメンテーションの自動化

音声認識ツールによる音素セグメンテーションの自動化に関する検討した。韻律コーパスを作成する上で音素ラベルを付すことが必要になり、そのために自動ラベリングの手法の検討を行った。音声認識手法としてはJuliusとHTKを用い、書起こしテキストから音素ラベルを付す際に、無声化やショートポーズを考慮した

処理を施すことにより、自動認識の推定精度が大幅に向上した。(市川)

### ジェスチャー・顔表情付の対話音声収録

音声対話におけるジェスチャーを記録・分析するためには、正面からの映像が不可欠である。しかし、話者も対話中は相手の顔を正面から見るができなければ、不自然な会話となってしまう。そこで、設備として導入した二台のプロンプター（ハーフミラーを用いて、話者の正面にテレビ画像を表示しながら、同時に正面からの顔画像が録画できる）を用いて音声対話収録を行ない、音声・ジェスチャーの予備的収録、タグ付け作業を行なった。

二台のプロンプター（映像提示装置）を介したマルチモーダルな自然対話の音声対話収録を行い、音声・ジェスチャーの収録、タグ付け作業を行った。視線と話者交替の関係は、発話終了時・話者交替時に相手を見る傾向がある。話者交替時にうなずく頻度は高い。「視線をそらす」ジェスチャーが話者継続の意思表示に用いられている可能性が示唆された。話者継続時における発話単位終了時での話し手の頷きによって聞き手が相槌や頷きを9割近く返していること、話し手が頷かない場合でも言語情報や音声の韻律情報によって聞き手は反応を3割程度返していることを明らかにした。聞き手のときは「あいづち」と同様、うなずくことが多い（予想された結果）、話し手のときにもうなずくことが多い（予想されなかった結果）、話し手のうなずきを詳細に分析したところ、発話の最後にうなずくことが多く、聞き手の「あいづち」や「うなずき」をうながしていることが示唆された。しかし、話し手のうなずきについては、まだまだ未解明の部分が多く、今後、さまざまな研究の発展が期待できる。(市川) [5]

### 3. 今後の発展

本研究では、男女各3名による MULTEXT 日本語版の朗読音声・模擬自然発話音声（EGG 信号同時録音）を収集し、これに音素ラベル、音節ラベル、韻律構造ラベル（韻律文：韻律節：韻律句：韻律語：）を付与した朗読・模擬自然発話音声の韻律コーパスの開発、文献検索をタスクとした既存の音声対話コーパスのラベリングと韻律記述モデルの基本周波数分析を行い、韻律分析および基本周波数パターンの記述モデルを作成、二台のプロンプターを用いてマルチモーダルな自然対話の音声対話収録を行い、音声・ジェスチャーの収録、タグ付け作業、等の成果を達成した。これらは、学会誌論文、国際会議論文等として発表する他、韻律コー

パスとして公開している。

特定領域研究「韻律と音声処理」平成12年度～15年度において「韻律コーパス班」として日本語韻律コーパスの作成に着手し、試作データを評価用として配布したところ好評を得ている。このため、内外の多くの研究者から、本データベースの拡張と、更なる付加情報の充実が強く要望されているところである。韻律ラベリングには世界的な標準とされる ToBI (Tone Break Index) の日本語版 J-ToBI によってラベル付けを行った。研究の主眼はデータベースのもととなる、正確な言語・韻律情報の記述したラベリングデータである。特定領域研究「韻律と音声処理」で試作した韻律コーパスの内、既に昨年度配布した「日本語 MULTEXT 韻律コーパス」はその改訂版を作成する。加えて未公開の「日本音響学会の案内・模擬対話タスク音声」、「ジェスチャー付の対話音声(話者の映像付き)」、「地図課題対話音声」について音素ラベル、F0抽出、韻律ラベル付けを整備した上で CD-ROM を作成して配布する。(平成16年度科学研究費補助金(研究成果公開促進費)「特定領域研究日本語韻律コーパス」(レコード数25000件、容量120MB))

### 参考文献

- [1] Kitazawa Shigeyoshi, Kitamura Tatsuya, Mochiduki Kazuya, and Itoh Toshihiko "Preliminary Study of Japanese MULTEXT: a Prosodic Corpus" International Conference on Speech Processing, Taejon, Korea pp.825-828 (2001.8)
- [2] Shinya Kiriyama, Yoshifumi Mitsuta, Yuta Hosokawa, Yoshikazu Hashimoto, Toshihiko Ito, and Shigeyoshi Kitazawa, "Japanese Prosodic Labeling Support System Utilizing Linguistic Information," Proceedings of the 8th European Conference on Speech Communication and Technology, Geneva, pp.181-184, (2003-9).
- [3] Kitazawa Shigeyoshi, Kiriyama Shinya, "Acoustic and Prosodic Analysis of Japanese Vowel-Vowel Hiatus with Laryngeal Effect," Proceedings 8th International Conference on Spoken Language Processing, Jeju, (to appear) (2004-10).
- [4] S-H Bu, M. Yamamoto, S. Itahashi, "Evaluation of a Method for Automatic Determination of Fo Model Parameters," Proc. International Conference: Speech Prosody 2004, Nara, Japan, pp. 431-434 (2004)
- [5] 堀内靖雄, 庵原彩子, 西田昌史, 市川薫, "自然対話における聞き手の反応と話し手のうなずき・言語情報・韻律情報との関係に関する予備的検討," 情報処理学会研究報告, Vol.2004, No.74 (2004-SLP-52-18), pp.93-98 (2004.7).