Vol.2020-MUS-127 No.34
Vol.2020-SLP-132 No.34
2020/6/7

PJS：音素バランスを考慮した日本語歌声コーパス

小口 純矢[2,a)]　高道 慎之介[,b)]

概要：本稿では，高い応用可能性と高い再現性をもった歌声合成研究の実現に向け，フリーの歌声コーパスを構築する．既存の歌声コーパスは，話声コーパスのように音素バランスを担保している保証がなく，また，著作権等の理由により，オープンコーパスとして公開することが困難である．これらの問題に対し本稿では，フリーの音素バランス話声コーパスである声優統計コーパスに対しメロディを付けて歌唱した PJS (Phoneme-balanced Japanese Singing voice) コーパスを構築する．本稿では，その設計方針と構築結果を述べる．

キーワード：コーパス, 歌声合成, 音楽情報処理, 音素バランス

1. はじめに

信号処理と深層学習の発展に恩恵を受け，昨今の歌声合成エンジン（例えば，Sinsy [1]，NEUTRINO [2]）の発展がめざましい．歌声合成研究の更なる発展を見据え，我々はその応用可能性と再現性に着目すべきである．それらを支える要因の一つが，オープンな歌声コーパスであり, 東北きりたん歌唱データベース [3] はその先駆け的存在である．当該データベースは童謡・アニメソング 50 曲から成る大規模歌声データベースであり，前述した NEUTRINO の構築にも利用されている．そのような大規模コーパスが利用される一方，音声収録と機械学習がより容易な小規模コーパスも必要とされる．HTS demo [4] や JVS-MuSiC [5] はその代表例である．しかしながら，これらの歌声コーパスは，小規模データベース特有の問題である音素バランス偏りを考慮しておらず，合成歌声品質の音韻的明瞭性を保証しない．

そこで本研究では，音素バランスを考慮した歌声コーパス PJS (Phoneme-balanced Japanese Singing voice) を構築する．フリーの音素バランス話声コーパスである声優統計コーパス [6] に対しメロディを付けて歌唱することで，音素バランスを担保した歌声コーパスを設計可能である．さらに本コーパスは，以下のような特徴を持つ．

話声収録: 歌声だけでなく，同一テキストを読み上げた話声を収録することで，話声歌声の横断的研究（例えば [7]）にも利用できる．

作曲情報: 楽譜情報だけでなく作曲時の付随情報（音楽ジャンルやスケール）を含むことで, 音楽情報処理研究にも利用できる．

フリーで使用可: 本コーパスに含まれる全データは CC BY 4.0 ライセンスで公開される．そのため，研究用途に限らず商用利用・音楽活動等にも利用できる．

以降では，本コーパスの設計方針と構築結果を述べる．本コーパスはウェブページからダウンロード可能である [8].

2. コーパスの設計

2.1 構成

以下に PJS コーパスのディレクトリ構造を示す．ディレクトリ PJS100_*[SENTENCE_ID]* における *[SENTENCE_ID]* は, 歌唱した声優統計コーパスの音素バランス文の ID に対応している [6].

```
PJS100_001
    PJS100_001_song.wav
    PJS100_001_speech.wav
    PJS100_001.mid
    PJS100_001.xml
    PJS100_001.lab
    PJS100_001.txt
PJS100_002
...
PJS100_100
```

また, ディレクトリ PJS100_*[SENTENCE_ID]* には以下の

[1] 明治大学
Meiji University, Nakano, Tokyo 164–8525, Japan
[2] 東京大学
The University of Tokyo, Bunkyo, Tokyo 113–8654, Japan
[a)] cs202027@meiji.ac.jp
[b)] shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp

ファイルが格納されている.

- **PJS100_*[SENTENCE_ID]*_song.wav**: 音素バランス文にメロディを与えて歌唱した歌声を収録した WAV ファイル
- **PJS100_*[SENTENCE_ID]*_speech.wav**: 音素バランス文を読み上げた音声を収録した WAV ファイル
- **PJS100_*[SENTENCE_ID]*.mid**: 歌声収録時にガイドメロディとして使用した MIDI ファイル
- **PJS100_*[SENTENCE_ID]*.xml**: 歌声の楽譜情報を記述した musicXML ファイル
- **PJS100_*[SENTENCE_ID]*.txt**: 作曲するうえで参考にした音楽ジャンルなどの付随情報

2.2 作曲条件

作曲は, 20 代の男性日本語話者 1 名によって行われた. 彼はプロの作曲家ではないが自身の歌声と作曲能力を利用した業務経験を持つ. 作曲者は, 音素バランス文に合致するメロディを地声で歌唱可能な音域内で当てはめた. その際に, 作曲者が暗黙的に仮定した, あるいはなるべく多様なメロディが含まれるように制約として加えた音楽的付随情報 (楽曲・ジャンル・スケールなど) を PJS100_*[SENTENCE_ID]*.txt に記述した.

2.3 収録条件

歌声の歌唱者, 話声の発話者は作曲者と同一人物とした. 歌唱者は, あらかじめ作成した midi ファイルから生成されたメロディを聴きながら, ピッチとタイミングがずれないように歌唱した. また, 近接効果の影響を低減するため, 極力, マイクロフォンと音源間の距離を 15 cm に保ちながら歌唱するように指示を与えた. 音声収録は, 吸音材を壁面に貼りつけた簡易防音室の中で行われた. 無響室やレコーディングスタジオで収録したわけではないため, 後処理による雑音抑圧を行いやすいように, この簡易防音室内の暗騒音を, 収録日毎に 15 秒程度収録した. マイクロフォンには Lewitt LCT 441 FLEX の Cardioid モードを, ウインドスクリーンには JZ MICROPHONES Pop Filter, オーディオインターフェースには RME Fireface UCX を用いた [9], [10], [11]. 歌声および読み上げ音声ファイルは, ともに 48 kHz サンプリングおよび 16 bit 量子化の RIFF WAV 形式とした.

3. 構築結果

3.1 全体の傾向

ここでは, コーパス全体としての音楽的な性質を議論する. まず, キーについて, **Figure 1** に示すように, 主音はバランス良く含まれている一方で, 短調の楽曲がやや少ない. また, テンポについては, **Figure 2** に示すように, 80 から 160 BPM (Beat Per Minute) の範囲に分布している. したがって, 極端にテンポが遅いまたは速く, 短調の楽曲を合成する目的には本コーパスは適さないことが示唆される.

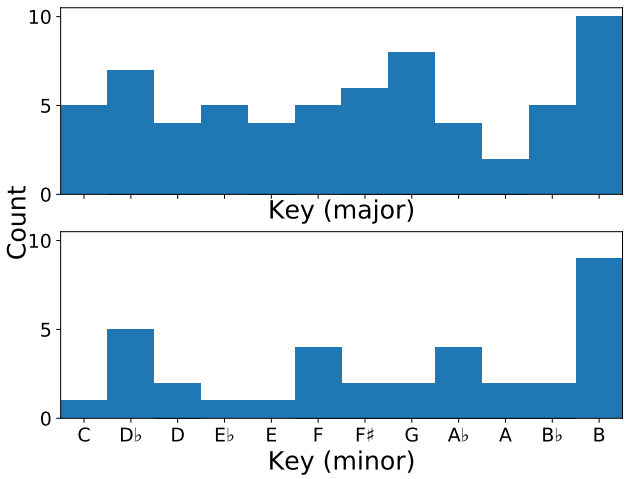

Figure 1 PJS コーパス全体のキーのヒストグラム. 短調のメロディが長調より少ない傾向にある.

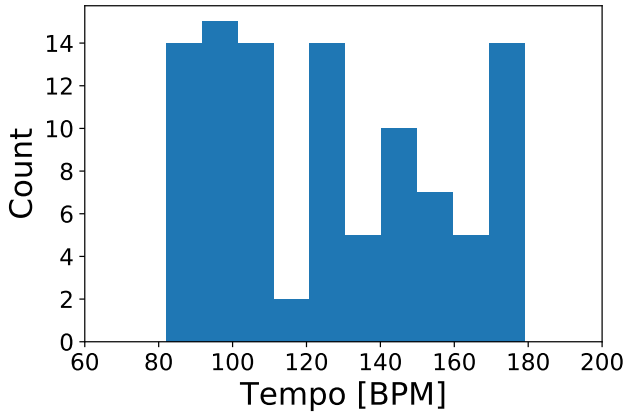

Figure 2 PJS コーパス全体のテンポのヒストグラム. 80 から 160 BPM (Beat Per Minute) の範囲に分布しているが, 80 BPM より遅いまたは 180 BPM より速いメロディは含まれない.

3.2 楽譜情報のアノテーション

musicXML の作成には歌唱者自らが人手で行った. 音符の長さは, 作曲時に作成されたガイドメロディ用 MIDI ファイルを基準とした. 一例として, PJS100_001.xml から生成した楽譜を **Figure 3** に示す. 「またとうじ」や「みょうおうの」のように, 複数の音節に 1 つのノートを対応させた例が散見される. つまり, 音節とノートが一対一対応しておらず, 音素ラベルへ変換する際には, ノートを音節の数に合わせて複製するなどの処理 (例えば [12]) を加える必要があることに注意されたい.

4. おわりに

本研究は, 音素バランスを考慮した歌声コーパスである PJS コーパスを構築した. 声優統計コーパスの音素バランス文にメロディを付与し歌唱することで音素バランスを担保した. 本コーパスには, 歌声だけでなく読み上げ音声や作

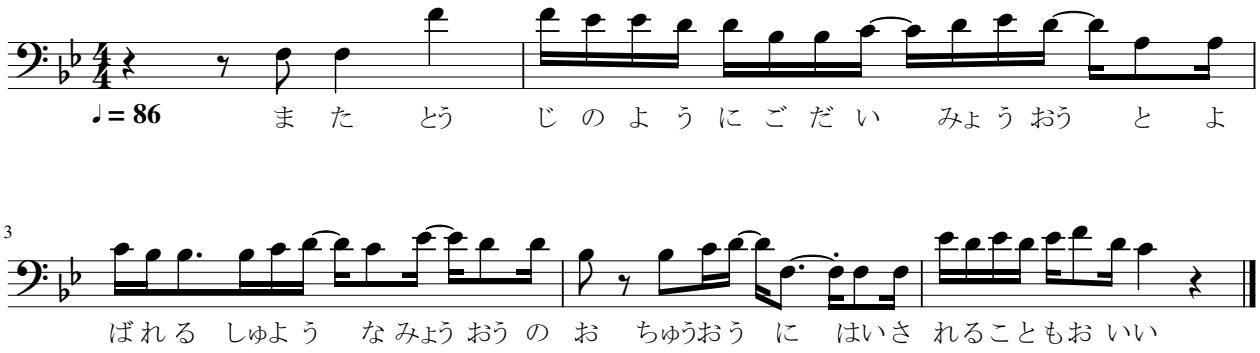

Figure 3 PJS100_001.xml から生成した楽譜. 元の音素バランス文は「また，東寺のように，五大明王と呼ばれる，主要な明王の中央に配されることも多い」である．1 小節目「またとうじ」や 2 小節目「みょうおうの」のように, 複数の音節に 1 つのノートを対応している例がある.

曲情報といった歌声以外の情報が豊富に付属しており, 歌声合成にとどまらない幅広い応用が期待される. 今後の展望として, 音素だけでなく, 音楽的なバランスの考慮や地声だけでなく裏声やグロウル歌唱など様々な声質の追加収録などが挙げられる.

PJS コーパスに含まれる全データは，CC BY-SA 4.0 でライセンスされており，[8] より入手可能である.

謝辞 本研究は，東京大学 GAP ファンドプロジェクト「音声合成技術の研究開発・商用利用を加速させる音声コーパスの設計・構築」の支援を受けて実施した.

参考文献

[1] “Sinsy,” http://www.sinsy.jp/.
[2] “NEUTRINO,” https://n3utrino.work/.
[3] 森勢将雅, “東北きりたん歌唱データベース,” https://zunko.jp/kiridev/login.php.
[4] “HMM-based speech synthesis system (HTS),” http://hts.sp.nitech.ac.jp/.
[5] H. Tamaru, S. Takamichi, N. Tanji, and H. Saruwatari, “JVS-MuSiC: free Japanese multispeaker singing-voice corpus,” *arXiv preprint 2001.07044*, Jan. 2020.
[6] y_benjo and MagnesiumRibbon, “声優統計コーパス,” http://voice-statistics.github.io.
[7] Y. Ohishi, M. Goto, K. Itou, and K. Takeda, “Discrimination between singing and speaking voices,” in *Proc. EUROSPEECH*, Lisbon, Portugal, Sep. 2005, pp. 1141–1144.
[8] “PJS: Phoneme-balanced japanese singing voice corpus,” https://sites.google.com/site/shinnosuketakamichi/research-topics/pjs_corpus.
[9] Lewitt, “440 FLEX, howpublished = https://www.lewitt-audio.com/microphones/lct-recording/lct-441-flex,.”
[10] JZ MICROPHONE, “Pop filter,” https://intshop.jzmic.com/collections/accesories/products/pop-filter.
[11] RME, “Fireface UCX,” https://www.rme-audio.de/fireface-ucx.html.
[12] K. Nakamura, K. Oura, Y. Nankaku, and K. Tokuda, “HMM-based singing voice synthesis and its application to Japanese and English,” in *Proc. ICASSP*, Florence, Italy, May 2014, pp. 265–269.