絵本読み聞かせ音声コーパスの構築とそのラベリングに関する検討*

☆百武恭汰, 齋藤大輔, 峯松信明 (東大)

1　はじめに

近年、文字情報を音声情報に変換する音声合成技術は、スマートフォンやコミュニケーションロボットの普及により活躍の場を広げている。そうした中で、テキストを単調に読み上げるだけでなく、多様な読み方を実現する音声合成が求められている。代表的な検討例としては、喜びや怒りなどの感情が込められた感情音声の合成が挙げられる。

ところで、我々が実際に話す場面を振り返ると、感情だけでなく、目上の人には敬語を使って丁寧な口調で話すなど、相手によっても話し方を変えていることが伺える。前段で挙げた例を expressive な音声合成とすると、こうした相手との関係性を反映したものは social な音声合成と言うことができ、これも多様な音声合成を実現するための軸と捉えることができる。

本研究では、相手によって話し方を変える例として、子どもに向かって（主に母親が）発話するときに見られる対乳児音声（Infant-directed speech; IDS）を取り上げる。IDS には一般にピッチが高くなりそのレンジが広がる、ポーズが多く、長くなるなど、IDS 特有の特徴が見られることが知られている [1]。IDS 風音声合成の応用例としては絵本の読み聞かせが考えられ、その実現を目標として絵本読み聞かせコーパスの構築を行った。なお、英語の audiobook を題材とした音声合成はこれまでに Blizzard Challenge でも課題となっている [2]。

本稿では、絵本読み聞かせ音声コーパス構築の様子、および音声合成に向けた収録音声へのラベリングに関して述べる。

2　対乳児音声（IDS）

IDS は子どもの言語獲得と関係があると考えられており、これまでに様々な分析研究例が存在する。特に日本語 IDS に関しては理化学研究所により『理研母子会話コーパス（R-JMICC）』 [3] が構築され、様々な分析が行われている。

一般に IDS では話速が遅くなると思われがちであるが、実際には話速はほぼ変わらず、発話が短くポーズが多いことからそのように感じられるとされている [6]。また、アクセント句末において上昇を伴う局所的なピッチの変動が見られる複合境界音調（boundary pitch movement; BPM） [4] について、IDS では通常の発話に比べて大きな変動が見られ、ピッチレンジが拡大していることが確認されている [3]。

Table 1　絵本読み聞かせコーパスの概要

話者	女性保育士 2 名
スタイル	アナウンサー調、読み聞かせ調
文数	各話者・各スタイル 919 文

3　絵本読み聞かせコーパスの構築

今回構築したコーパスの概要を Table 1 に示す。

話者については、絵本の読み聞かせに熟達した保育士から候補を募った結果、6 名の立候補者が集まった。立候補者の読み聞かせ音声サンプルを 43 名の日本人に web 上で聴取させ、「自分の子どもに読み聞かせを依頼する場合、誰を選ぶか」という観点から選定させた。この結果、上位 2 名を収録用話者として採用した。1 名は地の文とセリフの文との読み方の変化が特徴的であった。また、もう 1 名は句末に BPM が頻繁に見られた。

コーパスの文章としては、絵本の文章をそのまま用いた。今回用いた絵本は 7 冊（うち一冊は前半部のみ）、文数は 919 文である。登場頻度の低いモーラをカバーするため、かるた遊びのように各々のひらがなについて例文を提示しているものを 1 冊採用した。また、文数については、「『おはよう』と、おかあさんが言いました」のような引用文は基本的に 2 文としてカウントしている。

これらの文章に関して各音素の出現頻度をカウントし、ATR 音素バランス文 [5] と比較したものを Fig. 1, 2 に示す。音素数、モーラ数共に合計数は ATR 音素バランス文を上回っており、なおかつ音素バランス性もある程度保たれていると言える。なお、モーラごとに見ると絵本の文章には「し」「た」「ま」の 3 つが特に多く現れていた。これは、絵本において「〜ました。」「〜しました。」という文が多いことを示唆していると考えられる。

以上の条件で、各話者について「アナウンサーのような単調な読み方（読み上げ調）」「子どもに向かって

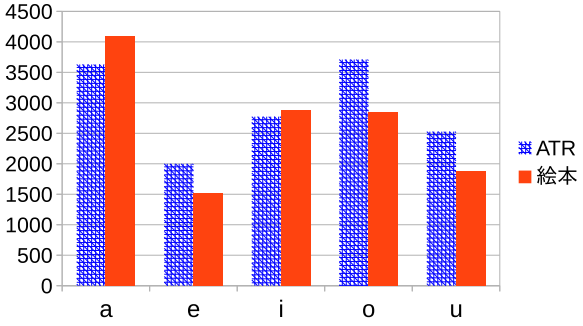

Fig. 1　母音の出現頻度

*Construction of a corpus of infant-directed storytelling speech and its labeling by HYAKUTAKE Kyota,SAITO Daisuke, and MINEMATSU Nobuaki, (The University of Tokyo)

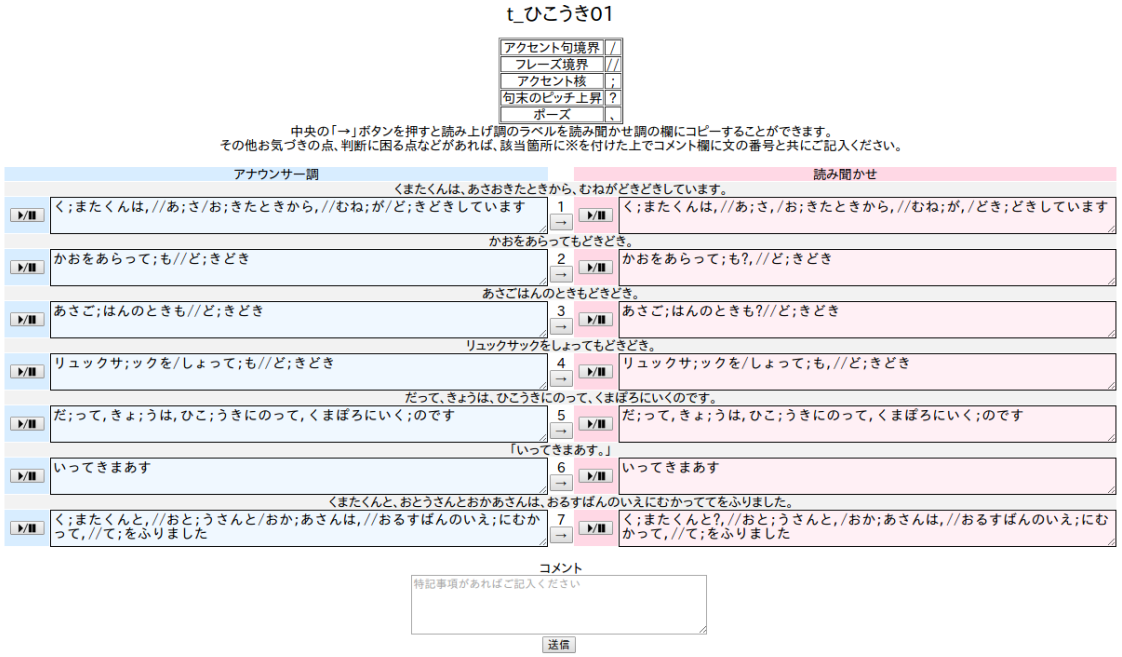

Fig. 3　ラベリングインターフェースの様子

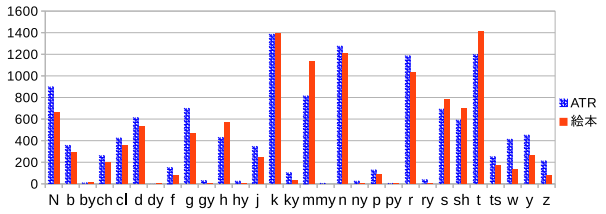

Fig. 2　子音の出現頻度

読み聞かせるような読み方（読み聞かせ調）」の 2 通りのスタイルによる絵本の読み上げ音声を収録した。なお、前後の文脈による特徴が損なわれないよう、原則として見開きページごとに収録を行った。

4　収録音声へのラベリング

今回収録した音声に対し、音声合成に利用するためのラベリングを行っている。収録に参加した保育士のうち 1 名と発達心理学を専攻する学生 6 名がラベリング作業にあたっている。HMM 音声合成による合成を想定し、HTS-2.2[1] において利用されているラベルを参考として韻律情報に関するラベルの付与を行っている。それに加え、より読み聞かせらしい音声の合成を実現するため、絵本の読み聞かせ音声に特徴的に見られる現象についてもラベルを付与している。

実際のラベリング作業の様子を Fig. 3 に示す。読み上げ（図中「アナウンサー調」）と読み聞かせの各音声を聞き比べながら、両者をシンボリックにラベリングする。特に、読み聞かせ特有の現象へのラベリングを効率的に行うことができるよう、インターフェースを構築した。現時点では図中にも示されている通り、「アクセント句境界」「アクセント核」「フレーズ（イントネーション句）境界」「句末のピッチ上昇（上昇調 BPM）」「ポーズ」の 5 項目に関してラベルを付与しているが、新たに考慮すべき特徴が見られた際には随時ラベリング項目の追加を行っていく予定である。これまでに「いってきまあす」のようなセリフ文に対して長音が極端に長く発音される現象が確認されており、ラベルの追加を検討をしている。

[1]HTS, http://hts.sp.nitech.ac.jp

5　おわりに

本稿では、多様な音声合成の一つとして絵本読み聞かせ風音声の合成を目標とし、そのために構築したコーパスの概要およびラベリングに際しての検討事項について述べた。

今後は新たなラベリング項目の追加を検討しつつラベリングを進めていく。ラベリング作業終了後、音声合成に取り組んでいく予定である。現時点ではラベルを付与したテキストからの読み聞かせ音声の合成を目標としている。読み聞かせ音声から HMM を構築する際に今回導入したラベルを明示的に利用した場合、HTS-2.2 で採択されているラベルのみを用いた場合に比べ、合成音声の品質にどの程度の変化が現れるかについても検討していきたい。

参考文献

[1] M. Soderstrom, Developmental Review, 27, pp. 106–119, 2007.
[2] S. Takaki *et al.*, The Blizzard Challenge, 2013.
[3] Y. Saikachi *et al.*, 第 3 回コーパス日本語学ワークショップ, pp. 383–392, 2013.
[4] K. Maekawa *et al.*, “『日本語話し言葉コーパス』のイントネーションラベリング Version 1.0,” 2004.
[5] A. Kurematsu *et al.*, Speech Communication, vol. 9, pp. 357–363, 1990.
[6] Y. Igarashi *et al.*, 信学技報 SP2006-90, pp.31–35, 2006.