6M-09

英語音声における連音の自動検出の検討

佐藤 玄† 加藤正治† 小坂哲夫†
†山形大学大学院理工学研究科

1 はじめに

日本人は英語ネイティブ話者の発音を聞いたとき分からない発音があった場合，その音を知らない音であると考え，簡単な単語であっても聞き取れなくなる場合がある．特に，連音と呼ばれる発音変化の聞き取りが困難であると言われている．このような連音を聞き取れるようになるには，まず連音を発音できるようになる必要がある [1] という主張がある．よって，本研究ではこの主張に従って，連音に関して調査を行い，さらに連音自動検出システムの検討を行う．

2 連音の種類と発生確率の調査

連音とは，特定の音素が連続する場合にその組み合わせによって発音が変化することである．日本人英語では単語と単語が連結されず単語ごとに区切られた発音になっている場合が多く存在する．しかし，米国人英語の発音では連音が発生する場合が多い．例えば，hand in を日本人は「ハンド イン」と発音する傾向があるのに対し，米国人は「ハンディン」のように発音する．米国人は連音を用いて発音していることが日本人が米国人英語を聞き取りづらいことの一因となっている．外連声とは単語境界で起こる連音であり，内連声とは形態素内で起こる連音である．連音の種類を表 1 に示す．下線部は連音が発生する箇所を示している．表中の LINK は Linking(連結)，DEL は Deletion(脱落)，FLAP は Flapping(弾音化)，ASS は Assimilation(同化)，INS は Insertion(挿入)，RED は Reduction(弱化) を示している．また，'B-' は外連声，'W-' は内連声を表している．

従来どのような種類の連音があるかについては種々の文献や書籍で述べられている [2][3] が，どの程度の頻度で連音が出現するかについては調査されていなかった．このため，実際のネイティブ話者の文発声でどの程度連音が発生しているのか調査を行った．調査では TIMIT コーパス [4] を用いた．TIMIT コーパスには英語教師が音声を聞き，聞こえたように音素ラベルを付与した手動ラベリングデータが存在する．この調査では，TIMIT コーパスの単語辞書の発音表記と発話ごとの手動ラベリングの発音表記を比較した．結果を表 2 に示す．表中の発生は連音が発生する可能性のある箇所で連音が発生した割合，変化なしは音素が変化しないもの，その他は連音以外の音素変化の割合を示している．調査の結果，B-FLAP1，W-FLAP1，W-INS1 などが高確率で連音が発生し，W-RED1，W-DEL2 などが比較的連音が発生する確率が低いことが分かった．

表 1: 連音の種類

規則	内容	例
B-LINK1	子音と母音が連結	hand in
B-LINK2	二重母音と母音が連結	be on
B-DEL1	破裂音と子音が連結し破裂音が脱落	good time
B-DEL2	同じ子音同士が連結し 1 つが脱落	hot tea
B-FLAP1	前後が母音の/t/，/d/が/r/に変化	not at all
B-ASS1	/s/，/z/，/t/，/d/と/y/が同化し，それぞれ/ʃ/，/ʒ/，/tʃ/，/ʤ/に変化	would you
B-ASS2	/s/，/z/と/ʃ/または/n/と/ð/が同化し，それぞれ/ʃ/，/n/に変化	dance show
W-DEL1	破裂音が連続する場合，前の無声破裂音が脱落	outdoor
W-DEL2	破裂音 + 子音の場合，破裂音が脱落	lastly
W-DEL3	鼻音と破裂音が連続する場合，破裂音が脱落	endless
W-DEL4	単語末の破裂音が脱落	help
W-FLAP1	前後が母音の/t/，/d/が/r/に変化	rider
W-ASS1	/str/，/ntr/，/t(ə)r/の/t/の後ろに/ʃ/が挿入され/stʃr/，/ntʃr/，/tʃ(ə)r/に変化	destroy
W-INS1	/ns/の間に/t/が挿入	sense
W-RED1	機能語 (人称代名詞など) が弱化し，音が変化	to

3 連音の自動検出

3.1 検出システム

本研究で用いる検出システムについて図 1 に示す．このシステムは基本的には発音誤り検出システムと同様の構成となっている [5]．発音例文は標準的な発声に対応する音素列に変換された後，連音規則を適用し，複数の音素列を生成する．これらの音素列は母国語話者音素の HMM を用いて音素 HMM 列に変換される．一方，発音例文に基づく学習者音声からは音声分析によって認識用パラメータ (MFCC) が抽出される．各音素 HMM 列と学習者音声からの特徴を用いて強制アライメント (Forced Alignment) が行われる．複数の音素 HMM 列の中から最大尤度の HMM 列が出力され，学習者音声が対応する音素区間に分割される．本システムでは連音規則と強制アライメントを利用することによって連音を検出することができる．連音検出の流れを図 2 に示す．実線が連音が検出された発音のルート，点線が連音が検出されない発音のルートである．この例では，尤度により/r/か/t/を選択する．/r/が選択された場合，連音検出 (W-FLAP1) と判定される．

3.2 実験条件

本研究では英語音響モデルの学習データとして ERJ コーパスの話者数 20 名 (男性 8 名，女性 12 名) が 399〜403 文づつ発話した 8051 文の General American 話者英語音声を用いる [6]．また，米国人話者の評価データとして話者数 630 名 (男性 438 名，女性 192 名) が 10 文づつ発話した 6300 文の TIMIT 音声データを用いる．今回は内連声に関して連音検出率を検討する．図 3 はネイティブ話者の発音における

Automatic Detection of Sandhi in English Speech
Haruka Sato†, Masaharu Kato†and Tetsuo Kosaka†
†Graduate School of Science and Engineering, Yamagata University
992-8510, Yonezawa, Japan
trx29924@st.yamagata-u.ac.jp

表 2: 連音発生確率

連音条件	発生	変化なし	その他
B-LINK1	32.58%	0%	67.42%
B-LINK2	35.39%	0%	64.61%
B-DEL1	46.79%	40.70%	12.51%
B-DEL2	65.05%	5.09%	29.86%
B-FLAP1	90.61%	0%	9.39%
B-ASS1	38.18%	31.25%	30.57%
B-ASS2	28.68%	48.84%	22.48%
W-DEL1	71.65%	4.12%	24.23%
W-DEL2	18.11%	72.30%	9.59%
W-DEL3	39.92%	37.64%	22.44%
W-DEL4	53.68%	34.79%	11.53%
W-FLAP1	81.49%	0%	18.51%
W-ASS1	0%	76.56%	23.44%
W-INS1	89.58%	0%	10.42%
W-RED1	19.92%	30.66%	49.42%

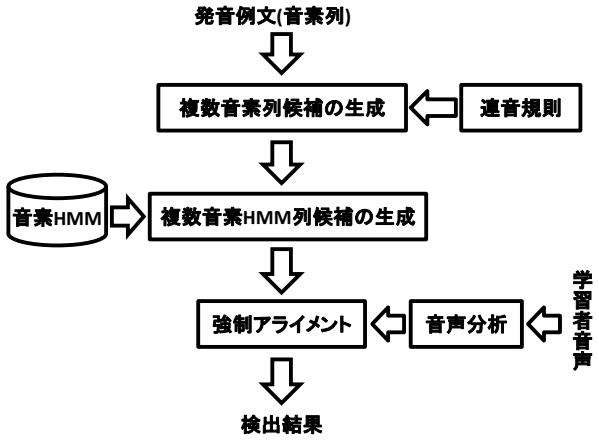

図 1: 連音検出システム

内連声が発生する可能性のある箇所の割合を示したものである．なお，一部ルール間の重複があるが，二重にカウントした (表 2 も同様)．図 3 を見ると挿入の可能性がある箇所が非常に少ないことが分かる．また，表 2 より弱化は発生確率が低いためこれら 2 つの規則を除き，脱落とフラッピングに関して連音検出率の算出を行う．

3.3 実験結果

本実験の結果を表 3 に示す．連音検出率は約 64%となった．また，種類別で見ると脱落の検出率が約 65%，フラッピングの検出率が約 52%となっている．結果を見ると非連音→連音の誤りが多い．代表的な誤りの例を図 4 に示す．dark の k は実際には該当の音素が聞こえるが，発話時間が短いため

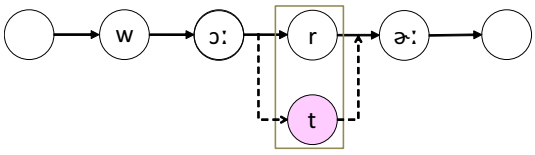

図 2: 連音検出の流れ (例：water)

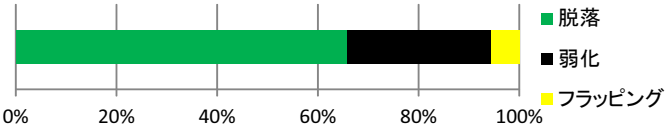

図 3: ネイティブ話者の発音における内連声の発生可能性箇所の割合 (挿入は 0.14%のため表示していない)

表 3: 内連声の連音検出率

		自動検出結果	
		連音	非連音
手動ラベリング	連音	94.58%	5.42%
	非連音	66.05%	33.95%
連音検出率		64.26%	

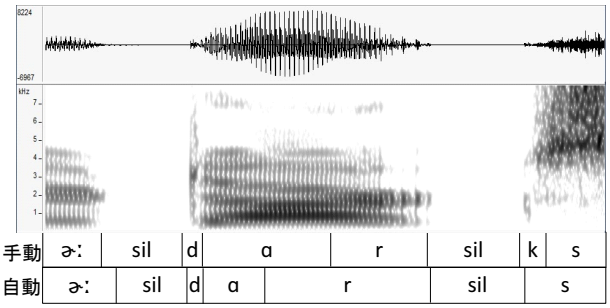

図 4: 脱落の例 (発話内容：She had your dark suit〜の dark の部分) 上：手動ラベル，下：自動検出結果，sil は無音を表す

検出することができなかったと考えられる．その結果，脱落の連音検出率が低下し，全体としての連音検出率が低下したと考えられる．

4 まとめ

本研究ではまず，連音の条件を調査し，それぞれの連音についてネイティブ話者の発音における発生確率を調査した．また，連音の自動検出システムを構築して評価実験を行ったところ，約 64%の検出率が得られた．

参考文献

[1] 藤田 英時：”知ってる英語なのになぜ聞き取れない？”，ナツメ社 (2003)
[2] 竹林 滋，清水あつ子：”英語音声学・音韻論入門”，研究社 (2002)
[3] 長尾 和夫，アンディ・バーガー：”聴こえる！話せる！ネイティヴ英語発音の法則”，DHC(2006)
[4] TIMIT Acoustic-Phonetic Continuous Speech Corpus - Linguistic Data Consortium, https://catalog.ldc.upenn.edu/ldc93s1
[5] 河合剛，石田朗，広瀬啓吉，”2 言語間の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価”，日本音響学会誌 57 巻 9 号，pp.569-580 (2001)
[6] 峯松 信明，富山 義弘，吉本 啓，清水 克正，中川 聖一，壇辻 正剛，牧野 正三，”英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築”，日本教育工学会論文誌 27 巻 3 号，pp.259-272 (2004)