情報処理学会論文誌 Vol.54 No.2 452–462 (Feb. 2013)

日本語の未知語に対する発音付与のための多対多アライメント

久保 慶伍[1,a)] 川波 弘道[1,b)] 猿渡 洋[1,c)] 鹿野 清宏[1,d)]

受付日 2012年5月31日, 採録日 2012年11月2日

概要：音声ドキュメント検索や音声検索のような固有名詞や新語を扱うアプリケーションの発達とともに，未知語に対する頑健な自動発音付与の必要性は増加している．未知語への自動発音付与手法として統計的アプローチやWebテキストマイニングによるアプローチがある．これらには単語の表記と発音を単語（形態素）よりも小さい単位で対応付けたデータが不可欠である．本論文では日本語の未知語に対する発音付与の性能向上を目的として，表記と発音の対応付けの精度を劣化させずに，未知語を表現する能力が高い小さい単位での対応付けを求めるアライメントアルゴリズムを提案する．また，Webテキストマイニングを用いた日本語の未知語に対する自動発音付与により提案手法の評価を行った．評価実験の結果，提案手法は従来手法が持つ精度をほとんど劣化させずに，未知語に対する汎化能力を表す再現率を約3.9ポイント改善した．

キーワード：文字列アライメント，Joint multigram，未知語，自動発音付与，Webテキストマイニング

Many-to-many Alignment Algorithm for Automatic Pronunciation Annotation on Japanese Unknown Words

Keigo Kubo[1,a)] Hiromichi Kawanami[1,b)] Hiroshi Saruwatari[1,c)]
Kiyohiro Shikano[1,d)]

Received: May 31, 2012, Accepted: November 2, 2012

Abstract: The need for robust pronunciation annotation on unknown words has been increasing with the development of an application that deals with proper nouns and brand-new words, such as Spoken Document Retrieval and Voice Search. In robust pronunciation annotation on unknown words, the alignment between graphemes and phonemes is vital data. In this paper, for the purpose of the improving pronunciation annotation on Japanese unknown words, we propose the alignment algorithm that requires a mapping with small unit having high expression ability on unknown words while avoiding degradation of the accuracy of a mapping between graphemes and phonemes. An evaluation experiment of a many-to-many alignment by automatic pronunciation annotation using Web text mining is also performed. That experimental result shows that the proposed many-to-many alignment obtains 3.9 point improvement on recall rate that represents the generalization ability for unknown words while avoiding degradation of the accuracy of the pronunciation annotation compared with the conventional approach.

Keywords: character alignment, Joint multigram, out-of-vocabulary, pronunciation annotation, Web text mining

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630–0101, Japan
[a)] kcigo-k@is.naist.jp
[b)] kawanami@is.naist.jp
[c)] sawatari@is.naist.jp
[d)] shikano@is.naist.jp

1. はじめに

近年，Web上に存在する音声データの言語情報を検索する音声ドキュメント検索[1]や音声でWeb検索を行う音声検索[2]など，ドメインを限定しない大語彙の音声認識

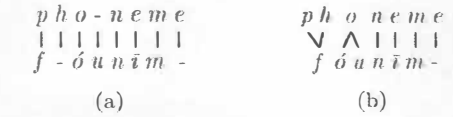
© 2013 Information Processing Society of Japan

アプリケーションの実用化が進んでいる．音声ドキュメント検索や音声検索において固有名詞や新語は検索結果を効率的に絞り込むための検索語として使われることが多いため，これらのアプリケーションの性能は固有名詞や新語に対する音声認識性能に大きく依存する．そのため，新たに表れるこれらの単語（未知語）に対処するためには認識辞書と言語モデルを継続的に更新しなければならず，未知語獲得の自動化は重要な研究課題となっている．

Yahoo! 検索ランキング [3] などの Web 上の言語資源を継続的に自動収集すれば，未知語の表記は自動獲得できるが，音声認識や規則音声合成で必要な発音（読み）は必ずしも提供されないため，自動発音付与技術が必要となる．従来，日本語の自動発音付与には Kytea [4] などの形態素解析器が用いられてきた．これは入力された文字列を形態素へと分割し，形態素ごとに形態素辞書に登録された発音を付与する．これは辞書ベースの発音付与といえる．しかし，このような辞書ベースの手法では辞書に登録されていない形態素（未知語）に対して発音を付与することはできない．音声ドキュメント検索や音声検索のタスクにおいてよく出現する固有名詞や新語は未知語である可能性が高く，辞書ベースの手法では正確な発音付与が困難である．そのため，固有名詞や新語に対して誤った発音を付与する可能性が高く，誤った発音が付与されると音声ドキュメントなどの音声認識時にそれらを認識することができなくなる．音声ドキュメント検索の場合，検索語になりやすい固有名詞や新語が，音声ドキュメントにおいて正しく認識されないと検索性能は劣化すると考えられる．

また，単語単位とサブワード単位を併用する音声ドキュメント検索システムの検索を考えた場合，入力された検索語に関する適切なサブワード系列を得るために，固有名詞や新語が多い検索語に対して正しい発音を付与する必要がある（検索語入力時にユーザが入力したひらがなの系列を獲得できる場合があるが，必ずしも正しい発音とは限らず，またコピー&ペーストや英字・記号ではその系列を得られないため，検索語に対して発音を付与する必要がある）．これらのことから，上記の音声認識アプリケーションにおいて未知語（固有名詞や新語）に頑健な自動発音付与技術が必要となる．

未知語に対する自動発音付与の研究として，オンライン識別訓練 [5] などの統計的アプローチや，括弧表現に基づく Web テキストマイニングを用いたアプローチ [6] が提案されている．これらの手法では，図 1 のように単語（形態素）より小さい単位で表記と発音を対応付けたデータが必要となる．本論文では，図 1 の *ph/f* のような表記の部分文字列と発音の部分文字列の対応関係を「対応付け」と呼び，「ph/f o/óu n/n e/ī m/m e/-」のような対応付けの系列を「アライメント」と呼ぶ．このデータを単語（形態素）辞書から自動生成することを目的として，表記と発音の自動アライメントの研究が行われている．



(a) (b)

図 1 表記 *phoneme*，発音 *fóunīm* のアライメント例．(a) が 1 対 1 アライメントの例．(b) が多対多アライメントの例．「-」は対応する部分文字列がないことを表す．本論文では「-」が発音側に表れる対応付けを削除文字，表記側に表れる対応付けを挿入文字と呼ぶ

Fig. 1 Example of the alignment in the word *phoneme* pronounced as *fóunīm*. (a) is the example of the one-to-one alignment. (b) is the example of the many-to-many alignment. The "-" means that there is no the corresponding substring. We call a mapping that the "-" appeared on the pronunciation side "deleted character", and a mapping that the "-" appeared on the notation side "inserted character" in this paper.

表記と発音の自動アライメントの先行研究として EM アルゴリズム [7] の概念を用いた教師なしアライメントが提案されている [8], [9], [10], [11]．この自動アライメント手法の代表例として，図 1 (a) のように対応付けを表記 1 文字，発音 1 文字に制限した 1 対 1 アライメント [8] や図 1 (b) のように表記，発音の両方において複数文字との対応付けを許した多対多アライメント [9], [10], [11] がある．なお，図 1 において，「-」は対応する文字が存在しないことを示し，これが発音側に表れる対応付けを本論文では削除文字と呼び，逆に「-」が表記側に表れる対応付けを挿入文字と呼ぶ．表記と発音の多対多アライメントの場合，挿入文字の出現頻度は限りなく少ないため，本論文では挿入文字を考慮しない．図 1 のようにアライメントされたデータは発音付与のために利用される．

文献 [9] において，多対多アライメントは Joint multigram と呼ばれており，文献 [10] で提案されている手法と本質的に同じである．本報告ではこの従来の多対多アライメントの手法を Joint multigram と呼ぶ．Jiampojamarn らの報告 [10] によると Joint multigram は 1 対 1 アライメントよりも自然な表記と発音の対応付けを行い，統計的アプローチを用いた自動発音付与において 1 対 1 アライメントを上回る性能を実現している．

ただし，Joint multigram はアライメントに含まれている対応付けの数の違いによる影響を考慮していないため，相対的に大きい単位での表記と発音の対応付けが優先される．たとえば，図 1 の表記 *phoneme*，発音 *fóunīm* の対データの場合，Joint multigram は図 1 (b) の *ph/f* のような小さい単位での対応付けよりも，*phon/fóun* や *phoneme/fóunīm* といった大きい単位での対応付けを優先する傾向にある．もし，このような大きい単位での対応付けがアライメント推定時に選ばれると，たとえば，未知語 *nome* (発音 *nóum*) に対して発音を付与する際に，アライメントで得られた対

© 2013 Information Processing Society of Japan

応付けだけでは *nome* を表現することができず，発音は付与できなくなる．つまり，大きい単位での対応付けは未知語を表現する能力（汎化能力）が低い．そのため，文献 [10] では対応付ける最大文字数を表記と発音の各々において 2 文字に制限し，大きい単位で対応付けされることを防いでいる．この方法では対応付けの最大文字数を固定するため，最大文字数を超える表記と発音の対応関係（たとえば「従兄弟（イトコ）」などの熟字訓）に対して不正確な対応付けが行われてしまう．これにより，発音付与の精度が劣化する可能性がある．逆に，最大文字数を増やすと，前述のように汎化能力が低い大きい単位での対応付けが選ばれてしまう．

対応付けの最大文字数を制限することなく汎化能力の高い小さい単位での対応付けを可能とする多対多アライメントの実現のために，我々はアライメントの学習時と推定時に，表記と発音の個々の対応付けに関するパラメータを表記と発音の長さの和（表記と発音のマトリックス上での市街地距離）でべき乗することでアライメントに含まれている対応付けの数の違いを解消する方法を提案している [11]．その結果，統計的アプローチによる自動発音付与において従来の Joint multigram を超える性能を示した．ただし，その手法では削除文字が発生しやすく，また，データセットによっては熟字訓などの特殊な発音を持つ単語に対して不正確な対応付けを行うという問題があった．本論文では前者の問題に対してアライメントの学習時に N-best Viterbi トレーニングを用いて解決する方法を提案し，後者の問題に対しては不正確な対応付けを隣の対応付けと結合させることで正確な対応付けに修正する方法を提案する．本手法は Joint multigram における最大文字数のような性能に大きな影響を与えるヒューリスティックなパラメータの最適値を求める必要がない．

本論文の提案手法はアライメント手法であるため，直接，表記に発音を付与することはできない．提案手法を用いてアライメントしたデータを，統計的アプローチや Web テキストマイニングによるアプローチで用いることではじめて発音付与が可能となる．また，表記と発音のアライメント手法の評価方法としてはアライメント内における対応付けの正解率で評価すべきであるが，表記「would」，発音「ウッド」のような正解の対応付けを一意に決定することが難しいデータが存在するため，表記と発音のアライメントの先行研究 [8], [9], [10], [11] においては，統計的アプローチの学習データに提案手法と従来手法でアライメントしたデータを用いて，その自動発音付与性能を評価することで間接的にアライメント手法の性能を評価している．そのため，本論文においても提案するアライメント手法を自動発音付与手法により間接的に評価する．評価に用いた自動発音付与手法は，括弧表現に基づく Web テキストマイニング [6] によるアプローチで，発音を付与されるデータは日本語の未知語のデータである．この評価は間接的であるが，我々のアライメント手法を研究する動機が未知語に対する自動発音付与性能の向上であるため，我々のアライメント手法が未知語に対する自動発音付与性能にどれだけ貢献できるかを評価することは，我々のアライメント手法の評価方法として妥当だと考えられる．

本論文の構成は以下のとおりである．2 章では従来手法である Joint multigram と提案手法の概要を説明する．3 章では提案手法において用いられるパラメータの学習方法を説明する．4 章では N-best Viterbi トレーニングを用いた削除文字推定の改善手法を示し，5 章では不正確なマッピングの結合手法を示す．6 章では提案手法の未知語に対する自動発音付与における有効性を評価し，7 章でまとめを述べる．

2. 多対多アライメント

2.1 変数の導入

はじめに本論文で用いる変数を定義する．表記 *phoneme*，発音 *fóunīm* のように単語の表記と発音からなるデータを対データ d，対データの集合をデータセット D と定義する．表記の部分文字列と発音の部分文字列の 1 つの対応関係を対応付け u と定義する．また，対データ d は対応付け u の系列により構成することができる．この系列をアライメント $\boldsymbol{u}$ とし，対データ d において可能なアライメント $\boldsymbol{u}$ の集合を $\boldsymbol{U}_d$ と定義する．各変数の例をまとめたものを以下に示す．

$$
\begin{aligned}
&d = \langle phoneme, fóunīm \rangle \\
&D = \{\langle phoneme, fóunīm \rangle, \langle pheasant, féznt \rangle, ...\} \\
&u = \text{ph/f} \\
&\boldsymbol{u} = \text{ph/f o/óu n/n e/ī m/m e/-} \\
&\boldsymbol{U}_d = \{phoneme/fóunīm, phonem/fóunīm \text{ e/-}, \ldots, \\
&\qquad\qquad \text{ph/ph o/óu n/n e/ī m/m e/-}\}
\end{aligned}
$$

2.2 Joint multigram

Joint multigram によるアライメントの推定は以下のように定式化される [9]．

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u} \in \boldsymbol{U}_d} P(\boldsymbol{u}) \tag{1}$$

$$\simeq \arg\max_{\boldsymbol{u} \in \boldsymbol{U}_d} \prod_{u \in \boldsymbol{u}} P(u) \tag{2}$$

$$\simeq \arg\max_{\boldsymbol{u} \in \boldsymbol{U}_d} \prod_{u \in \boldsymbol{u}} p_u \tag{3}$$

式 (1) から式 (2) において，アライメントの確率 $P(\boldsymbol{u})$ をアライメント $\boldsymbol{u}$ に含まれる対応付け u の確率 $P(u)$ の積と仮定している．式 (3) の p_u は対応付け u に関するパラメータである．式 (3) から，適切な p_u を得ることができれば，Viterbi アルゴリズムを用いて正確なアライメント

© 2013 Information Processing Society of Japan

を推定することができる．J●int multigram ではEMアルゴリズムを用いて p_u を最尤推定する．

1対1アライメントとは違い，多対多アライメントでは複数文字の対応付けが可能になるため各アライメントに含まれている対応付けの数は異なりうる．式(3)から分かるように，大きい単位で対応付けられた対応付けを多く含む（含んでいる対応付けの数が少ない）アライメントは小さい単位で対応付けられた対応付けを多く含むアライメントよりも乗算回数が少なく，$P(u) \leq 1$ であることから $\prod_{u \in \boldsymbol{u}} P(u)$ は高い値をとりやすい．そのため，J●int multigram は未知語に対する汎化能力が低い大きい単位での対応付けを優先する傾向にある．これに対処するために，J●int multigram では表記と発音それぞれにおいて対応付けの最大文字数を制限し，大きい単位での対応付けを防いでいる．

2.3 提案手法

J●int multigram では対応付けの最大文字数を固定するため，最大文字数を超えた表記と発音の対応関係に対し，不正確な対応付けを行ってしまう．また，このことを考慮して最大文字数を増やすと，汎化能力の低い大きい単位での対応付けが行われてしまう．

そこで，最大文字数を固定せずに大きい単位での対応付けを抑制するために，提案手法では対応付けに関するパラメータをその対応付けの表記の文字数と発音の文字数の和（表記と発音のマトリックス上での市街地距離）でべき乗することで，乗算数を表記の文字数と発音の文字数の和で統一する．提案手法は以下のように定式化できる．

$$\hat{\boldsymbol{u}} \simeq \arg\max_{\boldsymbol{u} \in \boldsymbol{U}_d} \prod_{u \in \boldsymbol{u}} P(u) \simeq \arg\max_{\boldsymbol{u} \in \boldsymbol{U}_d} \prod_{u \in \boldsymbol{u}} w_u^{s_u} \tag{4}$$

ここで，w_u は対応付け u に関するパラメータであり，s_u は表記と発音のマトリックス上での市街地距離を意味し，以下のように定義される．

$$s_u = i_u + j_u \tag{5}$$

ここで，i_u は対応付け u における表記の文字数，j_u は対応付け u における発音の文字数である．

3. 提案手法におけるパラメータの学習

提案手法ではJ●int multigramと同様にEMアルゴリズムを用いてパラメータ w_u を学習する．以下に提案手法のパラメータ w_u を学習するためのEMアルゴリズムを示す．

(1) w_u に初期値を設定（均一の値を設定）．

(2) 式(6)により，学習データにおける対応付け u の出現回数の期待値を計算（E-step）．

$$\gamma_u = \sum_{d \in \boldsymbol{D}} \sum_{\boldsymbol{u} \in \boldsymbol{U}_d} \frac{\prod_{u \in \boldsymbol{u}} w_u^{s_u}}{\sum_{\boldsymbol{u} \in \boldsymbol{U}_d} \prod_{u \in \boldsymbol{u}} w_u^{s_u}} n_u(\boldsymbol{u}) \tag{6}$$

$n_u(\boldsymbol{u})$ は $\boldsymbol{u}$ における u の出現数である．

(3) 式(7)により尤度を最大化（M-step）．

$$\hat{w}_u = \frac{\gamma_u}{\sum_{u \in \boldsymbol{U}} \gamma_u} \tag{7}$$

$\boldsymbol{U}$ は対応付け u の異なり集合である．

(4) $\hat{w}_u$ に w_u の値を代入．

(5) 収束条件を満たせば終了，満たさない場合は(2)へ戻る．

4. N-best Viterbi トレーニングを用いた削除文字推定の改善

多対多アライメントにおいて削除文字は通常の発音が対応付けられた対応付けと性質が異なる．発音が対応付けられた対応付けはその表記と発音が同時に対データに出現した際に3章のE-stepにおいてその出現回数の期待値が得られるのに対して，削除文字は対データにその表記が出現しただけでその出現回数の期待値が得られる．そのため，対応付けとして不正確な削除文字でも有利に学習され推定されやすくなる．

たとえば $D = \{\langle \text{many}, \textit{ménı} \rangle, \langle \text{ask}, \textit{æsk} \rangle\}$ を考えると，表記「a」は各データに存在するため，式(6)において削除文字 a/- の各データにおける出現回数の期待値が足し合わされる．一方で，正しい対応付けの a/*é* は $\langle \text{many}, \textit{ménı} \rangle$ にしか出現しないため，このデータにおける（a/*é* の）出現回数の期待値しか得られない．これにより削除文字 a/- のパラメータ w_u が，a/*é* の w_u よりも高い値となり，アライメントを推定する際に a/- が選ばれやすくなる．

J●int multigram では削除文字が出現すると式(2)の乗算回数が増加するため，そのアライメントの確率が小さくなり，それに合わせて，その削除文字の出現回数の期待値も小さくなって，削除文字が抑制される．一方で，提案手法では削除文字が出現しても乗算回数は同じため，削除文字を抑制できない．

この問題を改善するために，EMアルゴリズムのようにすべての可能なアライメントを考慮するのではなく，有望なアライメントだけを考慮するN-best Viterbi トレーニング[9]を学習に導入する．N-best Viterbi トレーニングは3章のEMアルゴリズムとほぼ同じ学習方法であるが，式(6)の $\boldsymbol{U}_d$（すべての可能なアライメントの集合）の代わりに，有望なN個のアライメントの集合（N-best のアライメント）を使用する．有望でないアライメントには不正確な削除文字が多く含まれるため，有望なアライメントのみを学習に用いることで，不正確な削除文字を抑制すること

© 2013 Information Processing Society of Japan

ができる．以下にN-best Viterbiトレーニングを導入したパラメータの学習方法を示す．

(1) 初期の段階では有望なアライメントを求められないため，市街地距離を導入した3章のEMアルゴリズムでパラメータを学習する．この際，削除文字を許可すると削除文字が有利に学習されるため削除文字は禁止する．

(2) 手順(1)で学習したパラメータを初期値とし，式(8)を基準にN-bestのアライメントを求めるN-best Viterbiトレーニングを行う．ただし，そのN-best Viterbiトレーニングを用いてパラメータを更新する回数は1回とする．式(8)は削除文字の対応付けのパラメータ w_u をそれ以外の対応付けのパラメータ w_u の加重相乗平均としている．各パラメータ w_u の重みはそれに対応する s_u である．

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} \prod_{u \in \boldsymbol{u}'} w_u^{s_u} \times \Big(\prod_{u \in \boldsymbol{u}'} w_u^{s_u} \Big)^{\frac{D_{\boldsymbol{u}}}{I_{\boldsymbol{u}} + J_{\boldsymbol{u}} - D_{\boldsymbol{u}}}} \quad (8)$$

ここで，$\boldsymbol{u}'$ は $\boldsymbol{u}$ から削除文字を取り除いた対応付けの系列，$D_{\boldsymbol{u}}$ は全削除文字の表記の総文字数，$I_{\boldsymbol{u}}$ は表記全体の文字数，$J_{\boldsymbol{u}}$ は発音全体の文字数である．

(3) 手順(2)で学習したパラメータを用いることで有望なアライメントを求めることができる．その学習したパラメータを初期値として，式(4)を基準にN-bestのアライメントを求めるN-best Viterbiトレーニングを行う．

5. 不正確な対応付けの結合

EMアルゴリズムは対データにおいて出現するすべての対応付けをカウントする．そのため，表記と発音の誤った対応付けも学習される．データセットによっては，誤った対応付けが学習されることにより熟字訓などの特殊な発音を持つ単語に対して不正確な対応付けが行われる可能性がある．

その例を**表1**に示す．⟨*AAA*, *tripléi*⟩の誤った対応付けである *AA/tripl* が含まれているアライメントには *A/éi* の対応付けが存在する．この *A/éi* の対応付けは他のデータのアライメント候補にも出現している．これにより，*A/éi* の対応付けの確率が学習により高い値をとり，⟨*AAA*, *tripléi*⟩のアライメントとして「*AA/tripl A/éi*」が推定される．結果として，*AA* に対して *tripl* が対応付けられた不正確な対応付けが得られる．

J●int multigramは大きい単位での対応付けが優先されるため，このような問題はほとんど起こらない．一方で，提案手法は市街地距離の導入により，詳細な対応付けを行うため，このような問題を避けることができない．

この問題を改善するために，*AA/tripl* などの不正確な対応付けを自動的に検出し，隣接する対応付けと結合することを提案する．不正確な対応付けの検出にはアライメント

表1 ⟨*AAA*, *tripléi*⟩と⟨*Ace*, *éis*⟩の場合における不正確な対応付けの例．この例の場合，不正確な対応付けは *AA/tripl* である

Table 1 Example of irrelevant mapping in the case of ⟨*AAA*, *tripléi*⟩ and ⟨*Ace*, *éis*⟩. In this case, the irrelevant mapping is *AA/tripl*.

⟨*AAA*, *tripléi*⟩ → {…, *AA/tripl A/éi*, …}
⟨*Ace*, *éis*⟩ → {…, *A/éi c/s e/ϵ*, …}

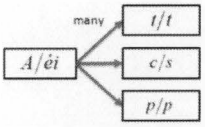


図2 正確な対応付けのコンテキスト

Fig. 2 Context of correct mapping.

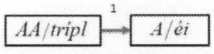


図3 不正確な対応付けのコンテキスト

Fig. 3 Context of irrelevant mapping.

推定後の対応付けが持つ前方または後方のコンテキストの種類数を用いる．**図2**にアライメントを行った際の正確な対応付けのコンテキストの様子を示す．図2のようにコンテキストを複数種類持つ対応付けは正確な対応付けであると考えられる．一方で，**図3**のようにコンテキストが1種類しかない（種類数が1の）対応付けは他にその対応付けに関する事例を持っていないため，不正確な対応付けである可能性が高い．このようなヒューリスティックな性質を用いることで *AA/tripl* などの不正確な対応付けを検出することができる．つまり，前方と後方のコンテキスト数を調べ，どちらかのコンテキストの種類数が1しかない対応付けを隣（種類数が少なかった方向）の対応付けと結合する．これにより，*AAA/tripléi* などの正確な対応付けを求めることができ，不正確な対応付けを避けることができる．

6. 評価実験

6.1 実験内容

1章で述べたように，括弧表現に基づくWebテキストマイニング[6]によるアプローチを用いて，提案手法の未知語に対する自動発音付与における有効性を評価する．**図4**に，括弧表現に基づくWebテキストマイニングによる自動発音付与の例を示す．提案手法と従来手法は工程(1)のアライメント処理に対応する．あらかじめ提案手法（または従来手法）により辞書データから抽出した表記と発音の対データをアライメントする．そして，アライメントされたデータから，ある表記がどのように発音されるかを定義した発音規則（アライメントデータの各対応付け）を獲得する（工程(2)）．

© 2013 Information Processing Society of Japan

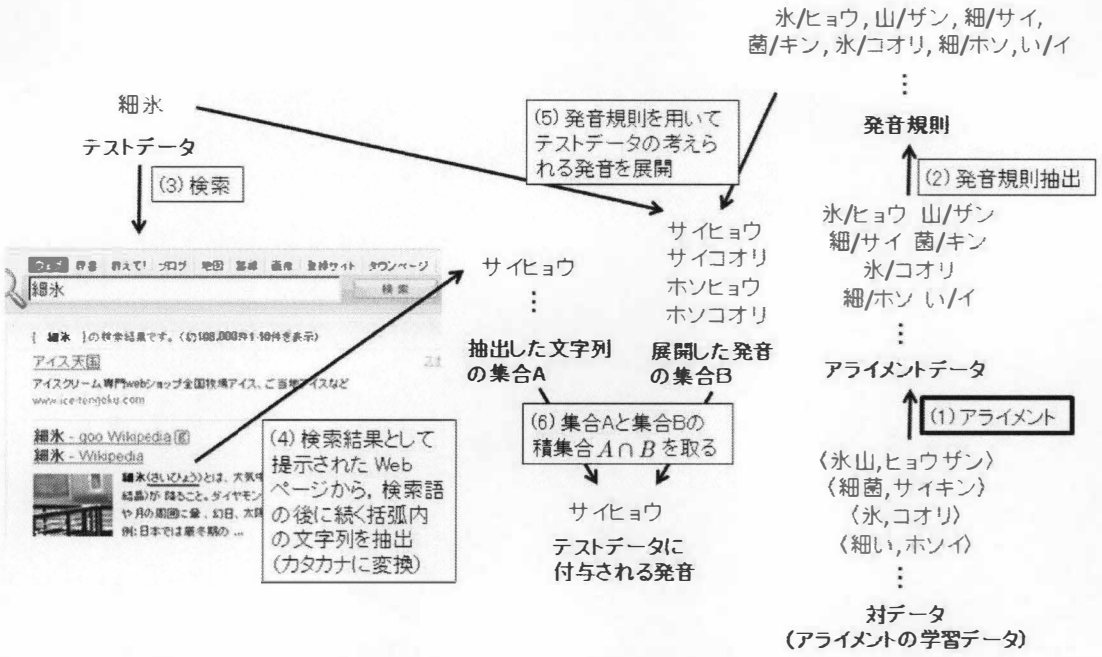


図 4 提案手法（または従来手法）は (1) アライメント処理に対応する，括弧表現に基づく Web テキストマイニングによる自動発音付与の流れ（「細氷」に対して発音を付与する）．Web 検索は goo 検索 [12] を使用．

Fig. 4 Flow of automatic pronunciation annotation using the proposed (or conventional) method and Web text mining based on the bracket representation (in the case of "細氷" as a test data). We used goo search [12] as Web search.

発音規則の獲得後，図 4 の工程 (3)，(4)，(5)，(6) を各テストデータに対して行う．工程 (3) ではテストデータを検索語として検索を行い，工程 (4) では検索結果として提示された Web ページから，検索語の後に続く括弧内の文字列を抽出し，抽出した文字列の集合 A を得る．これは日本語テキストにおいて，単語に続く括弧内にその読みを表記するという記法が用いられることを利用している．工程 (5) では獲得した発音規則を用いてテストデータの表記から考えられる発音をすべて展開し，展開した発音の集合 B を得る．工程 (6) では抽出した文字列の集合 A と展開した発音の集合 B の積集合 $A \cap B$ の中から，Web ページから抽出された頻度が最も高い発音を 1 つ選択し，テストデータの発音として付与する．つまり，本実験においては先行研究 [8], [9], [10], [11] と同様に異音同義語を考慮しない．もし，$A \cap B = \emptyset$ なら，誤った発音を付与する可能性が高いため，テストデータには発音を付与しない（抽出した文字列をすべて棄却する）．誤った発音が付与される可能性が高いデータに対して，自動で発音を付与しないことは実用的な利点がある．たとえば音声認識の認識辞書構築時に，実際に発音が誤っているデータを語彙として登録すると認識に悪影響を与える可能性がある．もし，あらかじめ誤った発音が付与される可能性が高いデータを分かっていれば，認識に悪影響を与えないよう，そのデータに対して人手で発音を付与して認識辞書に登録するか，人手で発音を付与せずに認識辞書に登録しないか選択が可能になる．そのため，$A \cap B = \emptyset$ の場合，発音を付与しない．これらの工程により，$A \cap B = \emptyset$ の場合を除いて，テストデータに対して自動で発音を付与する．

本実験では評価尺度として再現率と精度，F 値を用いる．文献 [6] では *generality*（一般性，正解したテストデータ数割る全体のテストデータ数）を用いているが，これは図 4 の工程 (4) のテキストマイニング処理において，どれだけ正解の発音を抽出できるかに依存する．我々が提案する手法はアライメント手法であり，Web テキストマイニングを用いた自動発音付与手法ではないため，アライメント手法の評価としてはテキストマイニングの性能に依存しない評価尺度を用いるべきだと考えられる．そのため，テキストマイニングの性能に依存しないよう，テキストマイニングで正解の発音を抽出できなかったテストデータは考慮せずに，正解の発音を抽出できたテストデータのみで評価する再現率を定義し，それを *generality* の代わりとして評価に用いる．各評価尺度は以下のように定義される．

$$\text{再現率} = \frac{R}{C} \tag{9}$$

$$\text{精度} = \frac{R}{N} \tag{10}$$

$$\text{F 値} = \frac{R}{0.5(N+C)} \tag{11}$$

ここで，R は正解したテストデータ数（正解の発音と自動

© 2013 Information Processing Society of Japan

付与した発音が一致するテストデータ数)，C は工程 (4) におけるテキストマイニングにより抽出された文字列の集合 A の中に，正解の発音が含まれているテストデータの数である．いい換えると，C はアライメント手法で改善することができるテストデータの数の上限値である．N は「(抽出した文字列の集合 A)∩(展開した発音の集合 B)$\neq\emptyset$ であるテストデータ」の数である．いい換えると，N は図 4 の「Web テキストマイニングによる自動発音付与」システムが発音を付与したテストデータの数であり，正解のテストデータ数 R をこれで割ることにより求まる精度はこのシステムの発音付与の正確さを表す指標である．

アライメントの観点から見ると，本実験の再現率は未知語に対する表現能力（汎化能力）の高さを表す．精度は不正確な対応付けが少ないことを表す．一般的に，汎化能力を上げるために小さい単位で対応付けを行うと不正確な対応付けが多くなり，正確な対応付けを行うために大きい単位で対応付けを行うと汎化能力が下がる．そのため，この 2 つの指標はトレードオフの関係にある．

評価する手法は以下のとおりである．

(1) ベースライン 1 (Web_freq)：工程 (6) の代わりに，Web から抽出した文字列の集合 A の中で最も抽出された頻度が高い文字列を発音として付与する方法
(2) ベースライン 2 [6] (without_align)：工程 (1) においてアライメントを行わずに単語の表記と発音を直接発音規則として用いる方法
(3) 従来手法 [9], [10] (joint)：工程 (1) において Joint multigram を使用
(4) 提案手法 1 (city)：工程 (1) において市街地距離を導入した多対多アライメントを使用
(5) 提案手法 2 (city+del)：工程 (1) において city に 4 章の「N-best Viterbi トレーニングを用いた削除文字推定の改善」を導入した多対多アライメントを使用
(6) 提案手法 3 (city+merge)：工程 (1) において city に 5 章の「不正確な対応付けの結合」を導入した多対多アライメントを使用
(7) 提案手法 4 (city+del+merge)：工程 (1) において city+del に 5 章の「不正確な対応付けの結合」を導入した多対多アライメントを使用

6.2 実験条件

実験条件を以下に示す．1 つの対応付けにおける表記と発音の最大文字数は without_align，city，city+del，city+merge，city+del+merge に関しては無制限，joint に関しては表記 1，2，3 文字と発音 3，6 文字の組合せ（計 6 組）を試行した．各手法とも削除文字の出現は許可し，挿入文字の出現は禁止した．また，図 4 の工程 (5) の発音の展開時において削除文字の発音規則が連続して選ばれるような展開は禁止した．city+del，city+del+merge における N-best Viterbi トレーニングでは，対応付けとして不正確な削除文字が学習されることを極力避けるために，2-best のアライメントを用いて学習を行った．Web 検索には goo 検索 [12] を使用し，マイニングする Web ページ数は 1 キーワードにつき 500 ページとした．

実験データは以下のとおりである．辞書データは単漢字辞書（Wnn [13] と三省堂 [14]），NAIST Japanese Dictionary [15]，英辞郎 [16] を用いた．辞書データから抽出した表記と発音の対データは合計で約 35 万個であった．テストデータは，音声ドキュメント検索や音声検索のタスクを意識し，固有名詞や新語が多く含まれている最新の検索キーワードを使用した．テストデータは以下のように構築した．まず，Yahoo! 検索ランキング [3]，Google 急上昇ワード [17]，Goo キーワードランキング [18] から収集した使用頻度の高い最新の検索キーワードの中から，重複を除いたのち，ランダムに 6,439 個の検索キーワードを抽出した．次にそれらのキーワードに対して，人手で発音の揺らぎも考慮し，1 つのキーワードにつき 1 つ以上の発音を正解の発音として付与した．そして，各キーワードに対して形態素解析器 Kytea [4] を用いて発音を推定し，人手で付与した正解の発音と一致しなかった（正解の発音が得られなかった）1,995 個のキーワードを未知語のテストデータとして用いた．辞書データとテストデータでは，表記が漢字，平仮名，カタカナ，英字，記号で構成され，発音はカタカナのみで構成されている．また，構築したテストデータの C の値（Web から正解の発音を抽出できたテストデータ数）は 1,403 であった．

6.3 実験結果

表 2 に Joint multigram の表記と発音の最大文字数の違いによる評価実験の結果を示す．各再現率と精度に関して，それらの値を求める際に用いた R，C，N の値も示している．joint N-M は最大文字数を表記 N 文字，発音 M 文字に制限した Joint multigram を意味する．表 2 から Joint multigram は表記と発音の最大文字数の違いによって性能が大きく変化することが分かる．joint 1-6 の場合，他の設定と比べ最も再現率が高かった．これは表記 1 文字という小さい単位でのアライメントを行うことにより，未知語に対する高い汎化能力が得られたからである．一方で，精度は最も低い値を示した．これは小さい単位でアライメントすることで不正確な対応付けが多くなったからである．逆に最大文字数を表記 3 文字，発音 6 文字と大きい単位でアライメントした場合 (joint 3-6)，不正確な対応付けが少なくなり精度は向上するものの，汎化能力が低下して再現率は劣化している．F 値が最も高かったのは joint 2-3 で，この結果を他の手法との比較に用いる．

表 3 に全手法の評価実験の結果を示す．表 3 から，city と city+del を比べると，再現率を約 1.0 ポイント，精度を

© 2013 Information Processing Society of Japan

表 2 従来手法 Joint multigram の表記と発音の最大文字数の違いによる評価実験の結果．joint N-M は最大文字数を表記 N 文字，発音 M 文字に制限した Joint multigram を意味する

Table 2 The result of the evaluation experiment due to differences in the maximum number of characters of notation and pronunciation for joint multigram approach of conventional method. The joint N-M denotes joint multigram approach with the maximum lengths of N graphemes and M phonemes.

	再現率 (%) (R/C)	精度 (%) (R/N)	F 値 (%)
joint 1-3	87.95 (1,234/1,403)	93.27 (1,234/1,323)	90.54
joint 1-6	**88.38** (1,240/1,403)	92.81 (1,240/1,336)	90.54
joint 2-3	86.46 (1,213/1,403)	95.81 (1,213/1,266)	**90.90**
joint 2-6	85.46 (1,199/1,403)	95.54 (1,199/1,255)	90.22
joint 3-3	85.10 (1,194/1,403)	95.83 (1,194/1,246)	90.15
joint 3-6	83.96 (1,178/1,403)	**95.93** (1,178/1,228)	89.55

表 3 全手法の評価実験の結果

Table 3 The result of the evaluation experiment for all evaluated methods.

		再現率 (%) (R/C)	精度 (%) (R/N)	F 値 (%)
ベース	Web_freq	59.73 (838/1,403)	45.64 (838/1,836)	51.74
ライン	without_align	76.19 (1,069/1,403)	**95.87** (1,069/1,115)	84.91
従来手法	joint 2-3	86.46 (1,213/1,403)	95.81 (1,213/1,266)	90.90
提案手法	city	89.24 (1,252/1,403)	93.43 (1,252/1,340)	91.29
	city+del	90.24 (1,266/1,403)	95.12 (1,266/1,331)	92.61
	city+merge	88.81 (1,246/1,403)	94.82 (1,246/1,314)	91.72
	city+del+merge	**90.31** (1,267/1,403)	95.33 (1,267/1,329)	**92.75**

約 1.7 ポイント改善したことが分かる．有意水準 5%の t 検定において，再現率の改善には有意な差があるとはいえないものの，精度の改善は有意であった．また，F 値で見ると約 1.32 ポイントの改善であった．city と city+merge は F 値で見ると約 0.43 ポイントの改善であったが，再現率と精度は有意水準 5%の t 検定において有意な差があるとはいえなかった．city+del と city+del+merge もまた，F 値で見ると約 0.14 ポイントの改善であったが，再現率と精度は有意水準 5%の t 検定において有意な差があるとはいえなかった．city+del+merge は joint 2-3 と比べて再現率を約 3.9 ポイント改善した．これは有意水準 1%の t 検定においても有意な差がみられた．一方で精度は約 0.5 ポイント劣化したが，これは有意水準 5%の t 検定において有意な差があるとはいえなかった．また，F 値で見ると約 1.85 ポイントの改善であった．以上の結果から，拡張された提案手法は Joint multigram が持つ精度をほとんど劣化させずに，再現率を大きく改善することが分かった．

6.4 考察

6.4.1 提案手法 city+del+merge と従来手法 joint の比較

テストデータにおける city+del+merge と joint 1-6 の発音付与の違いを比較すると，「tacica（タシカ）」の場合，city+del+merge は正解の発音を付与したのに対し，joint 1-6 は「ニコニコ」を付与した．この理由は joint 1-6 において「t/-」，「a/ニ」，「c/コ」，「i/ニ」，「a/-」の発音規則が存在したからである．「a/ニ」に注目すると，辞書データ内の「mechanic（メカニック）」に対して，「m/メ e/- c/カ h/- a/ニ n/- i/ック c/-」とアライメントした結果から得られており，これは明らかに不正確な対応付けである．joint 1-6 においてこのような対応付けになったのは，表記を 1 文字に制限したことにより，アライメント候補の中で発音が割り当てられない表記の割合が増え，削除文字の影響が強くなり，n/- が n/ニ や a/- よりも高いパラメータ値を得たからである．このことから，最大文字数を小さく設定することで不正確な対応付けが多くなることが分かる．

また，city+del+merge と joint 3-6 の違いを比較すると，「熊谷直実（クマガイナオザネ）」に対して，city+del+merge は正解の発音を付与したのに対し，joint 3-6 は発音を付与しなかった．この理由は joint 3-6 は大きい単位で対応付けを行うため，小さい単位での対応付けである「実/ザネ」を発音規則として得られなかったからである．このことから，従来手法は最大文字数を小さくすると不正確な対応付けが増えて精度が劣化し，大きくすると未知語に対する汎化能力が劣化することが分かる．一方で提案手法は最大文字数の適切な設定を探索する必要がなく，これは提案手法の利点の 1 つとしてあげられる．

© 2013 Information Processing Society of Japan

6.4.2 提案手法 city と提案手法 city+del の比較

city と city+del の性能を比べると，city+del が精度において約 1.7 ポイントの有意な改善を示した．この改善の要因は，N-best Viterbi トレーニングにより，不正確な削除文字の学習が抑制されたためである．たとえば，テストデータの「nakata.net（ナカタドットネット）」に対して，city+del は発音を付与しなかったのに対し，city は「レベルファイブ」を付与した．この理由を調査すると，city において「*n*/レ」，「*a*/ベ」，「*k*/-」，「*a*/ル」，「*t*/フ」，「*a*/ァ」，「./イ」，「*n*/-」，「*e*/ブ」，「*t*/-」の発音規則が存在していることが分かった．「*a*/ベ」に注目すると，辞書データ内の「deathbed（デスベッド）」に対して，city は「*d*/デ *e*/ス *a*/ベ *th*/- *b*/ッ *e*/- *d*/ド」とアライメントしており，これは明らかに不正確な対応付けである．city においてこのような対応付けになったのは，すべてのアライメントを学習に用いることで，削除文字の影響が強くなり，*th*/- や *e*/- が他よりも高いパラメータ値を得たからである．一方で，city+del は「deathbed（デスベッド）」を「*d*/デ *ea*/- *th*/ス *b*/ベ *e*/ッ *d*/ド」とアライメントしている．これは有望なアライメントのみ学習に用いることで，不正確な削除文字の影響を抑えたからだと考えられる．このことから，「N-best Viterbi トレーニングを用いた削除文字推定の改善」により精度が改善されることが分かる．

6.4.3 提案手法 city+del と提案手法 city+merge の相補性

city+del と city+merge を比べると，「不正確な対応付けの結合」よりも「N-best Viterbi トレーニングを用いた削除文字推定の改善」の方が F 値を改善する効果が高いことが分かる．これは不正確な削除文字が自動発音付与の性能に大きな影響を及ぼしており，「不正確な対応付けの結合」ではこの問題に対処できないからである．一方で，city+del は特殊な発音を持つ単語（熟字訓など）に対して不正確な対応付けを行っていたのに対し，city+merge はそれらに対して正確な対応付けを行っていた．たとえば辞書データにある「山葵（ワサビ）」のアライメントにおいて，city+del では「山/ワ 葵/サビ」と対応付けたのに対し，city+merge は「山葵/ワサビ」と正確に対応付けた．これらのことから，この 2 つの手法は相補的な関係にあり，この 2 つの手法を導入した city+del+merge は F 値において最も高い性能を示す結果となった．

6.4.4 総合的な発音付与性能

city+del+merge はテストデータの 1,995 個のキーワードのうち，1,267 個のキーワードに対して正解の発音を付与した．本実験ではあらかじめ人手で正解の発音を付与しているため，「形態素解析器により付与された発音が誤っているデータ」を自動的に得ることが可能であるが，実際にはこのデータを得ることは難しい．もしこのデータを得ることができた場合は，提案手法の city+del+merge を適用した Web テキストマイニングによる自動発音付与を形態素解析器と併用することで，Yahoo!検索ランキングなどから抽出した 6,439 個のキーワードの約 88.7%に対して正解の発音を付与できることになる．形態素解析器だけの場合，約 69%しか正解の発音を付与できないため，形態素解析器と提案手法の city+del+merge を適用した Web テキストマイニングによる自動発音付与を併用することで，音声ドキュメントの認識性能などを改善することが期待できる．また，従来手法の joint 2-3 ではその値が約 87.9%であり，提案手法の city+del+merge との差は約 0.84 ポイントである．これは有意水準 7%の t 検定において有意な差である．音声ドキュメント検索などでこの差がどの程度になるかは，音声ドキュメントの種類，音声ドキュメントシステムの構成，評価指標によって異なるが，Web 上の音声ドキュメント検索に関しては新語が継続的に出現することが想定されるため，提案手法を継続的に使用することにより，音声ドキュメント検索における提案手法と従来手法の性能差が漸次的に拡大していくと考えられる．「形態素解析器で付与した発音が誤っているデータ」の検出も含め，形態素解析器と Web テキストマイニングによる自動発音付与の併用に関しては今後の課題である．

また，本実験においては，先行研究 [8], [9], [10], [11] と同様に，異音同義語を考慮しなかった．異音同義語を考慮する場合，自動発音付与システムによって与えられた複数の有望な候補（統計的アプローチでは N-best の出力結果，Web テキストマイニングによるアプローチでは「(抽出した文字列の集合 A)∩(展開した発音の集合 B)」）に対して，それが正解の発音か誤った発音かを識別器によりさらに判定する必要がある．このような異音同義語を扱うことができる自動発音付与システムもまた，形態素解析器と Web テキストマイニングによる自動発音付与の併用と同様に今後の課題である．

7. まとめ

日本語の未知語に対する発音付与の性能向上を目的として，表記と発音の対応付けの最大文字数を制限せずに，汎化能力が高い小さい単位での対応付けを求める多対多アライメントを提案した．これまでに提案していた，市街地距離を導入した多対多アライメントでは削除文字が推定されやすく，またデータセットによっては熟字訓などの特殊な発音を持つ単語に対して不正確な対応付けを行うという問題を持っていた．本論文では前者の問題に対してアライメントの学習時に N-best Viterbi トレーニングを用いて解決する方法を提案し，後者の問題に対しては不正確な対応付けを隣の対応付けと結合させることで正確な対応付けにする方法を提案した．また，Web テキストマイニングを用いた日本語の未知語に対する自動発音付与により拡張された提案手法の評価を行った．評価実験の結果，拡張された提

© 2013 Information Processing Society of Japan

案手法は従来手法の Joint multigram が持つ精度をほとんど劣化させずに，未知語に対する汎化能力を表す再現率を約 3.9 ポイント改善した．

謝辞 本研究の一部は，戦略的創造研究推進事業「共生社会に向けた人間調和型情報技術の構築」（JST/CREST）などの援助を受けて行われた．

参考文献

[1] Ogata, J., Goto, M. and Eto, K.: Automatic Transcription for a Web 2.0 Service to Search Podcasts, *Proc. Interspeech 2007*, pp.2617–2620 (2007).

[2] グーグル株式会社：Android/iPhone 向け音声検索，グーグル株式会社（オンライン），入手先 ⟨http://www.google.co.jp/mobile/default/onsei.html⟩（参照 2012-04-27）.

[3] ヤフー株式会社：Yahoo!検索ランキング，ヤフー株式会社（オンライン），入手先 ⟨http://searchranking.yahoo.co.jp/⟩（参照 2012-04-27）.

[4] Neubig, G. and Mori, S.: Word-based Partial Annotation for Efficient Corpus Construction, *Proc. LREC 2010*, pp.2723–2727 (2010).

[5] Jiampojamarn, S., Cherry, C. and Kondrak, G.: Integrating joint n-gram features into a discriminative training framework, *Proc. NAACL-HLT 2010*, pp.697–700 (2010).

[6] Miyake, J., Takeuchi, S., Kawanami, H., Saruwatari, H. and Shikano, K.: Automatic Reading Annotation to Japanese Trendy Words based on Parentheses Expression, *Proc. Oriental COCOSDA 2008*, pp.81–86 (2008).

[7] Dempster, A., Laird, N. and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Vol.Series B, pp.1–38 (1977).

[8] Damper, R.I., Marchand, Y., Marsters, J.D. and Bazin, A.I.: Aligning text and phonemes for speech technology applications using an EM-like algorithm, *Journal of speech Technology*, Vol.8, No.2, pp.147–160 (2005).

[9] Deligne, S., Yvon, F. and Bimbot, F.: Variable-Length Sequence Matching for Phonetic Transcription Using Joint Multigrams, *Proc. EUROSPEECH*, pp.2243–2246 (1995).

[10] Jiampojamarn, S., Kondrak, G. and Sherif, T.: Applying Many-to-Many Alignments and Hidden Markov Models to Letter-to-Phoneme Conversion, *Proc. NAACL HLT 2007*, pp.372–379 (2007).

[11] Kubo, K., Kawanami, H., Saruwatari, H. and Shikano, K.: Unconstrained Many-to-Many Alignment for Automatic Pronunciation Annotation, *Proc. APSIPA 2011* (2011).

[12] エヌ・ティ・ティレゾナント株式会社：goo 検索，エヌ・ティ・ティ レゾナント株式会社（オンライン），入手先 ⟨http://search.goo.ne.jp/⟩（参照 2012-04-27）.

[13] Aono, T.: FreeWnn @ SourceForge.jp Web ページ，FreeWnn プロジェクト（オンライン），入手先 ⟨http://freewnn.sourceforge.jp/⟩（参照 2012-04-27）.

[14] 株式会社三省堂：三省堂★SANSEIDO，株式会社三省堂（オンライン），入手先 ⟨http://www.sanseido-publ.co.jp/⟩（参照 2012-04-27）.

[15] Yamane, H. and Asahara, M.: NAIST Japanese Dictionary プロジェクト日本語トップページ - SourceForge.JP，奈良先端科学技術大学院大学（オンライン），入手先 ⟨http://sourceforge.jp/projects/naist-jdic/⟩（参照 2012-04-27）.

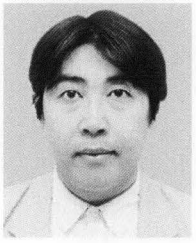
[16] EDP：英辞郎（えいじろう・EIJIRO）の最新情報，EDP（オンライン），入手先 ⟨http://www.eijiro.jp⟩（参照 2012-04-27）.

[17] グーグル株式会社：急上昇ワード，グーグル株式会社（オンライン），入手先 ⟨http://www.google.co.jp/m/trends⟩（参照 2012-04-27）.

[18] エヌ・ティ・ティレゾナント株式会社：検索ランキング ランキング - goo ランキング，エヌ・ティ・ティ レゾナント株式会社（オンライン），入手先 ⟨http://ranking.goo.ne.jp/keyword/⟩（参照 2012-04-27）.

久保 慶伍

平成 21 年近畿大学理工学部情報工学科卒業．平成 23 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．同年同大学院博士後期課程入学，現在に至る．音声認識のための自動発音付与の研究に従事．日本音響学会学生会員．



川波 弘道

平成 6 年東京大学工学部電気工学科卒業．平成 12 年東京大学大学院工学系研究科博士課程修了．博士（工学）．同年電子技術総合研究所入所．平成 13 年より奈良先端科学技術大学院大学情報科学研究科助手，平成 19 年同助教．現在，音声分析，音声対話の研究に従事．電子情報通信学会，日本音響学会各会員．

猿渡 洋

平成 3 年名古屋大学工学部電気工学科卒業．平成 5 年同大学大学院修士課程修了．平成 12 年同大学院博士課程修了．工学博士．平成 5 年セコム（株）入社．セコム IS 研究所音声情報処理研究室において，音響アレー信号処理に関する研究に従事．平成 12 年奈良科学技術大学院大学助教授．平成 19 年同准教授．音声信号処理，統計的信号処理等に関する研究に従事．平成 13，18 年電子情報通信学会論文賞受賞．平成 15，20，23 年電気通信普及財団テレコムシステム技術論文賞受賞．平成 23 年ドコモ・モバイル・サイエンス賞基礎科学部門優秀賞受賞．日本音響学会，日本 VR 学会，IEEE 各会員．

© 2013 Information Processing Society of Japan

鹿野 清宏 （フェロー）

昭和 45 年名古屋大学工学部電気工学科卒業．昭和 47 年同大学大学院修士課程修了．同年電電公社武蔵野電気通信研究所入所．昭和 59～61 年カーネギーメロン大客員研究員．昭和 61～平成 2 年 ATR 自動翻訳電話研究所音声情報処理研究室長．平成 4 年 NTT ヒューマンインタフェース研究所主席研究員．平成 6 年より奈良先端科学技術大学院大学情報科学研究科教授．音情報処理学講座を担当．工学博士．音声・音情報処理の研究および研究指導に従事．昭和 50 年電子通信学会米沢賞，平成 3 年 IEEE SP 1990 Senior Award，平成 6 年日本音響学会技術開発賞，平成 12 年情報処理学会山下記念研究賞，平成 13 年 VR 学会論文賞，平成 17，18 年電子情報通信学会論文賞，平成 17 年猪瀬賞．電子情報通信学会フェロー，IEEE フェロー，ISCA，音響学会各会員．

© 2013 Information Processing Society of Japan