J-018

聞き返しに対する誤聴箇所の推定
Mishearing part estimation corresponding to asking back

大谷 優果† 篠山 学‡
Yuka Otani Manabu Sasayama

1. はじめに

コミュニケーションロボットは医療機関や企業などで利用者や来訪者への対応を目的に導入され，急速に普及しつつある．しかし，これらのロボットが人間とスムーズな対話を行えているとは言い難い．その原因のひとつがコミュニケーションの断絶である．コミュニケーションの断絶とは，聞き間違いや言い間違いによって会話が成り立たなくなってしまうことである．ロボットと人間の対話がスムーズに行われるためには，コミュニケーションの断絶を防止する必要がある． 本研究ではコミュニケーションの断絶の中でも，聞き間違いとそれによる聞き返しに注目した．

例えば「ショートケーキ買って」という発話に対し「消毒液？」と返答された場合を取り上げる．人間の発話者は，相手の返答が聞き間違いによる聞き返しであると認識することができる．これは発話者が，相手に「ショートケーキ」を「消毒液」を聞き間違えられたことを理解できるためである．このように人間の発話者は，相手が聞き間違えた箇所を特定した上で，相手の返答が聞き返しであることを認識する．しかし，コミュニケーションロボットにはこれらのことができない．

そこで本研究では，発話を聞き間違った相手による聞き返しから，聞き間違えられた箇所を推定する手法を提案する．この推定が行えることで，聞き返しが発話を聞き間違えたことによるものであると判定できる．ここでの発話者はロボット，聞き間違えた相手は人間とする． 本論文では，相手の発話を別の文や単語に聞き間違えることを誤聴と呼ぶ．また，誤聴した文や単語を相手に再度発話してもらうことを期待する発話を聞き返しと呼ぶ．

2. 関連研究

関連研究では，発話者の質問に曖昧性が含まれていることを認識し，曖昧性を除去するための「聞き返し文」を生成する研究[1][2][3]がある．これらは人間の発話者と，人間またはロボットの質問者の会話を想定している．そのため，ロボットの発話者を想定している本研究とは根本的に異なる．

音声認識の誤認識を訂正する研究[4][5]では，音韻と意味の両面から訂正の必要性と訂正候補の妥当性を判断している．ここで訂正の対象となっている誤認識は意味の通らない発話のみであり，本研究で対象とする意味の通る発話は除外されている．

音声対話に関する研究[6]では，ユーザの聞き返しに対応して音量調節を行う手法が提案されている．この研究での聞き返しは「何？」や「え？」など音声が聞き取れなかった場合を対象としている．本研究では聞き返しの中でも，発話の中に対応する誤聴箇所をもつものを対象とする．

† 香川高等専門学校　情報工学科本科
‡ 香川高等専門学校　情報工学科

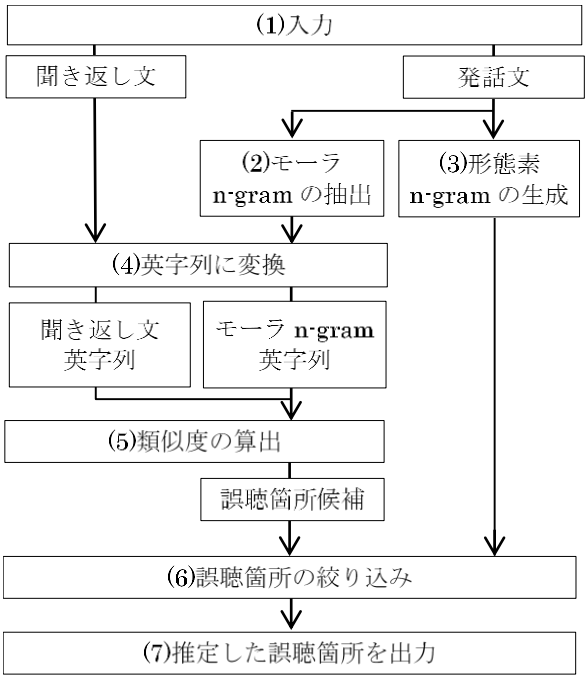

図 1　誤聴箇所の推定の流れ

また本研究では，発話文と聞き返し文をモーラ単位でとらえた．これは誤聴の原因を考察した研究[7]で，そのひとつにモーラのすり替えが挙げられているためである．モーラは音の文節単位の一つで拍とも呼ばれ，かな文字として意味を持つ音韻論上の最小単位である．拗音，促音，長音子はそれぞれ 1 モーラとみなす．本論文では，発話文と聞き返し文のモーラ単位でのレーベンシュタイン距離を用いた類似度から誤聴箇所を推定する手法を提案する．

3. レーベンシュタイン距離

レーベンシュタイン距離は，二つの文字列がどの程度異なっているかを示す距離の一つである．文字列 A と文字列 B のレーベンシュタイン距離は，文字列 B を文字列 A に変形するために必要な操作数の累加の最小値で定義される．一文字の挿入，削除，置換はそれぞれ操作数 1 とする．

本論文では二つの英字列の類似度と二つの文字列の類似度を求めるときに用いる．二つの英字列のレーベンシュタイン距離は次のように計算する．例えば「saqka-」（サッカー）と「saka」（坂）の 2 つのモーラ英字列を考える．「saka」は q と-を挿入する操作によって「saqka-」に変形できる．このことから，「saqka-」と「saka」のレーベンシュタイン距離は 2 である．

4. 提案手法

4.1 提案手法の流れ

提案する誤聴箇所の推定の流れを図 1 に示す.

(1) 発話文と聞き返し文の組を入力する．例えば，発話文「5 名様ご案内」と聞き返し文「お姉様」の組を入力する．本論文では，コミュニケーションロボットの発話者を想定している．そのため発話文には漢字かな交じり文に加えて，音声合成のために各単語の読み仮名や音韻記号などが存在している．提案手法では漢字かな交じり文と読み仮名を利用する．また聞き返し文は音声認識で得ることを想定し，文字の読み仮名とする．ここでの読み仮名はカタカナとする．またこのとき，ロボット側の音声認識は正確に行えているものと仮定する.

(2) 発話文からモーラ n-gram を抽出する．ここでの n は，聞き返し文のモーラ数 m に対して$m-2 \leq n \leq m+2$と定義する．モーラ n-gram を用いるのは，誤聴の文または単語の形態素区切りと発話文の形態素区切りが異なる場合を考慮するためである.

(3) 発話文の形態素 n-gram を生成する．これを用いて誤聴箇所の絞り込みの処理を行う．形態素 n-gram は，カタカナの形態素を各要素とした配列である．これは漢字かな交じり文を形態素解析し，読み仮名を抽出することで生成する．形態素解析器には MeCab[*1] を用いる．ここでの n は，発話文の形態素数 N に対して 1 から N の全ての整数とする． 「5 名様ご案内」の形態素 n-gram の生成を，図 2 に示す.

(4) 発話文のモーラ n-gram と聞き返し文をそれぞれ英字列に変換する．英字列は表 1 の定義に従って生成する．シやツは h や s を入れずに，si や tu と表記する．ザ行には z を用いる．拗音は表 1 に示すように，大文字と小文字に分けて定義する．例えばニュは「ニ」を大文字，「ュ」を小文字とし，「nilyu」と表記する.

(5) 全てのモーラ n-gram において，聞き返し文との類似度を算出する．類似度はレーベンシュタイン距離の逆数とする．また，類似度が最大となるモーラ n-gram を誤聴箇所候補と定義する．例えば「5 名様」(gomeisama)と「お姉様」(oneesama)の類似度は，レーベンシュタイン距離が 3 であることから 0.33 となる.

(6) (5)で得た誤聴箇所候補に対して絞り込みを行うため，誤聴箇所候補と発話文の形態素 n-gram を比較する．絞り込みによって候補を 1 つに絞り込めた場合，これを推定した誤聴箇所とする．候補が 2 つ以上残った場合は，推定が失敗したとする．詳しくは 4.2 節で説明する．絞り込みは誤聴箇所の補正の役割も持つため，候補が 1 つの場合にも実行する.

(7) 推定が行えた場合，誤聴箇所を出力する.

*1 MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/

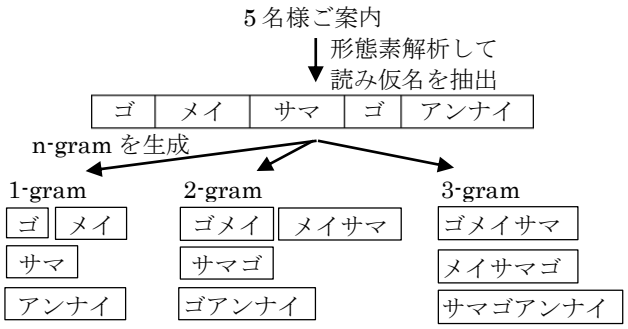

図 2 形態素 n-gram の生成

表 1 英字列の表記の定義

清音，濁音，半濁音	ローマ字による英字表記
拗音	（大文字）+l+（小文字）
促音	q
長音子	-（半角ハイフン）

4.2 誤聴箇所の絞り込み

図 1 の(6)誤聴箇所の絞り込みでは，誤聴箇所候補と発話文の形態素 n-gram を比較する．形態素 n-gram を用いるのは，発話文中の誤聴箇所を明確にするためである．以下の優先順位に従って，誤聴箇所候補を 1 つに絞り込む．まず，誤聴箇所候補と完全一致する形態素 n-gram を探す．これを全ての誤聴箇所候補に対して行い，完全一致した中でモーラ数が最大の誤聴箇所候補を誤聴箇所とする．モーラ数が最大の誤聴箇所候補が複数存在した場合は，推定が失敗したとする．もし完全一致する誤聴箇所候補が存在しなかった場合は，誤聴箇所候補と形態素 n-gram との類似度を計算する．この類似度が最大の形態素 n-gram を誤聴箇所とする．類似度が最大の形態素 n-gram が複数存在した場合は，推定が失敗したとする.

4.3 提案手法による誤聴箇所の推定

発話文が「5 名様ご案内」と聞き返し文が「お姉様」の場合を例に推定の流れを説明する.

(1) 発話文の漢字かな交じり文「5 名様案内」と読み仮名「ゴメイサマゴアンナイ」，聞き返し文「オネエサマ」が入力される.

(2) 聞き返し文のモーラ数が 5 であることより，発話文のモーラ 3-gram からモーラ 7-gram を抽出し，合計 30 のモーラ n-gram を得る.
   A) モーラ 3-gram：ゴメイ，メイサを含む 8 つ
   B) モーラ 4-gram：ゴメイサを含む 7 つ
   C) モーラ 5-gram：ゴメイサマを含む 6 つ
   D) モーラ 6-gram：ゴメイサマゴを含む 5 つ
   E) モーラ 7-gram：ゴメイサマゴアを含む 4 つ

(3) 発話文の漢字かな交じり文「5 名様ご案内」から形態素 n-gram を生成し，合計 14 の形態素 n-gram を得る.
   A) 形態素 1-gram
      ゴ，メイ，サマ，アンナイ
   B) 形態素 2-gram

ゴメイ，メイサマ，サマゴ，ゴアンナイ
C) 形態素 3-gram
ゴメイサマ，メイサマゴ，サマゴアンナイ
D) 形態素 4-gram
ゴメイサマゴ，メイサマゴアンナイ
E) 形態素 5-gram
ゴメイサマゴアンナイ

(4) (2)で得た合計 30 のモーラ n-gram と聞き返し文を英字列に変換する．

(5) 全てのモーラ n-gram において聞き返し文との類似度を算出する．最大の類似度 0.33（レーベンシュタイン距離 3 の逆数）を示す 2 つの n-gram「ゴメイサマ」と「メイサマ」を誤聴箇所候補とする．

(6) 誤聴箇所の絞り込みを行う．「ゴメイサマ」と「メイサマ」は，両方が形態素 n-gram と完全一致する．よって，モーラ数が最大値の 5 を示す誤聴箇所候補「ゴメイサマ」が誤聴箇所であると推定する．

(7) 推定の結果である「ゴメイサマ」を出力する．

5. 誤聴データの収集

提案手法の評価に用いる発話文と聞き返し文として，Web サイト*2 への聞き間違いに関する投稿記事から誤聴を含む会話（誤聴データ）を 1127 組収集した．誤聴データは，誤聴箇所を含む発話文と誤聴箇所，聞き返し文の三つで構成された二行一組のタブ区切りテキストとする．誤聴箇所は投稿記事に明示されている．例えば「お父さんが“芋洗い行ってきた”っていうからびっくりしたけど，エムアールアイだった」といった投稿記事からは，発話文「エムアールアイ行ってきた」，聞き返し文「芋洗い」，誤聴箇所「エムアールアイ」を抽出する．この抽出は人手で行った．発話文と誤聴箇所，聞き返し文はそれぞれ漢字かな交じり文と読み仮名を保持する．誤聴データの一部を表 2 に示す．

6. 評価実験

4 節で述べた提案手法が誤聴箇所の推定に有用であるかどうかを評価するため，評価実験を行った．評価実験では 982 組の誤聴データから入力を与え，誤聴箇所の推定を行う．入力は発話文の漢字かな交じり文と読み仮名，問い返し文の読み仮名とする．評価実験での入力は，ロボットによる発話と聞き返しに対する音声認識が行われたと仮定し，テキストとして与える．また，評価実験で得た推定結果と誤聴データが保持する誤聴箇所との対応を調査し，以下の三つに分類する．

- 完全一致
  誤聴データが保持する誤聴箇所と一致したもの
- ほぼ一致
  助詞の消失や追加のみが現れたもの
  例えば誤聴箇所「美術館」に対して「美術館に」，誤聴箇所「時計を」に対して「時計」を出力した場合が当てはまる
- 不一致

*2 ほぼ日刊イトイ新聞 - 言いまつがい，
http://www.1101.com/iimatugai/

表 2　誤聴データの一部

番号	発話文（カタカナ）	誤聴箇所 聞き返し文	カタカナ カタカナ
1	5 名様ご案内	5 名様	ゴメイサマ
	ゴメイサマゴアンナイ	お姉様	オネエサマ
2	次は三滝口	三滝口	ミタキグチ
	ツギハミタキグチ	豚キムチ	ブタキムチ
3	また後逸だ	後逸	コウイツ
	マタコウイツダ	コイツ	コイツ

表 3　評価実験での推定結果

分類	占める割合[%]
完全一致	76.8
ほぼ一致	4.3
不一致	18.9

完全一致とほぼ一致に当てはまらないもの

評価実験による推定結果と誤聴箇所との対応の内訳を，表 3 に示す．評価実験における完全一致率は 76.9%となった．また，ほぼ一致を含めた場合の一致率は 81.1%となった．

7. 考察

推定に成功した例を示す．発話文「コードシェア便をご利用いただきありがとうございます」と聞き返し文「幸田チャーミン」に対して誤聴箇所「コードシェア便」を推定することができた．長い発話文中で起こる誤聴でも誤聴箇所の推定が行えた．

また，実際の誤聴箇所と推定結果の不一致の原因について考察を行った．不一致の原因について評価実験に用いた事例を挙げ，期待する出力と推定された誤聴箇所の違いを述べる．さらに，提案手法の妥当性を評価するための検証実験と調査についても述べる．

7.1 形態素解析器の出力による不一致

不一致として分類されたもののうち，形態素 n-gram の生成において読み誤りや形態素区切りの違いによる影響を受けた不一致が 32.8%（全体の 6.2%）あった．これを，形態素解析器の出力による不一致とする．

7.1.1 読み誤りによる不一致

発話文の読み仮名と漢字かな交じり文から得た形態素 n-gram が異なるために発生する不一致である．例えば，正しい誤聴箇所「さすらい人」に対して「サスライジン」が出力される．形態素 n-gram の中に「サスライビト」が存在しないため，これが出力されることがない．

7.1.2 形態素区切りの違いによる不一致

形態素解析器に登録された単語の有無による不一致である．例えば，正しい誤聴箇所「フラペチーノ」に対して「抹茶フラペチーノ」が出力される．「抹茶フラペチーノ」が一つの固有名詞として認識され，形態素 n-gram の中に「フラペチーノ」が存在しないため，これが出力されることがない．

7.2 その他の不一致

形態素解析器の出力による不一致に当てはまらない不一致が 67.2%（全体の 12.7%）あった．その中で，完全一致と成り得る候補を含むが他の候補で形態素 n-gram との一致が発生することによる不一致について取り上げる．例えば，正しい誤聴箇所「輸出入」に対して「オリーブノ」が出力される．発話文と聞き返し文は「フランスでオリーブの輸出入やってる」と「ヨシツネ」である．類似度の算出で得られる誤聴箇所候補は，以下に示す合計 6 候補である．

A) モーラ数 2 の誤聴箇所候補：スデ，ユシュ
B) モーラ数 3 の誤聴箇所候補：ヤッテ，ンスデ，ユシュツ
C) モーラ数 5 の誤聴箇所候補：オリーブノ

候補の中で，ユシュとユシュツは正しい誤聴箇所「輸出入」を得られる候補である．しかし，誤聴箇所の絞り込みの優先順位に従うと，形態素 n-gram との完全一致が起こる候補「オリーブノ」が誤聴箇所として推定される．

7.3 絞り込みの必要性

4 節で述べた提案手法では，二つの理由で誤聴箇所の絞り込みが行われている．一つは誤聴箇所の補正するため，もう一つは発話文中の誤聴箇所を明確にするためである．この必要性を確認するため，検証実験を行った．

検証実験では図 1 中の (6)誤聴箇所の絞り込みの処理を除いて，誤聴箇所の推定を行う．入力は発話文と聞き返し文の読み仮名のみとする．これらを用い，図 1 中の(5)類似度の算出で得た誤聴箇所候補を推定した誤聴箇所として扱う．誤聴箇所と一致するただ一つの推定結果を得た場合を一致とした．検証実験には，99 組の誤聴データを用いた．

検証実験による推定結果と誤聴箇所の対応の内訳を，表 4 に示す．検証実験における一致率は 48.5%（評価実験における一致率は 81.1%）となった．表 4 における一つの候補を持つ不一致には，期待する誤聴箇所「小論文」に対して「ショウロンブ」を推定するように，誤聴箇所に対して助詞以外の 1 モーラの追加や不足を含む推定結果を出力する事例があった．これは推定された「ショウロンブ」に対して補正を行うことで，形態素 1-gram である「小論文」を出力できる．

また，複数の候補を持つ不一致には誤聴箇所「国立」に対して「コクリツ」と「クリツ」の二つを候補に持つように，片方の候補が完全一致となる事例があった．これは候補一つ「コクリツ」が，発話文の形態素 1-gram であることを明確にすることで，「コクリツ」を出力できる．

これらの事例は，形態素 n-gram との比較を用いた誤聴箇所の絞り込みを導入した評価実験の結果で完全一致またはほぼ一致となり，推定の精度の向上に繋がったと考えられる．以上のことから，提案手法における誤聴箇所の絞り込みは必要であると言える．

7.4 モーラ n-gram における n の範囲

提案手法では抽出するモーラ n-gram の n の範囲を，聞き返し文のモーラ数 m に対して$m-2 \leq n \leq m+2$と定義した．この妥当性を確認するため，誤聴箇所と聞き返し文のモーラ数の差についての調査を行った．調査結果を表 5 に示す．表 5 の「モーラ数の差±a」は，モーラ数の差が a 以内であることを表す．例えば，モーラ数が同じであるのは全体の 55.9%であり，モーラ数の差が 1 以内まで含めると全体の 87.5%を占める．

また，評価実験において n の値を変更して推定結果を得たが，±2 のとき最良の結果を示しそれ以上 n の範囲を大きくしても精度は向上しなかった．以上のことから，推定手法で定義した n の範囲は妥当であると言える．

表 4　検証実験での推定結果

分類	占める割合[%]
一致	48.5
一つの候補をもつ不一致	19.2
複数の候補をもつ不一致	32.3

表 5　誤聴箇所と聞き返しのモーラ数の差

モーラ数の差	±0	±1	±2	±3	±4	±5
該当する誤聴データの割合[%]	55.9	87.5	97.4	97.4	99.3	99.9

8. おわりに

本論文ではロボットと人間の対話におけるコミュニケーションの断絶を防止するために，聞き返し文から発話文中の誤聴箇所を推定する手法を提案した．まず最初に，モーラを単位として発話文と聞き返し文の英字列の類似度を求めて誤聴箇所候補を得た．次に絞り込みを補正するため，発話文と誤聴箇所候補の文字列の類似度を求めて誤聴箇所を推定した．また評価実験を行い，提案手法による推定では 81.1%の精度を得られることを示した．

今後の課題は，類似度の算出に音素の類似を組み込むことである．正しい誤聴箇所を得られる候補がそれ以外の候補に比べて大きな類似度を示すことで推定の精度向上を目指す．

謝辞

本研究は，科学研究費補助金若手研究(B)(No.16K16134)の助成を受けて行われました．

参考文献

[1] 渡辺 靖彦, 横溝 一哉,西村 涼,岡田 至弘, “メーリングリストを利用した質問応答システムのための知識獲得”, 自然言語処理,Vol.12, No.6,pp.25-44, (2005).
[2] 白井 清昭, 徳江 英範, “対話型質問応答システムにおける聞き返し文生成に関する基礎研究”, 情報処理学会研究報告, pp.53-58, (2005).
[3] 清田 陽司, 黒橋 禎夫, 木戸 冬子, “大規模テキスト知識ベースに基づく自動質問応答:ダイアログナビ”, 自然言語処理,Vol.10,No.4,pp.145-175, (2003).
[4] 石川 開, 隅田 英一郎, “テキストデータを使った音声認識誤りの訂正”, 自然言語処理, Vol.7, No.4, pp. 205-227, (2000).
[5] 中谷 良平, 岩橋 直人, 中野 幹生, 滝口 哲也, 有木 康雄, “未知語とその周辺単語の音声認識誤りを考慮した CRF による音声認識誤り訂正”, 研究報告音声言語情報処理（SLP）, 2011-SLP-89(24), pp. 1-6, (2011).
[6] 小暮計貴 , 吉永眞宏 , 鈴木光 , 北原鉄朗, “周囲の雑音やユーザーの聞き返しに基づいて音量調節を行う音声対話システム”, 研究報告音楽情報科学（MUS） 2014-MUS-103(28), pp.1-5,(2014)
[7] 柳田 益造, “誤聴とそのメカニズム”, 信学技報, TL97-1, PP. 1-8, (1997).