

対話特徴可視化ツールの開発

小野寺佐知子 落谷亮

株式会社 富士通研究所

1. はじめに

複数の音声対話データから評価や分析の対象としたい箇所をみつけ出すとき、順にすべての音声データを聴いていく、適当な時間間隔で頭出ししながら聴いていくなどの方法では多くの時間・労力を要する。例えば、コールセンタにおけるオペレータの対話スキル評価作業では大量の録音コールを対象とするので大きなコストがかかる。そこで、話者属性、対話状況、音声データから抽出される韻律特徴等の情報を利用して対象箇所を選び出すことができると、効率的に作業が行える。

筆者らは、複数の音声データを指定した項目で分類し、その結果をみて別の項目を追加して分類する操作を繰り返し、データの絞り込みが行える対話特徴可視化ツールの開発を行った。本稿では、このツールの概要を紹介し、このツールをコールセンタにおける音声対話データに適用し、評価したい箇所の選別を行う方法を報告する。

2. コールセンタにおける対話スキル評価

現在、多くのコールセンタはオペレータと顧客との全通話録音を行っている。その録音コールを評価者がモニタリングして、コミュニケーションスキル、顧客対応スキル、問題解決技術等の評価を行い、オペレータの対話スキルを測定している。評価者の具体的な作業内容は以下ようになる。

- (1) オペレータごとにいくつかの録音コールを任意に選ぶ。
- (2) 選んだ録音コールを聴き、あらかじめ用意してある評価項目「わかりやすい質問の仕方か」「わかりやすい説明の仕方か」「親身になって対応する態度が表せているか」等について、対象箇所を見つけ出し、よし悪しの判定を行う。
- (3) 選んだ対話の評価得点を集計してオペレータのスキル評価得点とする。

3. 必要な機能

この中心となる(2)の作業は、オペレータの話し方として良いとされる発話、あるいは、その対極にある悪い発話を機械的に選ぶことができると、効率的に行える。そこで、話し方の特徴は複数の韻律特徴の組合せとして表れることから、各韻律特徴の観点で順にデータ整理ができる必要がある。まず、一つの韻律特徴（基本周波数、パワー等）で整理するために「音声データのある観点で分類する」、また韻律情報それぞれについて「分類した結果を比較す

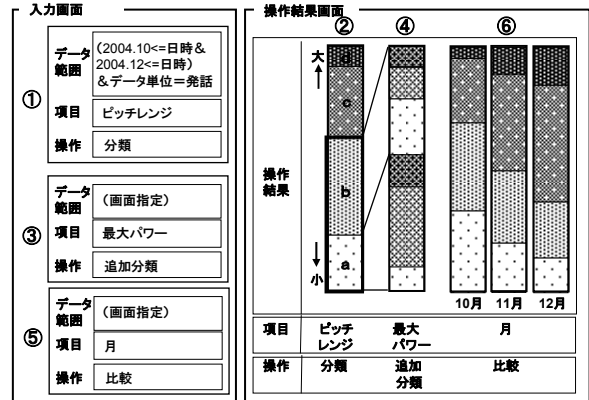


図1. 対話特徴可視化ツールの操作例

る」機能が必要である。さらに、特徴のある箇所を選び出すためには「分類した結果に、さらに異なる観点を追加して分類し、データの絞り込みを行う」機能が必要である。また、得られた結果の確認のためには「指定した箇所の音声を再生する」機能も必要である。

4. 対話特徴可視化ツール

以上の機能を実現する対話特徴可視化ツールの開発を行った。入力データは、音声データと、コールセンタの録音コールであれば音声データに付随する日時、話者等のデータとなる。音声データからは韻律特徴（基本周波数、パワー等）を算出し、各対話および対話を構成する各発話と対応させておく。

4.1. 操作と結果表示

各操作内容と操作結果を図1で説明する。

4.1.1. 分類

「2004年10月から12月の発話データ」を「ピッチレンジ」で「分類」する(①)。すると、ピッチレンジで分類した結果が表示される(②)。下からピッチレンジの値が小さい順に、各領域に分類される発話数の分類対象とした全発話数に対する割合が各領域の高さとして表示される。

4.1.2. 追加分類

対象とするデータの範囲を②の a, b とし、「最大パワー」で「追加分類」する(③)。すると、a, b に属するデータを領域ごとに「最大パワー」で再分類した結果が表示される(④)。

4.1.3. 比較

対象とするデータの範囲を②のすべてとし、「月」で「比較」する。すると、先の分類結果と同様の図が月別に並列表示される(⑥)。

4.1.4. 音声再生

すでに得られた分類結果から音声再生したい領

域を指定すると、その領域に属する音声を再生する。

5. ツールを利用した分析

5.1. “悪い話し方発話”の選別

人が音声を聴いて良し悪しを選別する時には、知覚される抑揚、声の大きさ、話す速度から判定している。これらの要素はそれぞれ韻律特徴のピッチレンジ、パワー、モーラ長と関連があり、明らかに良い発話・悪い発話と判定できる発話のピッチレンジ、パワー、モーラ長の値は経験から得られるため、その値を利用して判別が可能である発話もある。しかし、話し方を制御するために主に利用する要素は話者によって異なり、同じような韻律特徴を持つ発話が、ある話者では悪い発話、別の話者では良い発話となることがあるため、すべての発話を一律に選別することは難しい。そのため、話し方の特性に応じて選別方法を変える必要があると考える。

5.2. ツールによる選別手順

図2に女性話者AとBの操作結果を示す。ここでは、図による説明のため、6段階で色の薄い順に値の小さいものの属する領域を表示している。

話し方特性

発話の「ピッチレンジ」「最大パワー」「モーラ長」の各項目で分類を行う。その結果を図2①～⑥に示す。話者Aは、ピッチレンジが大きい値に偏っており、全体に抑揚のある発話を行っていることが伺える。一方、話者Bは話者Aに比べてピッチレンジは小さい値に偏っているが、モーラ長は大きく、全体にゆっくりとした話し方であることが伺える。このように、異なる話し方の特性に応じて、最初に着目する韻律特徴を決め、対象とするデータ範囲を選出し、異なる韻律特徴でさらに分類を行う操作を繰り返して悪い発話となっている可能性の高い発話を絞り込んでいく。

抑揚をつけて話す話者の場合

全体に抑揚のある話し方（話者A）の場合、まず、悪い話し方である可能性の高い抑揚のない発話を選ぶ。次に、抑揚があっても極端に速い・遅い発話で声が小さい発話も、悪い発話である可能性が高いのでそのような発話を選ぶ。そこで、
 [1]ピッチレンジの小さい発話（図2①' a）と、ピッチレンジが比較的大きい発話（図2①' b, c）を含めた領域（図2①'の太枠）を対象範囲として、「モーラ長」で追加分類を行う。
 [2]その結果（図2⑦）、モーラ長が極端に大きい・小さい発話およびピッチレンジの小さい発話の領域（図2⑦の太枠）を対象範囲として、「最大パワー」で追加分類を行う。
 [3]その結果（図2⑧）、最大パワーが極端に小さい発話およびピッチレンジの小さい発話の領域（図2⑧の太枠）を対象範囲として、再び「ピッチレンジ」で分類する。
 [4]その結果（図2⑨）は悪い発話である可能性の

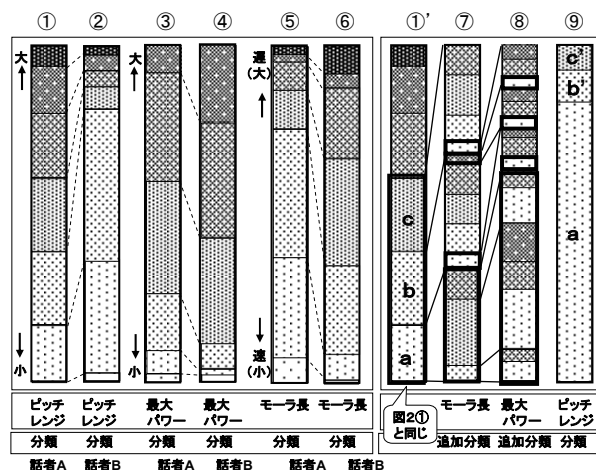


図2. “悪い話し方発話”の選別操作

高い発話となっており、抑揚のない発話（⑨ a : ①' a と同じ発話の属する領域）、および、抑揚は比較的あるが話速が速い・遅い、かつ、小さい声の発話（⑨ b', c' : ①' b, c のうち[2][3]の過程で絞り込まれた発話の属する領域）が選別される。

ゆっくり話す話者の場合

全体にゆっくりとした話し方（話者B）の場合、まず、悪い話し方である可能性の高い、極端に速い発話を選ぶ。次に、話速が適切であっても抑揚がなく音量が小さい発話も悪い発話である可能性が高いのでそのような発話を選ぶ。そこで、Aと同様に、まず、モーラ長が小さい発話と比較的適切なモーラ長である発話の領域を対象範囲として、「ピッチレンジ」、さらに「最大パワー」で追加分類を行い、再び「モーラ長」で分類を行って、悪い発話である可能性の高い発話を選別する。

5.3. 選別精度

ツールによる選別の結果得られた発話で、かつ、対象となる発話を実際に聴いて良し悪し判定を行い悪い発話と判定した発話の数を x, 実際に聴いて悪い発話と判定した発話数を y, ツールによる選別の結果得られた発話数を z とし、再現率 = x / y, 適合率 = x / z として選別精度を調べると、A では再現率 59.3% 適合率 76.2%, B では再現率 79.2% 適合率 61.3% であった。音声を聴いて悪い発話と判定した数は対象発話数の約 20% であったので、ランダムに発話を選ぶより、ツールによって選別を行った方が効率的に悪い発話を選別できるといえる。

6. まとめ

本稿では、対話特徴可視化ツールの概要と、ユーザ対話に適用した場合の利用方法を示した。今回は韻律情報のみを利用したが、対話中の発話の役割や話題が選別に影響するので、それらを考慮することで選別の精度が向上すると考えられる。今後はこのような言語情報の利用に取り組んでいく予定である。