

スピーチの話し方と聴衆の興奮度の可視化

吉澤 晃太郎¹ 加納 徹¹ 竹島 由里子¹

東京工科大学

1. はじめに

スピーチコンテストや大統領演説, Apple 社の新商品発表会, 結婚式など, 用途や目的は違うが, 多種多様な場面でスピーチは行われている. スピーチで必要とされる話し方は, 場面や目的に応じて変化するが, 良いとされるスピーチには一定の共通項がある. 山崎 [1] によると, 一般的にスピーチに適しているとされる話し方は, 緩急をつけつつ, 適切な間を取りながら, 声の高低などで抑揚をつけて, 大きな声で話すというものである. しかし, 真にいいスピーチであるかを判断するには, スピーチをしている人だけでなく, それを聞いている聴衆の反応も調べる必要がある. 音声に関する研究は多く行われているが, スピーチの話し方とそれを聞いた聴衆の反応の相関に関する研究は存在しない. そこで本研究では, 聴衆の反応が顕著である米大統領のスピーチを対象に, 話し方と聴衆の興奮度の両方の変化に着目して分析する. 話者の抑揚や速さ, 声の大きさといった特徴と, 聴衆の出す音の高低と大きさを同時軸上に可視化することにより, 聴衆の盛り上がりとスピーチの話し方の関係性が明らかになる. さらに, さまざまなスピーチの可視化結果を比較分析することで, 良いとされるスピーチの共通点の獲得, および聴衆の心を惹きつける話し方の解明が期待される.

2. 処理の流れ

2.1 音声の取得, 加工

本研究では, ABC News や The New York Times などの米国ニュースサイトが提供している無編集の大統領演説動画から, ステレオ音源を抽出して利用する. まず, 数値解析ソフト MATLAB を使用し, ステレオ間での音の位相差に基づいて, 話者のスピーチ部分を分離する. この処理により発生したノイズは, Adobe 社が提供している音声編集ソフト Audition のノイズキャンセリング機能を用いて削減する. そして, 話者のスピーチ部分の逆位相をフィルタにかけることで, 聴衆の反応部分を抽出する.

2.2 特徴量抽出

次に, スピーチの重要な要素であるテンポ, 声の高低, 声の大きさを抽出する. これら特徴量の抽出には, 数値解析ソフト MATLAB 上に実装される楽曲特徴分析パッケージ MIRToolbox を使用する. ただし, 聴衆の反応部分については, 不特定多数の人間が一斉に声だけでなく拍手や口笛といった音の生じる行動をするため, 一定の周期を確認するためのテンポは用いないものとする. 抽出した特徴量は, 外れ値や各要素にとって必要のない帯域などの削除加工を行う.

また, ある時刻 (t) における聴衆の反応は, 音の高低 (F_t) と音量 (V_t) から興奮度 (E_t) として, 式 (1) に基づいて算出する.

$$E_t = V_t \log_2(wF_t) \quad (1)$$

音量に対数をとった音の高低を乗じることで, 興奮度と定義した. 対数をとっているのは, 音の高低の値が指数関数的に増加するため, そのまま使用すると音量と比較して音の高低の影響が大きすぎてしまうためである. 周波数感度 (w) は音の高低の影響を制御する数値であり, 値が大きいほど興奮度に対する音の高低の影響が高くなる.

2.3 可視化

得られた特徴量を元に, 話者の話し方と聴衆の反応を同時軸上に可視化していく. 話者のテンポ, 音の高低, 声の大きさは, 折れ線を用いて表現する. 折れ線の色がテンポに対応し, 色相環で青に近いほどテンポが遅く, 赤に近いほどテンポが早くなっていることを表す. 折れ線の縦軸が声の高低に対応し, 画面の上に行くほど高い声で, 下に行くほど低い声で話していることを表す. 折れ線の下に伸びる線の長さが声の大きさに対応し, 線が下に長いほど声が大きく, 短いほど声が小さいことを表す.

聴衆の反応は式 (1) で算出した興奮度に基づき, 背景色として表現する. 背景色は興奮度が高いほど白に近づくように設定する.

3. 結果

ドナルド・トランプ米大統領就任演説を可視化した結果

*Visualization of speech patterns and audience reactions

¹ Tokyo University of Technology

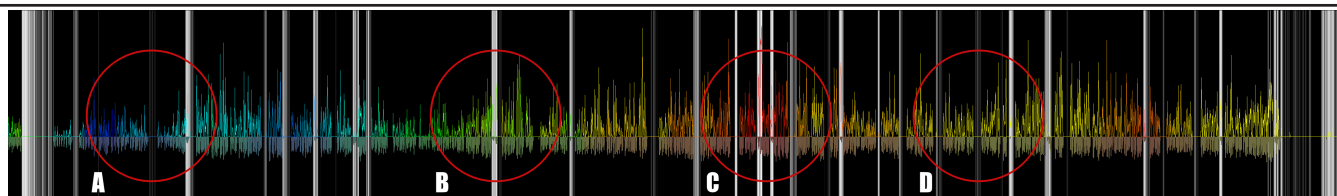


図1 ドナルド・トランプ米大統領就任演説の可視化結果

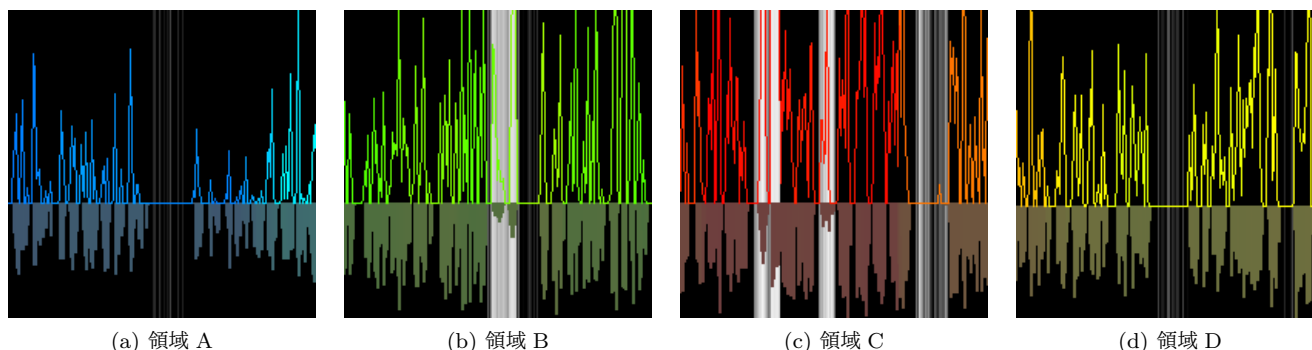


図2 図1における聴衆の反応が特徴的な部分の拡大図

を図1に示す。この演説は約17分間で、可視化における時間分解能は1ピクセルあたり約0.25秒である。スピーチの折れ線から、序盤から中盤までの間はテンポにあまり変化がなく、声の高低の変化はあまりないことがわかる。中盤からはテンポが早くなっており、序盤に比べて声が高くなっている。中盤から終盤にかけては、声を発していない時間が多く見られ、間を多くとった話し方をしていることがわかる。音量の線の長さを見ると、序盤から終盤にかけて徐々に大きく変化しているが、局所的に音量が著しく変化している箇所は見られない。一方、背景の変化から、聴衆の反応が見て取れる。強い反応がある箇所は全体的に散らばっているが、終盤になるにつれ、聴衆が反応を示している箇所は増えている。また、聴衆の反応は主にスピーチの間が空いている箇所で発生していることがわかる。

図2に、図1において聴衆の反応が特徴的な箇所の拡大図を示す。領域Aでは、間が大きくあるにも関わらず、聴衆の反応が少ない部分である。これは、言葉尻が小さくなることで聴衆が冷静になり、反応を返すことができなかつたためだと考えられる。実際にこの領域は、スピーチは全体から見てやや遅めのテンポで話しており、間がある直前に声の高低と音量の値が小さくなっていることが確認できる。領域Bでは、話し終わる前に聴衆が強く反応している部分である。序盤に比べて、テンポが若干早く、音量の値から言葉尻もはっきりと発声している。話し始めの声を低くし、段々と高低の激しい話し方へと変化していることも見て取れる。このような話し方により聴衆の興奮を誘発させることで、聴衆が小さな間にも反応しやすくなったと推測できる。聴衆の興奮が多く現れているのが領域Cである。特徴としてテンポはかなり早く、声の高低も激しい。また、聴衆の反応中に話し始めていることから、話者自身も興奮していると推測できる。領域Dは間が大きく開い

ているが聴衆の反応がよくなかつた部分である。テンポが比較的早く、言葉尻がはっきりしている傾向は領域Bと類似しているが、間の直前の声の高低があまり変化していない。このことから、声の高低の変化によって聴衆の反応も変化すると推測できる。

4. おわりに

本研究では、動画サイトから米大統領就任演説を取得し、話し方の特徴をテンポ、声の高低、声の大きさから、それに伴う聴衆の反応を声の高低、音の大きさから抽出し、可視化した。これにより、スピーチの話し方の変化によって、聴衆の反応に差が生じることが直感的に理解できるようになった。実験から、聴衆の反応は間がある部分に集中しており、テンポが早いほど、音の高低差が大きいほど、反応が良い傾向がことがあった。また、遅いテンポや声の高低の変化が少ないと、聴衆の反応があまりないことや、全体を通して話者の声の大きさのみに対する、過度な聴衆の反応はないことがわかった。スピーチ終盤に行くほど聴衆の反応が良いのは、前半に比べて音量が大きくなっていることに起因する可能性も示唆された。

本研究で得られた、結果の有用性を示すためには、他大統領スピーチとの比較分析が必要不可欠である。また、本研究ではステレオ音源のみを抽出して利用したが、任意のスピーチからスピーチ話者の音声と聴衆の出す音を完全に分離する技術は現段階では確立されていない。そのため、解析精度を高めるためにはこれらの技術について検討をする必要がある。

参考文献

[1] 山崎広子：8割の人は自分の声が嫌い，KADOKAWA (2014).