

フラクタル次元を用いたパラ言語解析

奈良工業高等専門学校 ○馬淵 拓実 ◎西田 茂生

要 旨

近年、人間と機械の音声による情報交換の機会が増えている。機械から人間への情報伝達をより円滑に行うために、機械音声に感情パラメータを含ませることを考える。第一段階としてフラクタル次元解析による新たな評価方法により、音声とパラ言語の相関を解析する。喜びや怒りなどの感情をこめて同じ文章を読み上げた音声について、フラクタル次元解析を行った。解析した結果、フラクタル次元の時間変化について違いがみられた。

1. 緒言

近年、音声認識システムが発展を遂げ、機械と人間の音声による情報交換が可能である。人間は機械の発する音声を聞き取ることによってコミュニケーションとしてきた。しかし、さらなる技術の発展により人間と機械のコミュニケーションの機会は非常に増えており、機械音声の機械らしさがコミュニケーションの受け取り手である人間との大きな壁となっている。聞き取りやすく分かりやすいだけでは、コミュニケーションとして不十分である。

音声情報から得られる言語以外の情報を、感情と音声のパラメータと考え、声との相関関係を解析する。また、このような音声からなる言語情報以外の要素をパラ言語と呼称する。音声に含まれる感情を定量的に示すことが出来れば、ある音声に特定の感情を含ませることが出来る。本研究では、機械音声に感情パラメータを含ませる第一段階として「フラクタル次元解析」という新たな評価方法を提案し、音声とパラ言語の相関関係の解析と比較を行う。

2. 解析方法

本研究では、複雑な形状に対して有効なフラクタル⁽¹⁾という概念を用いた、フラクタル次元解析によって音声に含まれる感情を解析する。フラクタル次元の定義には複数種類があるが、本研究ではボックスカウント次元を用いたボックスカウンティング法⁽²⁾、相関次元を用いた G-P 法⁽³⁾により解析を行う。

2.1 ボックスカウンティング法

ボックスカウンティング法の簡単な手順を図1を用いて以下に示す。

- (1)図形を格子で覆う。
- (2)図形の構成要素が含まれる格子の数をカウントする。
- (3)格子の大きさを変化させながら、同様に格子の数をカウントする。
- (4)格子の大きさと格子の数を両対数グラフにプロットし、直線近似する。
- (5)直線近似した時の傾きの絶対値をフラクタル次元とする。

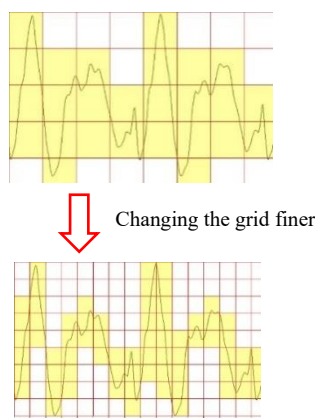


Fig.1 Box counting algorithm

感情パラメータは時間変化すると仮定し、時間依存型のフラクタル次元解析⁽⁴⁾を行う。試験音声として JSUT (Japanese speech corpus of Saruwatari-lab., University of Tokyo)の JUST コーパスを使用した。この音声コーパスは日本人女性声優により 100 パターンの音素バランス文を通常、喜び、怒りのそれぞれの感情によって読み上げられている。試験音声に対し、100ms の区間を設ける。この区間におけるフラクタル次元を算出し、その数値を瞬間のフラクタル次元とする。区間を 20ms 毎にずらし、各区間の瞬間のフラクタル次元をデータとして取る。通常の声と喜びの声に分けて、解析と比較を行う。

2.2 G-P 法

ランダムな点の集合が占める空間の次元の尺度を相関次元と称する。G-P 法は時間変化により物理系が変化する集合アトラクタに着目することで、相関次元を求める手法である。時系列データ $x(t)$ を一定時間遅れ τ でシフトさせたデータ群、ベクトル X_i を考える。埋め込み次元を与え、アトラクタを再構成する。 X_i を中心とした半径 r の超球内にアトラクタを構成する要素が含まれている数を数え、 N^2 で割ることで相関積分 $C(r)$ を求めることが出来る。 $C(r)$ の導出方法を式(1)に示す。

$$C(r) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N H(r - |X_i - X_j|) \quad (1)$$

相関次元 D の導出方法を式(2)に示す。

$$D = \frac{\log C(r)}{\log r} \quad (2)$$

r と $C(r)$ の対数をプロットしたときのグラフの傾きを相関次元として扱う。

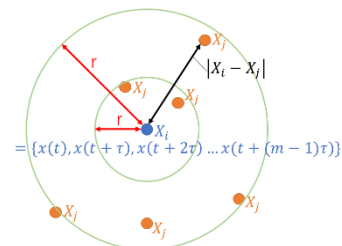


Fig.4 G-P algorithm

3. 解析結果

3.1 ボックスカウンティング法による解析結果

図2にボックスカウンティング法による解析結果を示す。喜びの感情を込めた音声は全体的に次元が高いことが分かる。また、怒りの感情を込めた音声は他2つと比べて次元の変動が大きいくことが分かる。

3.2 G-P 法による解析結果

図3に G-P 法による解析結果を示す。3種類のデータについて各感情の音声解析を行った。G-P 法でも同様に喜びの感情を込めた音声で最も次元が高いという結果が得られた。また、同じ音声に対して振幅を変化させて同様に解析を行ったが、次元の変化は見られなかった。

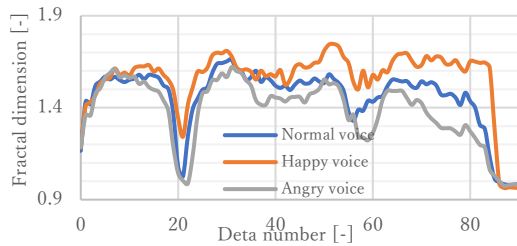


Fig.2 Fractal dimension of each emotion

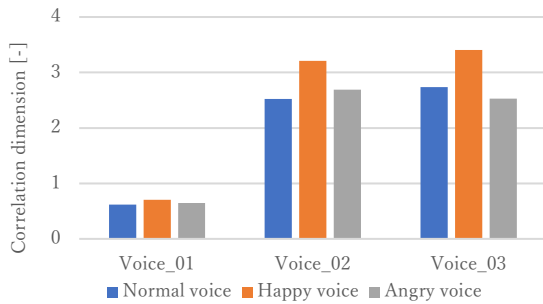


Fig.3 Correlation dimension of each emotion

4. 結言

今回パラ言語解析を行う上で、フラクタル次元という新たな解析方法を提案し比較した。音声に含まれる感情パラメータは時間変化すると考え、時間変位を観測できるように解析を行った。ボックスカウンティング法と G-P 法のいずれの解析方法でも喜びの感情を込めた音声は次元が高くなる傾向にあることが分かった。また、音声の振幅が次元に与える影響は少ないものであることが分かった。これはデータの大きさが変化してもアトラクタの構造そのものには影響を及ぼすことがないためだと考えられる。G-P 法は音声全体のデータについて解析しているが、同様に時間依存型の次元解析を行うことが出来ると考えられる。また、G-P 法によって解析するにあたって、膨大なデータ量を扱うため解析に時間がかかってしまうことが課題である。

参考文献

- (1) 高安秀, “フラクタル”, 朝倉書店, (1986).
- (2) 塩谷昌典, “ボックスカウント次元の計算”, JAPLA2003, シンポジウム, (2003).
- (3) 蛭田幸太郎, 他, “時系列データのフラクタル次元: 音声の解析”, 東京学芸大学紀要出版委員会, (1987).
- (4) 丸山貴司, 他, “脳波フラクタル次元解析を用いた感性解析”, 知能と情報 (日本知能情報ファジィ学会誌) Vol.24, No.6(2012), pp.1137-1153.