

モーメントを用いた音声の統計的解析

電気・化学専攻 医用工学領域 2181223 河野 慶太

(主査: 桐生 昭吾 教授 副査: 桃沢 愛 准教授 島谷 祐一 准教授)

1. はじめに

私たちの身の回りには様々な音が存在し、音から様々な情報や状況を判断している。例えば、喋っている声による意思の疎通や、道を歩いているときに視覚外から来る車の音による危険の察知など、多くの音による情報を意識的、もしくは無意識的に行っている。その中で人間はカクテルパーティ効果という優れた能力を持っており、二つ以上の声や、騒音下で特定の音を感覚的に聞き分ける事ができる。そしてこれを機械的に区別する方法として音源強調と音源分離という二つの技術が存在する。私はその中でも、音声認識機能向上といった製品への応用を考え、音源分離という技術に着目した。信号解析においては近年、周波数領域によって信号の特性を把握する周波数解析やフーリエ変換を用いるケプストラム解析など様々な解析方法で解析が行われている。しかし音声信号はノイズなどの部分的に予測できない変動を持っており、観測したい音源の位置や大きさが未知の場合、分離をすることは困難である。そのため、信号の独立性を利用したブラインド音源分離という方法を用いる必要がある。これは複数の音源が互いに独立であるという仮定から統計的に母数の推定が可能なモーメント法を用いる解析方法であり、現在は四次のモーメントが音源分離に利用可能な評価関数として使用されている。一方で五次以降の高次モーメントはあまり使用されていない。本研究では四次では区別不可能な音声が高次以降の高次モーメントを用いることで区別が可能になると考え、高次のモーメントが新たな評価関数として利用可能か評価した。

2. 研究方法

2.1 モーメントの規格化と理論

高次モーメントによる解析の基本式は、確率変数 x 、サンプルの個数 N 、モーメントの次数 $M[k]$ として (1) 式のように示される。

$$M[k] = \frac{1}{N} \sum_{i=0}^{N-1} (x[i] - \bar{x})^k \quad (1)$$

(1)式に振幅に対する不変性、単位の統合等を考慮し標準偏差 σ を使い(2)式のように規格化した。本研究では(2)式を用いて解析を行う。式に振幅に対する不変性、単位の統合等を考慮し標準偏差 σ を使い(2)式のように規格化した。本研究では(2)式を用いて解析を行う。

$$M_n[k] = \sqrt[k]{\frac{1}{\sigma^k} \frac{1}{N} \sum_{i=0}^{N-1} (x[i] - \bar{x})^k} \quad (2)$$

正規分布などのようにいくつかの確立分布はモーメント母関数から解析的に規格化モーメントを計算できる。本研究では、基準として正規分布を使用する。現在では正規分布を主に参考としているが、信号によっては一様分布やラプラス分布なども参考にする場合がある。正規分布の理論値は、平均値に対して対称な場合は0となる。(2)式より1次モーメント、2次モーメントは必ず同じ値になるため、3次以降の高次モーメントによって判断する。また偶数次のモーメントは(3)式で表される。

$$M_n^k = \sqrt[k]{(k-1)(k-3)\cdots(k-(2i-1))} \quad (3)$$

ただし、 $k < (2i-1)$ とする。

次数	正規分布の理論値
1	0
2	1
3	0
4	1.316074013
5	0
6	1.570417802
7	0
8	1.789157867
9	0
10	1.984006912
11	0
12	2.161401131
13	0
14	2.325323698
15	0
16	2.478441623
17	0
18	2.62264395
19	0
20	2.759326433

Table.1 正規分布の理論値

正規分布の理論値は上記の Table.1 の通りである。

2.2 解析方法

音声の収録には既存のスマートフォン用アプリ「簡単ボイスレコーダー (Digipom 社)」を使用した。このアプリを用いて、計 4 名の 20 代男女が異なる文学作品を朗読する音声を録音した。録音した音声の長さは各 5 分程度である。また、収録条件は実環境での音源分離を想定し、残響のある静かな部屋の中で収録を行った。本研究における騒音の解析には Python のプログラムを使用した。これは、騒音の音声信号をプログラム上に読み込むと、データの解析範囲の波形と規格化したモーメントの値、正規分布の理論値がグラフで表示されるというプログラムになっている。表示されるグラフの例を図 1、図 2 に示す。図 1 は騒音の解析範囲の波形で縦軸が振幅、横軸がサンプル数を表す。図 2 は騒音の解析結果の例であり、縦軸は規格化したモーメントの値、横軸はモーメントの次数を表す。

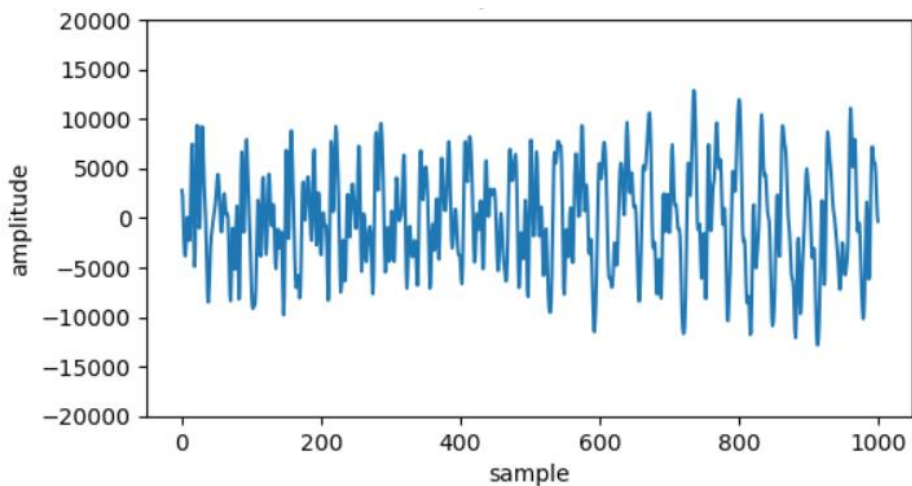


Fig.1 解析した波形の例

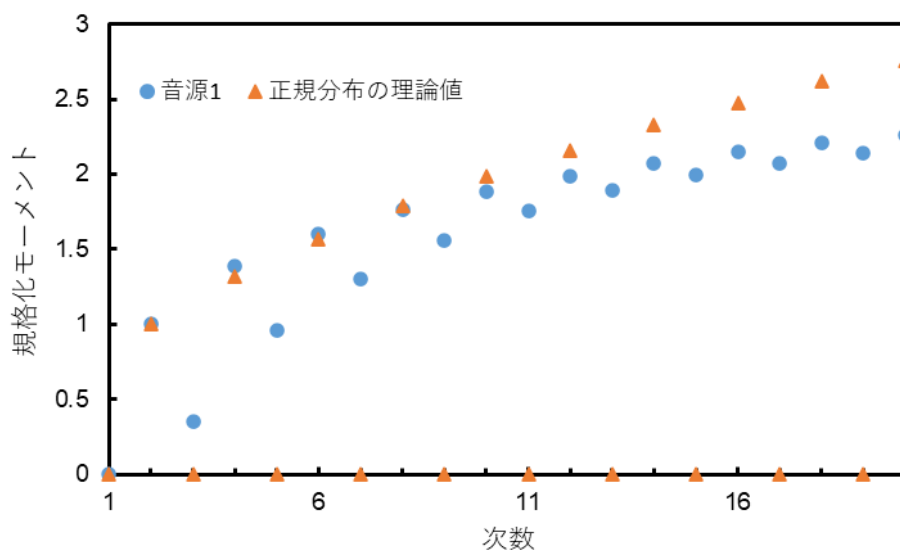


Fig.2 解析結果の例

3. 実験結果

3.1 使用音源

被験者 1 の朗読音源(C)

被験者 2 の朗読音源(D)

被験者 3 の朗読音源(E)

被験者 4 の朗読音源(F)

各朗読音源を混合した音源(mix)

3.2 解析方法

20 代男女 4 名が異なる文学作品を朗読した音声を用いた。また、朗読した音声に加え、朗読した 4 種類の音声を混合した音声を用いた。音声のデータは各音声でランダムに選んだ 10 データをそれぞれ解析し、各音のモーメント値の平均値を求め比較し、評価を行った。音の長さは先行研究で最も各音間で差異がみられた 10000 サンプルを用いた。(サンプリング周波数: 44.1kHz)

今回は混合する音声における音の比率を変化させた場合、どのような結果が得られるか比較、評価した。話者 D, E, F の音声を元の音源から -12dB した音圧で統一し、C の音声を ± 0 dB(元の音源), ± 6 dB, -12dB, -24dB と音圧を変化させた。

3.3 解析結果

各朗読者の音声の規格化モーメント値の比較結果を Fig.8, 9, 10, 11, 12 に示す。被験者 C の音声のみ ± 0 dB(元の音源), ± 6 dB, -12dB, -24dB と音圧を変化させ、混合した音声にどのような変化が表れるか評価した。

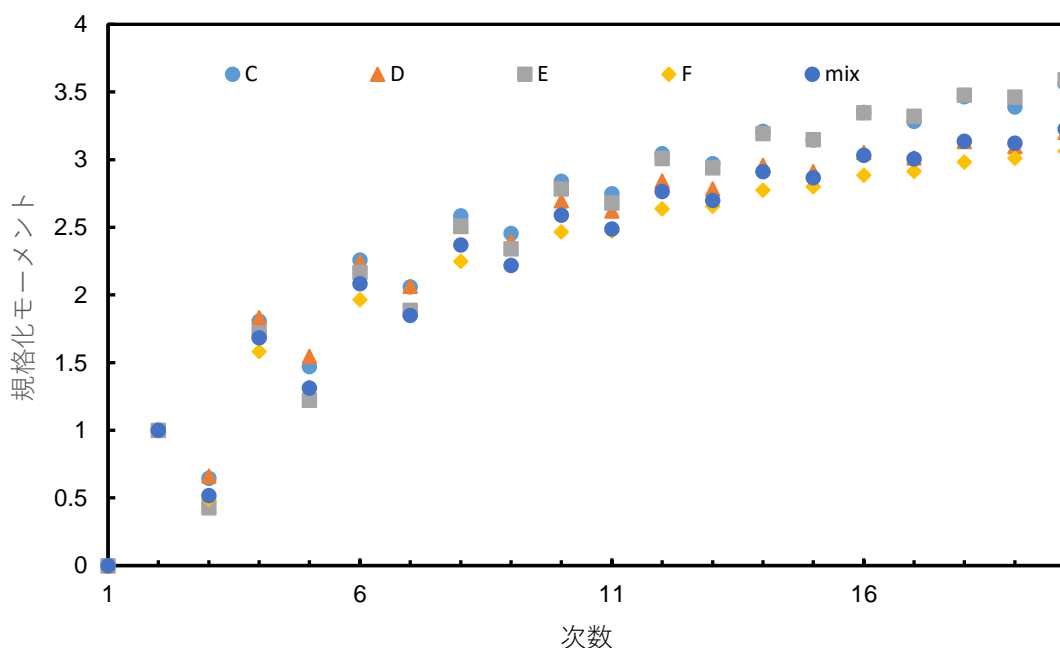


Fig.3 解析結果(Cのみ+6dB)

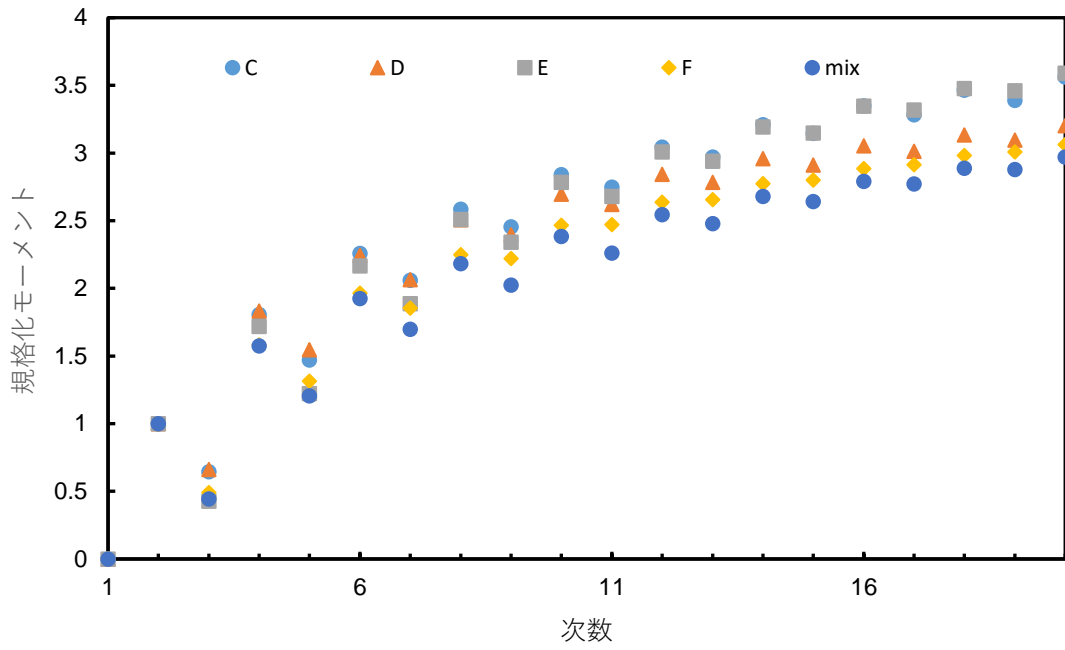


Fig.4 解析結果(Cのみ±0dB)

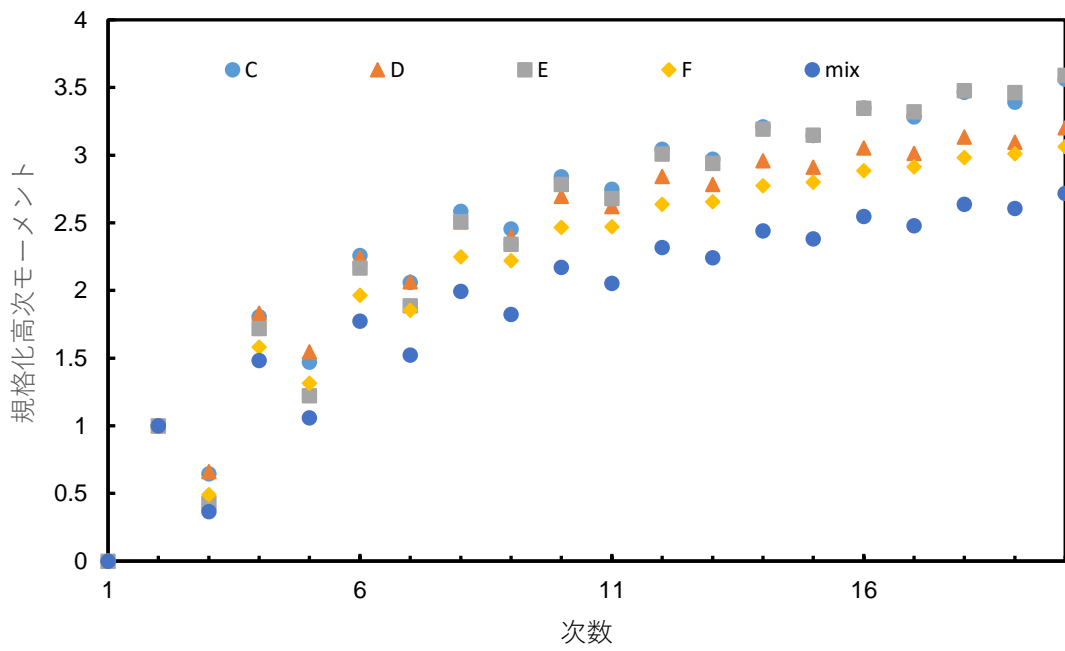


Fig.5 解析結果(Cのみ-6dB)

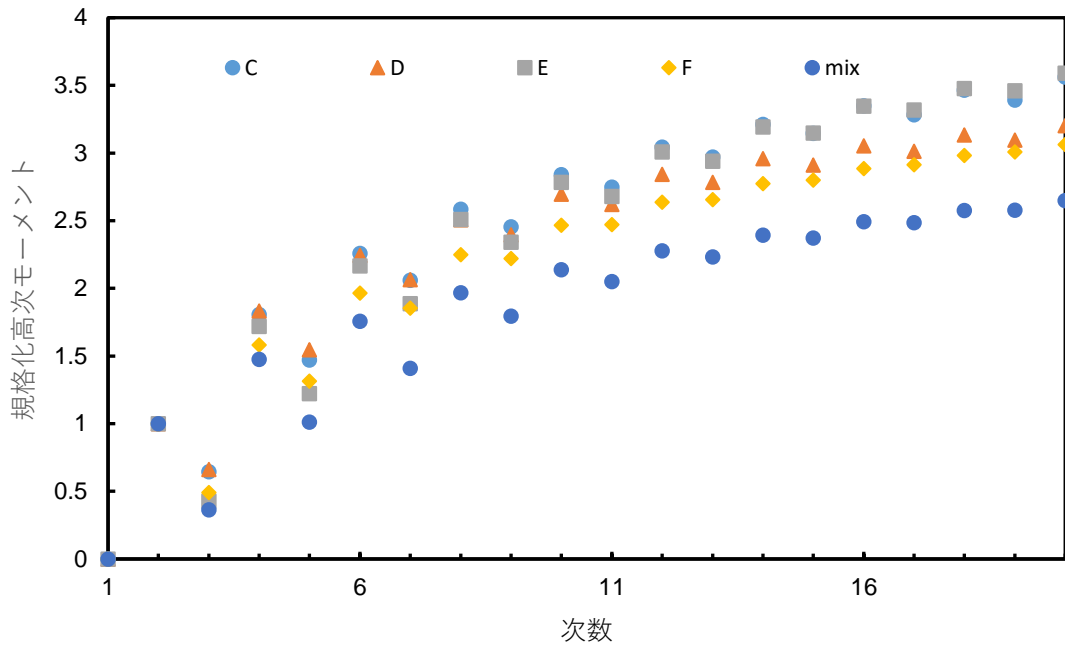


Fig.6 解析結果(Cのみ-12dB)

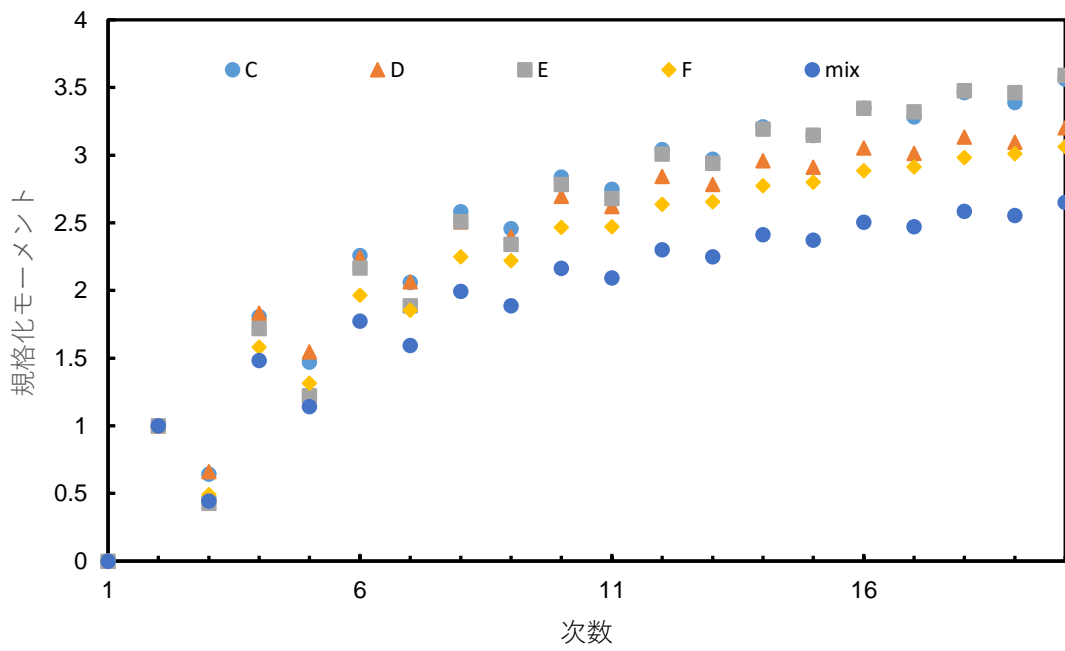


Fig.7 解析結果(Cのみ-24dB)

Fig.3, 4, 5, 6, 7 より、4 次では各朗読音源と混合音声の間で差異は小さかったが、高次では+6dB の時を除いて変化させた C と mix の音声間で差異が見られたが、±0dB では変化させていない D, F の音声と mix の音声に差異は見られなかった。+6dB では、混合音声のモーメントの値より、D, F の単独の音声の値が小さくなった。C 単独の音声は音圧を変化させても値に変化は見られなかった。また、混合音声のモーメントの値は音圧を下げると値が小さくなった。ただし、-6dB より音圧を下げても値に変化は見られなかった。

4. 考察

Table.2, 3 より、被験者単独の音声は音圧を変化させても規格化しているためモーメントの値に変化は見られないが、混合音声の値は音声間の比率によって変化することが分かった。一方で Table.2 より、C の音声を-6dB, -12dB, -24dB と音圧を低くしたところ、それぞれの混合した音声の値に変化は見られなかった。このことから、もともと元の音源から 12dB 低く設定した D, E, F の音声は、Fig3, 4 のように C の音圧を+6dB, ±0dB に変化させている場合、C の音声の音圧の大きさに対して音圧が小さくなりすぎてしまっている可能性が考えられる。一方で、C の音圧を低くすると、他の D, E, F の音声との差異がはっきりと表れ、Fig.5, 6, 7 のような-6dB, -12dB, -24dB における混合した音声のモーメントの値に変化が見られなくなったと考えられる。また、音声は混合すれば混合するほどモーメントの値が小さくなると言われている。よって、音圧を低くした際に各音声にはっきりと差異が見られ、モーメントの値が小さくなったと考えられる。

これらのことから、音声間の音圧の差が大きすぎるとモーメントによる音源分離は難しいと考えられる。ただし、どの程度の音圧から認識が難しくなるのか、どの程度混合音声の値に変化が見られるかなど細かな評価が必要である。

		次数			
		4	16	18	20
音圧(dB)	6	1.80699	3.34863	3.46443	3.56288
	0	1.80699	3.34870	3.46451	3.56296
	-6	1.80699	3.34883	3.46465	3.56312
	-12	1.80688	3.34900	3.46483	3.56329
	-24	1.80565	3.34727	3.46325	3.56182

Table.2 被験者 1(C)の朗読音源のモーメント値

		次数			
		4	16	18	20
音圧(dB)	6	1.6838	3.0327	3.1368	3.2268
	0	1.5757	2.7916	2.8873	2.9697
	-6	1.4836	2.5475	2.6383	2.7169
	-12	1.4745	2.4925	2.5767	2.6496
	-24	1.4827	2.5055	2.5839	2.6511

Table.3 混合音声(mix)の朗読音源のモーメント値

5. まとめ

本研究では異なる話者の朗読音源を用いてモーメントによる解析を行った。単独の音声のうち一つの音声の音圧を変化させて、混合した音声にどのような変化が見られるか解析した結果、音圧を下げ、他の音声と同等の大きさに変化させると混合した音声のモーメントの値が低くなるということが分かった。ただし、ある一定の音圧からモーメントの値が変化しなくなったため、どの程度音圧を変化させると、どの程度モーメントの値が変化するかなど検討する必要がある。

今後の研究計画としては、異なる音声を用いて同様の結果が得られるか、また、音圧を低くした場合、異なるサンプル数でも話者単独の音声と混合した音声に差異が見られるかといった比較、評価の必要がある。

6. 参考文献

[1] 佐藤他,“工学基礎演習シリーズ 3 確率・統計・ランダム過程”,森北出版株式会社,p.115 ,(1980)

[2] 石川 他,“多層型並列処理回路を用いた n 次モーメントの検出方法,”計測自動制御学会論文集 第 25 卷 第 8 号, pp76-78 (1989)

[3]根本 他,“詳解 独立成分分析 信号解析の新しい世界,”東京電機大学出版局, pp.39-43 (2005)