論　文

二つの周波数成分の分離知覚に関する工学的モデル
——複数の要因の評価と統合——

正　員　柏野　邦夫†　　正　員　田中　英彦†

A Computational Model of Auditory Segregation of Two Frequency Components —— Evaluation and Integration of Multiple Cues ——

Kunio KASHINO† *and* Hidehiko TANAKA†, *Members*

あらまし　知覚的音源分離システムを実現するための基礎として，二つの周波数成分の分離知覚に関するモデルを提案する．知覚的音源分離を周波数成分のクラスタリング問題として定式化すれば，二つの周波数成分がシステムに入力されたとき，これらがいかなる条件を満たした場合に別の音として認識するべきかを定めることが，最も基本的な問題となる．この問題に対し，スペクトログラム上の複数の特徴と分離知覚の生じる割合との定量的対応を与えるのが本モデルである．本モデルは，特徴の評価と統合とに基づいている．まず評価に関して，スペクトログラム上の2種類の特徴に着目して聴覚実験を行い，その結果に基づいて特徴の評価関数を定めた．次に評価値の統合に関して，Dempsterの結合規則を用いる方法を提案し，聴覚実験によってその妥当性を確認した．従来，スペクトログラム上の特徴に基づく音源分離処理の試みは，特徴を定性的に扱うにとどまっていたが，本論文によって，二つの周波数成分の分離知覚に関して，複数の特徴と分離知覚の生じる割合との関係が定量的にモデル化された．

キーワード　音源分離，音によるシーン解析，周波数成分，Dempsterの結合規則，クラスタリング

1. まえがき

本論文は，複数の音源が同時に存在している状況下で，それぞれの音源に由来する音響的情報を分離抽出する問題を念頭において，この問題に関する最も基本的なモデルを提案するものである．本論文ではこのような問題を音源分離問題と呼ぶ．音源分離を工学的に実現することは，実環境における音声言語や音楽など音響的情報を処理する上で極めて重要であり，音声認識システムのフロントエンドとしての応用や，複数種類の楽器演奏を対象とする自動採譜システムなどへの応用が期待できる．また音源分離は，音響的情報によるシーンの理解を目的とした「音によるシーン解析」(auditory scene analysis)[1]を計算機上で実現するための要素技術でもある．

この問題を考えるに際して，本論文では，物理的な音源と，知覚的な音とを区別する．物理的な音源が実際の発音体を意味するのに対し，知覚的な音とは，人間が一つの音と知覚(または認識)するような音を意味する．音源分離問題に対して従来試みられてきた方法には，音源の位置の情報を用いたもの[2]~[5]や，音源の具体的性質を利用したもの[6]があるが，これらは主に物理的な音源の分離(物理的音源分離)を前提としたものである．これらの方法では，音源の位置の情報または具体的性質のどちらかが必要であった．

一方，本論文では，知覚的な音の分離を行うシステム(これを知覚的音源分離システムと呼ぶ)の構築を念頭においている．知覚的音源分離は，人間の知覚的特性に即して音響的情報の分離抽出を行うものであり，将来的には，利用できるチャネル数(マイクロホン数)や音源の位置に対して自由度が高く，かつシステムにとって未知の音が入力された場合にも，人間の知覚的特性と整合性の良い結果を与えるような処理が期待できる．

従来の研究のうち，高調波選択によって目的音を分離する方法[7],[8]や，ボトムアップ処理に基づく音源分離の試み[9]は，知覚的音源分離を扱った例と見ることもできる．しかし，これら従来の研究においては，知覚的音源分離のもととなる特徴量の扱いは定性的なも

† 東京大学工学部電気工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

表 1 周波数成分の分離知覚または融合知覚の手掛りになり得る要因

周波数成分の高調波関係のずれ	(分離)
周波数成分の立上り時刻のずれ	(分離)
周波数成分の立下り時刻のずれ	(分離)
周波数成分に共通な FM	(融合)
周波数成分に共通な AM	(融合)
継時的なつながり	(分離 / 融合)
音源の方向	(分離 / 融合)
音の記憶	(分離 / 融合)

(分離)：分離知覚を促進し得る要因
(融合)：融合知覚を促進し得る要因

のに限られていた．そこで本論文では，特徴量の扱いに関する定量的なモデルを提案することを主題とする．

知覚的音源分離を計算機によって行うには，人間がスペクトルパターンからどのような基準でそれぞれの知覚的な音に対応する音響的情報を抽出しているのかを調べる必要がある．このとき最も基本的な問題は，ある二つの周波数成分がどのような条件を満たしたときに別々の音として認識されるかという問題である．この問題に関し，人間の音源分離知覚に影響を与える要因としては，音響心理学の分野において表1のようなものが挙げられている[1],[10]．ここでは詳細な議論は省略するが，周波数成分の特徴について見れば，これらの中で音源分離知覚に特に大きな影響を与える特徴は，周波数成分の高調波関係のずれと，周波数成分の立上り時刻のずれである．そこで本論文では，第1段階として，これら二つの特徴に着目する．なお本論文では，周波数が一定の二つの周波数成分があるとき，一方の周波数成分の周波数が他方の周波数の整数倍であるとき，これらの周波数成分は高調波関係にあるというものとする．

音響心理学の分野においては，これら二つの特徴に対していくつかの実験が行われてきているが[11]~[15]，これらはしきい値の測定や聴覚系の処理過程の推察を主眼としている．しかし，本論文のように工学的にシステムを構成するという観点からは，これまでに報告されている実験結果は十分なものではない．この理由は，システムの構成においては複数の要因の有効な統合を考える必要があり，それぞれの特徴を，特徴量と分離知覚の確実性(分離知覚が生じる割合)との関係という観点から評価する必要が生じるためである．そこで本論文では，まず，人間の聴覚系について，周波数成分のもつ特徴と分離知覚の生じる割合との関係を測定し，これを近似する評価関数を得る．次に，評価値を統合するモデルを提案して，評価実験を行う．

2. 知覚的音源分離システムの構成

本章では，本論文で考える音源分離問題の定式化を行った後，知覚的音源分離システムの全体像を簡潔に示すことによって，本論文に述べる工学的モデルがシステム構築においていかなる位置を占めるかを明らかにする．

2.1 音源分離問題の定式化

一般に音響信号を，その性質を表す特徴量(パラメータ)の集合で表すことができるとする．このとき，混合音を表すパラメータの集合をもとに，これに含まれる知覚的な音を表すパラメータの集合を得る問題を，知覚的音源分離問題と考える．本論文では，音響信号の性質を表すパラメータとして特に周波数成分を用いることにする．すなわち，音響信号 $S(t)$ は，L 個の周波数成分 $F_j(t)$ の集合

$$F(t)=\{F_1(t),\ F_2(t),\ \cdots,\ F_L(t)\} \qquad (1)$$

を用いてその性質を表現できるとする．ここで，周波数成分 $F_j(t)$ は，対象とする音響信号のサウンドスペクトログラムにおいて周波数方向の極大点(ピーク)を時間的に接続したものであって，少なくともピークのパワー $p_j(t)$ とピークの周波数 $f_j(t)$ を要素としてもち，

$$F_j(t)=\{p_j(t),\ f_j(t),\ \cdots\} \qquad (2)$$

のように表されるとする．なお，衝撃音や白色雑音などのように，式(2)の周波数成分による性質の表現が必ずしも適切でない音源については，本論文では考慮しないものとする．このとき音源分離問題は，

[問題1] 混合音を分析して得たサウンドスペクトログラム上で，式(2)の形の周波数成分 $F_j(t)$ を抽出する問題(周波数成分の抽出問題)

[問題2] 得られた周波数成分 $F_j(t)$ を，重複を許して，ある個数のクラスタに分類する問題(周波数成分のクラスタリング問題)

と定式化することができる．問題2において，分類されたクラスタが個々の音源に対応している．重複を許すのは，異なる音源が同一の周波数成分を共有する場合(重複周波数成分：shared component)を考慮するためである．重複を許すことは，重複周波数成分を分解すること(decomposition)に対応していると考えることもできる．

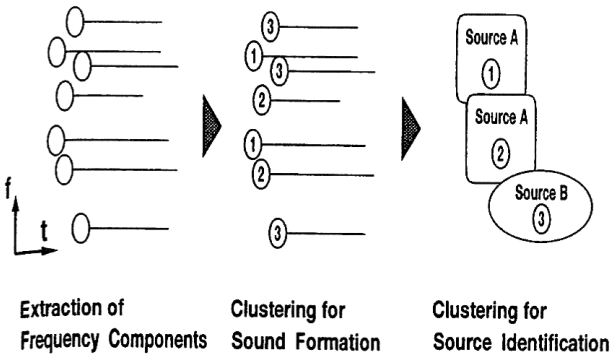


図 1　知覚的音源分離システムの処理の流れ
Fig. 1　A process flow of perceptual sound source separation system.

問題2は，更に2段階に分けて考えることができる．これを図1に示す．第1段階は，人間が一つの音として知覚しやすい周波数成分同士をまとめる操作である．これを単音形成クラスタリング(clustering for sound formation)と呼び，その結果を単音クラスタ(sound cluster)と呼ぶことにする．単音クラスタは，例えば処理の対象として音楽演奏を考えると，一つの音符(note)の音に対応すると考えられる．

第2段階は，単音クラスタについて，同じ音源に由来すると考えられるもの同士をまとめる操作である．これを音源同定クラスタリング(clustering for source identification)と呼び，その結果を音源クラスタ(source cluster)と呼ぶことにする．音楽演奏の場合には，同じ種類の楽器の音をまとめることに相当する．音源同定クラスタリングは，各単音クラスタのもつ特徴の類似や，複数の単音クラスタにまたがる継時的な情報に基づいて行うことが考えられる．

同一の音源クラスタに属する単音クラスタを時間の順に並べることにより，ある音源に由来する音響的情報を抽出することができる．例えば，単音クラスタの基本周波数を抽出することにより，ある楽器の演奏情報を抽出することが可能となる．

2.2　単音形成クラスタリングの処理モデル

単音形成クラスタリングを行うためには，周波数成分間に距離を定義することが必要である．このとき，知覚的音源分離の立場から，人間が別の音として聴く可能性の高さ，すなわち分離知覚の生じる割合に応じて，周波数成分間に距離を定めることを考える．この場合，表1に示すように，周波数成分のもつ特徴に限っても，複数の特徴に基づいて距離を定める必要がある．

そこで，我々は単音形成クラスタリングにおける評価統合モデルを提案する．これは，図2に示すように，まず複数の特徴が独立に評価され，次にその評価値が統合されるとするモデルである．ここで，特徴の評価とは，ある周波数成分にその特徴が存在したときに分離知覚が生じる割合を評価することである．各特徴についての評価値の統合に関しては，**5.**において議論する．

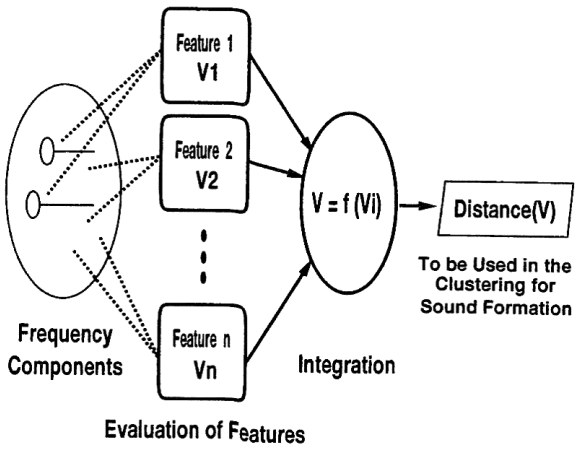


図 2　単音形成クラスタリングの評価統合モデル
Fig. 2　The evaluation-integration model of clustering for sound formation.

このような処理の流れの中で，本論文で扱うのは，単音形成クラスタリングにおいて必要となる，周波数成分間の距離の定義に関する検討である．すなわち，本論文は，スペクトログラム上の特徴と，二つの周波数成分の分離知覚の生じる割合との関係を実験的に考察することを主題とする．**1.**に述べたように，評価する特徴としては，周波数成分の高調波関係のずれと，立上り時刻のずれの2種類を扱う．そこでこれらの特徴の評価関数を得るため，以下の章で聴覚実験を行う．

3.　聴覚実験1：周波数成分の高調波関係のずれと分離知覚の関係

3.1　実験の目的

本実験の目的は，二つの周波数成分について，高調波関係のずれと分離知覚との関係を測定し，これをモデル化することである．なお，本実験では，高調波関係において，特に基本周波数成分(基音)と第2次高調波成分(第2倍音)との関係を考える．

基音と第2倍音の二つの周波数成分からなる音を聴いたとき，一般に，周波数成分の高調波関係のずれが比較的小さい場合には呈示した二つの周波数成分は一つの音として知覚される場合が多いが，ずれが大きくなるにつれて二つの音として知覚される割合が増加

し，ずれがある量を超えると分離知覚の割合は100%に達する．このモデルとして，ここでは，ずれの量に対する分離知覚の割合の測定データに対し，原点を通る直線(但し100%に達したら飽和する)を最小2乗法により当てはめ，これをモデルとして用いることにする．すなわち，モデルは

$$c_h(u)=\begin{cases}-\dfrac{1}{p_-}u & p_-<u<0\\ \dfrac{1}{p_+}u & 0\leqq u<p_+\\ 1 & \text{otherwise}\end{cases} \qquad (3)$$

である．ここで，c_hは，高調波関係のずれによって分離知覚が生じる割合であり，uは，周波数成分の周波数をf_1, f_2 (但し$f_1<f_2$)としたとき，

$$u=\left(\frac{f_2}{2f_1}-1\right)\times 100 \qquad (4)$$

で定義されるパラメータである．すなわち，uは，周波数の高調波関係からのずれを百分率で表したものである．このモデルのパラメータはp_+, p_-であるので，結局これらの値を定めることが，本実験の目的となる．

3.2 方　　法

本実験では，二つの周波数成分からなる音を試料として被験者に呈示する．低い方の成分は200 Hzおよび1,000 Hzの2通りで周波数一定とし，高い方はそれぞれ400 Hzおよび2,000 Hzを中心に公比1.005の等比数列に従って周波数を15段階に変化させた．二つの周波数成分の振幅は等しく，立上りおよび立下りは同時であって，2 dB/msの傾きで対数的に立ち上げ，−2 dB/msの傾きで対数的に立ち下げる．音の継続時間は1,000 msである．この試料をランダムな順序で被験者に呈示し，それぞれが二つの音に聞こえたか一つの音に聞こえたか二者択一(強制選択)の回答を記録した．すなわち，本実験は，二つの音として知覚されたか否かという絶対的判断を被験者に要求し，その判断の揺らぎを含めて分離知覚の生じる確実さとして測定するものである．

被験者は3名とした．被験者1名当り同一試料の呈示する回数(繰返し回数)は，少なすぎれば安定した結果が得られにくいが，あまり多すぎても，実際上は必ずしも測定結果の分散は減少しない．そこで予備実験を行ったところ，本実験の場合には，繰返し回数が6回から8回程度で得られるパラメータ値の分散の減少は頭打ちとなることがわかった．この結果に基づいて，被験者1名当り8回の繰返しを1セットと定めた．本章の実験では，休息を挟んで2セットの実験を行った．

試料は計算機上であらかじめ作成し，これをランダムな順序で選択再生してDAT(Digital Audio Tape)レコーダに録音した．試料の呈示間隔は約5秒おきとした．被験者には，録音した試料を静かな部屋においてヘッドホンにより両耳に呈示した．音圧はおよそ65 dB SPLとした．また被験者には，事前に実験課題に習熟させ，安定した判断が可能となるように配慮した．また本実験では，いわゆる分析的な態度で試料を聴取することは適切でないので，通常の聴取態度において二つの音が知覚された場合に「分離」と回答するよう，被験者に教示した．

3.3 結果と考察

本実験の結果を図3に示す．これは，3名の被験者についての結果を合わせたものである．本実験では，2成分のうち高い方の周波数成分の周波数を変化させたが，図3の横軸のnは，周波数をずれがないときから何段階ずらしたかを表している．これは，実際の周波数f_1, f_2($f_1<f_2$)を次式によってnに変換したものである．

$$n=\frac{|\log f_2-\log(2f_1)|}{\log 1.005} \qquad (5)$$

また，本実験によって得られたパラメータp_+, p_-の値を表2に示す．

表2の値について，「条件」(基本周波数およびずれの

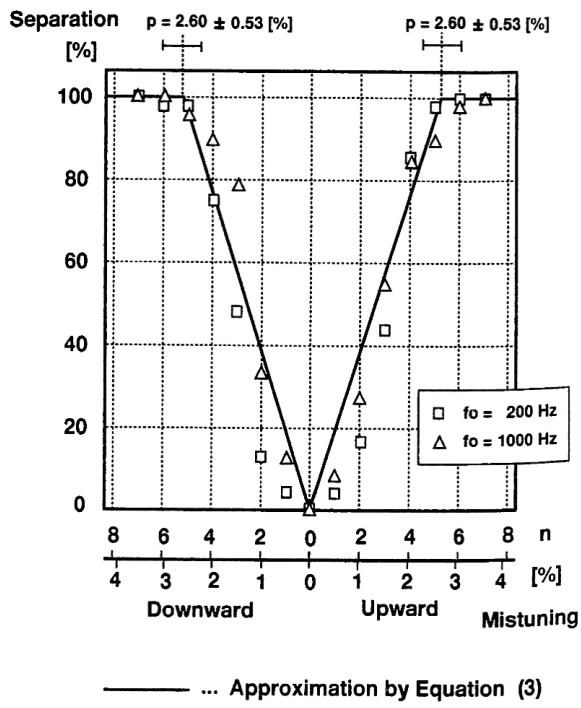


図3　聴覚実験1の結果
Fig. 3　The result of Experiment 1.

方向)と,「被験者」の二つの要因について,繰返しのある2元配置の分散分析をった結果を表3に示す.表3によれば,被験者については危険率1%で有意であるが,条件(基本周波数およびずれの方向)および条件と被験者の交互作用に関しては有意とならなかった.

このことから,一般に基本周波数やずれの方向がパラメータ p_+, p_- の値に無関係であるとの結論を導くことはできない.しかし,ここでは,基本周波数およびずれの方向については,評価関数のパラメータとして含めないことにする.また,評価関数は,個人別ではなく3名の被験者の結果を合わせたものを考える.このとき,周波数成分の高調波関係のずれに対する評価関数は,式(3)において $p=p_+=p_-$ としたとき,表2の値の全平均値をもって,

$$p=2.60 \quad [\%] \qquad (6)$$

と定めることができる.また,p の値の95%信頼区間を求めると,

$$[2.07\%,\ 3.13\%] \qquad (7)$$

となる.図3には,これらの結果をまとめて示した.

また,式(3)と図3の実測データとの相関係数 R を求めたところ,f_0=200Hz と f_0=1,000 Hz のデータについて,それぞれ

$$R_{200}=0.98 \qquad (8)$$

$$R_{1000}=0.97 \qquad (9)$$

となった.ここで,相関係数 R の定義は,データ y_i および Y_i に対して

$$R=\frac{\sum_{i=1}^{N}(y_i-\bar{y})(Y_i-\bar{Y})}{\sqrt{\left\{\sum_{i=1}^{N}(y_i-\bar{y})^2\right\}\left\{\sum_{j=1}^{N}(Y_j-\bar{Y})^2\right\}}} \qquad (10)$$

である.但し $\bar{y}$, $\bar{Y}$ は,それぞれ y_i, Y_i の平均値を表す.

表 2　聴覚実験1で求められた p_+, p_- の値

	被験者1	被験者2	被験者3
F_0 = 200Hz	2.85	3.18	2.97
p_+ (Upward)	1.02	3.18	2.42
F_0 = 200Hz	2.93	3.39	3.57
p_- (Downward)	2.09	2.89	2.42
F_0 = 1000Hz	1.76	3.26	3.62
p_+ (Upward)	1.71	3.13	2.56
F_0 = 1000Hz	1.65	2.42	3.45
p_- (Downward)	1.71	1.65	2.59

表 3　表2の値分散分析によって得られた F_0 の値

	F_0
条件 (A)	1.20
被験者 (B)	6.95 (危険率 1% で有意)
A × B	0.82

(「条件」とは、基本周波数およびずれの方向を意味する)

4. 聴覚実験2：周波数成分の立上り時刻のずれと分離知覚の関係

4.1 実験の目的

本実験の目的は,二つの周波数成分について,立上り時刻のずれと分離知覚との関係を測定し,高調波関係の場合と同様の直線近似式におけるパラメータを定めることである.なお,予備的な検討により,周波数成分の周波数と立上りの傾きが結果に影響することがわかったので,基本周波数については2通り,立上りの傾きについては3通りに変えて測定する.

4.2 方　　法

被験者に呈示する試料は,二つの周波数成分からなる音である.試料の概念図を図4に示す.低い方の成分は200 Hzおよび1,000 Hzの2通り,高い方はそれぞれ400 Hzおよび2,000 Hzとして,いずれも周波数は一定とした.定常部分における二つの周波数成分の振幅は等しいが,立上り時刻を,低い方の周波数成分を基準として前後に100 msとなるまで15通りに変化させた.ここで,立上り時刻とは,周波数成分の振幅が増加し終わって振幅一定となる点の時刻とした.また,立上りの傾きを1 dB/ms, 2 dB/ms, および5 dB/msの3通りに変えた.立下りは同時的とし,傾きは −2 dB/msとした.音の継続時間は,基準となる低い方の周波数成分の立上りから立下りまでの時間が1,000 msとなるようにした.この試料をランダムな順序で被験者に呈示し,それぞれが二つの音に聞こえた

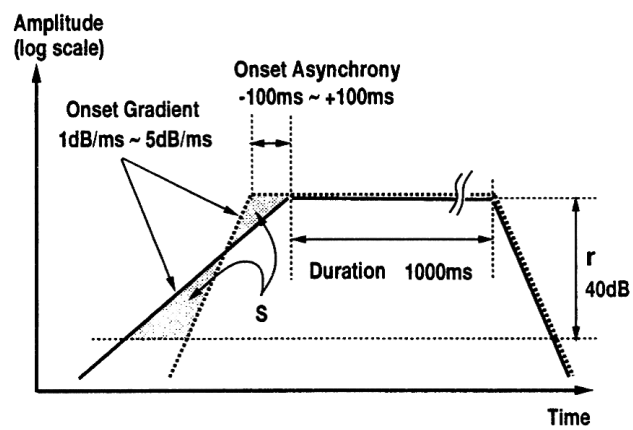


図 4　聴覚実験2で呈示した試料の概念図
Fig. 4　Stimuli presented in Experiment 2.

表 4　聴覚実験 2 で呈示した試料の種類

立上りの時間差 [ms]	-100	-70	-50	-40	-30
	-20	-10	0	10	20
	30	40	50	70	100
立上りの傾き [dB/ms]	5	2	1		
基本周波数 [Hz]	200	1000			

か一つの音に聞こえたか二者択一(強制選択)の回答を記録した．被験者は 3 名とし，1 名当り 8 回の実験を行った．

用いた試料の種類を表 4 にまとめる．表のうち，立上りの傾きは，三つの値の中から高低二つの周波数成分に値を割り当てるので，割当て方は 9 通りある．従って，試料の種類の合計は $15\times9\times2=270$ 種類となる．なお試料の作成や呈示の仕方は，前章の場合と同様とした．

4.3 結果と考察

本実験の結果の例を図 5 に示す．これは，200 Hz(立上りの傾き 5 dB/ms)の成分と 400 Hz(立上りの傾き 1 dB/ms)の成分とを試料としたときの結果である．これは，3 名の被験者についての結果を合わせたものである．表 4 からわかるように，実際にはこのような図が 18 枚得られるが，紙面の都合により 1 例のみを示した．

さて，本実験では，二つの周波数成分の立上り時刻の差をパラメータとして変化させて試料を作成したが，結果の分析にあたっては，立上り部分の様子を詳細に検討する必要がある．すなわち，ここでは立上り時刻を図 4 のように定めているが，図 5 に示されるように，二つの周波数成分の立上りの傾きが異なる場合，立上り時刻が同時であるときに最も分離知覚が生じにくくなるわけではない．

このように，立上りの傾きが異なる場合には，他方の周波数成分より振幅包絡が突出した部分の大きさ，すなわち図 4 の網掛け部分の面積 S を評価の基準とする方が自然である．そこで，ここでは立上り時刻を面積 S に変換してモデル化することにする．早く立ち上がる方の周波数成分の傾きを g_1，他方の周波数成分の傾きを g_2 とし(同時に立ち上がるときはどちらを g_1 としてもよい)，傾き g_1 の周波数成分を基準にしたときの，他方の成分の立上りの相対時刻を t とすれば，面積 S は

(i)　$g_1\leqq g_2$ のとき

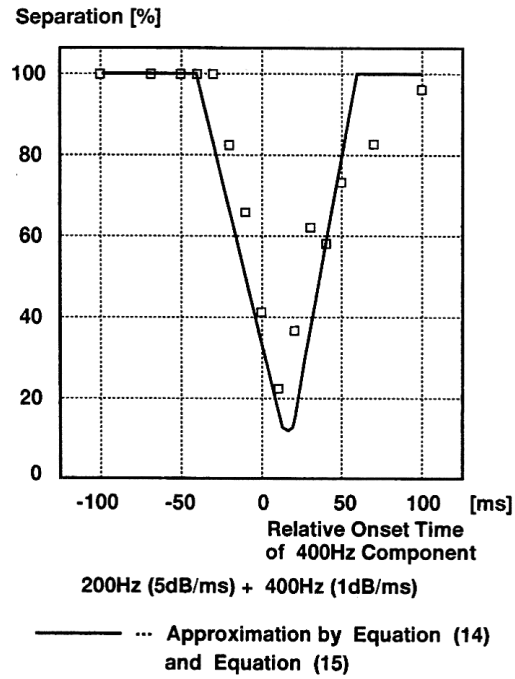


図 5　聴覚実験 2 の結果の一例
Fig. 5　An example of the result of Experiment 2.

$$S=r(t+t_a) \tag{11}$$

(ii)　$g_1>g_2$ のとき

$$S=\begin{cases} r(t-t_a) & t\geqq 2t_a \\ r\left\{\dfrac{1}{2t_a}(t-t_a)^2+\dfrac{t_a}{2}\right\} & \text{otherwise} \end{cases} \tag{12}$$

但し

$$t_a=\frac{|g_1-g_2|r}{2g_1g_2} \tag{13}$$

となる．ここで r は，面積計算のための基準となる量であり，ここでは周波数成分の周波数によらず $r=40$[dB] とした．

式(11)および式(12)によって時刻から面積への変換を施したとき，実験結果の直線近似式は，次のように考えることができる．

$$c_0(S)=\begin{cases} \dfrac{1}{S_p}S & S<S_p \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

実験結果から(14)の S_p を求め，これを早く立ち上がる方の周波数成分の周波数 f_1 と立上りの傾き g_1 に着目してまとめたものを表 5 に示す．表で，例えば左上の欄では，f_1 が 200 Hz で，g_1 が 1 dB/ms であるが，この場合遅く立ち上がる成分の周波数 f_2 は 400 Hz であり，その立上りの傾き g_2 は，表 4 にあるように

表 5　聴覚実験 2 で求められた S_p の値

表の数値は S_p [dB·s]。
f_1：早く立ち上がる周波数成分の周波数
g_1：早く立ち上がる周波数成分の立上りの傾き

f_1	g_1		
	1 dB/ms	2 dB/ms	5 dB/ms
200Hz	3.00	2.13	1.66
	2.80	1.90	1.60
	2.64	2.00	1.85
400Hz	2.17	1.55	1.02
	1.87	2.05	1.29
	1.58	1.56	1.54
1000Hz	2.29	1.27	0.53
	1.49	0.62	0.67
	1.56	1.15	0.66
2000Hz	2.11	1.29	0.60
	1.28	1.16	0.70
	0.76	1.24	0.64

表 6　S_p の分散分析によって得られた F_0 の値

	F_0
f_1	12.7 (危険率 1% で有意)
g_1	13.4 (危険率 1% で有意)
$f_1 \times g_1$	0.62
f_2	3.97 (危険率 5% で有意)
g_2	0.26
$f_2 \times g_2$	0.11

ここで、
f_1 ： 早く立ち上がる周波数成分の周波数
g_1 ： 早く立ち上がる周波数成分の立上りの傾き
f_2 ： 遅く立ち上がる周波数成分の周波数
g_2 ： 遅く立ち上がる周波数成分の立上りの傾き

1dB/ms, 2 dB/ms, および 5 dB/ms の 3 通りの場合について実験している．表 5 では，これらに対応する三つの値を，各欄に上から順に示した．

表 5 について，繰返しのある 2 元配置の分散分析を行った結果を，表 6 に示す．これによれば，f_1 と g_1 については，いずれも危険率 1%において有意となったが，これらの交互作用は危険率 5%においても有意とならなかった．また，f_2 と g_2 についても同様に分散分析を行ったところ，f_2 が危険率 5%において有意となったものの，g_2 およびこれらの交互作用については，危険率 5%においても有意とならなかった．

そこで，危険率 1%において有意となった要因に着目し，式(14)において，S_p は f_1 と g_1 をパラメータとしてもつと考える．ここで，その構造モデルを

$$S_p=\frac{a}{f_1}+\frac{b}{g_1}+c \tag{15}$$

と仮定して重回帰分析を行ったところ，

$$\begin{cases} a=250 \\ b=1.11 \\ c=0.317 \end{cases} \tag{16}$$

となった．重相関係数は 0.89 であった．以上により，周波数成分の立上り時刻のずれを評価する評価関数が式(14)および式(15)として定められたが，この評価関数と本実験における全測定結果との相関係数 R を求めたところ，

$$R=0.92 \tag{17}$$

であった．

5. 確実性の統合モデル

前章までに，周波数成分のもつ二つの特徴の評価値を求める評価関数が求められたが，本章では，これらの評価値を統合する手法について考察する．

このように，独立に評価された複数の評価値を統合する方法としては，最大値，代数積，平均値などに代表されるさまざまな演算が考えられる．しかし，単に統合するというだけではなく，何らかの合理的な根拠を見出し得る演算を考える必要がある．意味付けが可能な演算操作としては，ベイズの定理に基づく方法と，Dempster の結合規則による方法を挙げることができる．すなわち，前者では，ある特徴が存在したときに分離知覚が生じる事後確率を求るという意味付けが可能であり，後者では，特徴の評価値を，分離知覚を支持する「信用」ととらえていることになる．

このうち，ベイズの定理による方法では，分離知覚を促進する特徴が全く見られないとき，その余事象である融合知覚が起こる可能性を 100%支持することになる．しかし実際には，例えば周波数成分が完全な高調波関係にあっても，立上り時刻が異なれば分離知覚が促進されるのであり，分離知覚を促進する特徴がないことが直接融合知覚を促進すると考えるのは不自然である．

そこで本論文では，分離知覚を促進する特徴が見られないことは，分離知覚を支持する「信用の欠如」であって「不信用」ではないと考え，これらを区別して扱うことのできる Dempster の結合規則による方法を用いることを考える．Dempster の結合法則では，ベイズの定理に基づく方法と異なり，観測情報を基本確率関数

に変換するための一般的な方法がないことが問題とされている[19]が，本論文の方法は，基本確率関数を実験的手法により直接求めることに相当している．

Dempsterの結合規則は，一般に，次のようなものである．

$$m(A_k)=\frac{\sum_{A_{1i}\cap A_{2j}=A_k} m_1(A_{1i})m_2(A_{2j})}{1-\sum_{A_{1i}\cap A_{2j}=\phi} m_1(A_{1i})m_2(A_{2j})} \quad (A_k\neq\phi) \tag{18}$$

これは，A_{1i}，$A_{2j}(i,j=0,1,2\cdots)$なる焦点要素(focal element)につき，独立した証拠から得られた基本確率m_1およびm_2があったとき，統合された確実性尺度は$m(A_k)$となることを主張するものである[20]．

本論文における式(18)の具体的な適用法は，次のとおりである．式(18)において，A_1およびA_2は，それぞれ二つの周波数成分の分離知覚および融合知覚を表すものとする．ここでは分離知覚に着目し，A_1のみを考える．また，$m_1(A_{11})$と$m_2(A_{21})$は，それぞれ高調波関係のずれと立上り時刻のずれに基づく分離知覚の割合であって，それぞれ式(3)のc_hおよび式(14)のc_oであると考える．すなわち，

$$\begin{cases} m_1(A_{11})=c_h \\ m_2(A_{21})=c_o \\ m_1(\{A_{11},A_{12}\})=1-c_h \\ m_2(\{A_{21},A_{22}\})=1-c_o \end{cases} \tag{19}$$

となる．従って，式(18)より，統合された分離知覚の割合は

$$m(A_1)=1-(1-c_o)(1-c_h) \tag{20}$$

となる．

単音形成クラスタリングは，この値を周波数成分間の距離とみなすことによって行うことができる．我々は既に，音源分離の実験システムの実装を行っており，その中で単音形成クラスタリングのアルゴリズムについても具体的な検討を行っている．しかし，本論文の主題に照らし，単音形成クラスタリングの具体的なアルゴリズムの適用およびその評価等については，稿を改めて報告する．

6. 統合モデルの評価

評価値の統合において，前章ではDempsterの結合規則を用いることの妥当性を定性的に述べたが，本章では，簡単な聴覚実験(これを聴覚実験3とする)を行って，聴覚実験結果と，Dempsterの結合規則による処理結果およびベイズの定理に基づく処理結果との相関係数を求め，これらを比較することによって前章に述べたモデルの妥当性を確認する．

聴覚実験で被験者に呈示する試料は，二つの周波数成分からなる音である．低い方の成分は200 Hzおよび1,000 Hzの2通りとし，高い方はそれぞれ400 Hzおよび2,000 Hzを中心に公比1.005の等比数列に従って周波数を13段階に変化させた．定常部分における二つの周波数成分の振幅は等しいが，立上り時刻を，低い方の周波数成分を基準として前後に70 msとなるまで13通りに変化させた．各成分の立上りの傾きは2 dB/msとした．また立下りは同時的とし，傾きは−2 dB/msとした．音の継続時間は，基準となる低い方の周波数成分の立上りから立下りまでの時間が1,000

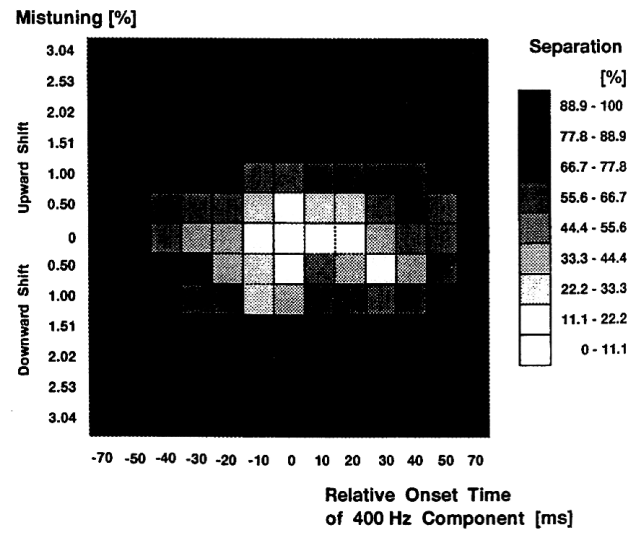


図 6 聴覚実験3の結果(f_0=200 Hzのとき)
Fig. 6 The result of Experiment 3 (f_0=200 Hz)

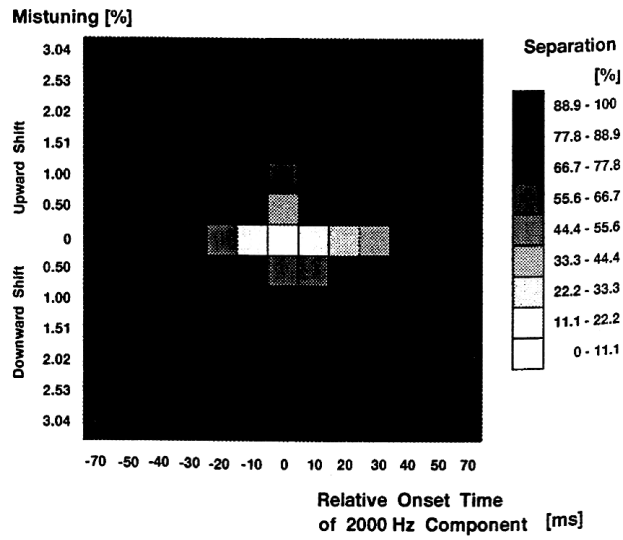


図 7 聴覚実験3の結果(f_0=1000 Hzのとき)
Fig. 7 The result of Experiment 3 (f_0=1000 Hz).

表 7　聴覚実験 3 の結果とモデルによる統合の結果との相関係数

f_0	式 (20) の場合 (Dempster)	式 (21) の場合 (Bayes)
200 Hz	0.84	0.76
1000 Hz	0.75	0.68

ms となるようにした．この試料をランダムな順序で被験者に呈示し，それぞれが二つの音に聞こえたか一つの音に聞こえたか二者択一(強制選択)の回答を記録した．被験者は 2 名とし，1 名当り 8 回の実験を行った．

一方，モデルを用いる場合については，各特徴の評価値を式(20)によって統合した場合と，ベイズの定理に基づいて統合した場合について計算した．ここで，ベイズの定理を用いた場合の統合を式(20)に対比して書くと，

$$m(A_1)=\frac{c_h c_o}{c_h c_o+(1-c_h)(1-c_o)} \qquad (21)$$

となる．この式において，c_h と c_o の一方が 0 で他方が 1 のときは値が不定となるので，この場合には $m(A_1)=0.5$ とする．

聴覚実験 3 の結果を図 6 および図 7 に示す．また，聴覚実験 3 の結果と式(20)および式(21)の統合による値との相関係数を求めると，表 7 のようになった．式(21)の場合に，式(20)に比べて相関値が低いのは，一方の特徴の評価値が 0 で他方が 1 の場合に主な原因があると考えられる．すなわち，図 6 および図 7 の実験結果では，一方の特徴が確実な分離知覚を生じさせる場合には，他方の特徴において分離知覚が促進されなくとも，ほぼすべての場合に分離知覚が生じていることがわかる．このことから，式(21)に比較して式(20)の方が聴覚実験 3 の結果をより良く表現していると言うことができる．

7. むすび

本論文では，知覚的音源分離システムの構築を念頭に，単音形成クラスタリングに着目した．単音形成クラスタリングにおいては，二つの周波数成分がいかなる特徴をもったときに分離知覚を生じるかを定めることが最も基本的な問題である．本論文では，まず，単音形成クラスタリングの処理モデルとして図 2 の評価統合モデルを提案した．次に，聴覚実験の結果をもとに，二つの周波数成分の分離知覚に関するスペクトログラム上の 2 種類の特徴の評価関数を導いて，聴覚実験結果との対応を示した．更に，評価値の統合を行う手法として Dempster の結合規則を用いる方法が妥当であることを指摘し，実験によってこれを確認した．これにより，特徴量と分離知覚の生じる割合との関係が定量的にモデル化された．本モデルの与える分離知覚の生じる割合は，単音形成クラスタリングにおいて，周波数成分間の距離として用いることができる．

これまでにも，Brown らによって，周波数成分の高調波関係および立上りと立下りの同時性という特徴を用いた音源分離システムが提案されている[9]．しかし，Brown らのシステムでは，特徴の評価における評価値の意味付けがあいまいであって，人間の知覚的特性との対応という観点からの検討も十分でない．また統合においても，高調波関係の評価値に対して，周波数成分の同時性が存在する場合に定数を加算するという簡易的な方法が用いられている．これに対し本論文は，実験の条件を限定した上での議論として，単音形成クラスタリングに用いる距離の定義における，複数の要因の定量的な評価と統合の方法を考察したものである．

本論文は，最も基本的な場合として，二つの周波数成分のみが存在する場合について考察した．聴覚実験は，一方の周波数成分の周波数が他方の 2 倍になっている場合を基準にして行った．また，二つの周波数成分の振幅が等しい場合についてのみ測定し，音の継続時間は一定として測定した．従って，一般の単音形成クラスタリング処理を実現する上では，本論文に述べた処理モデルに加えて，三つ以上の周波数成分や，振幅が異なる周波数成分の取扱い，音の継続時間の影響等について，なお考慮する必要がある．これらの点に関する検討を含め，知覚的音源分離の実験システムの具体的な構成および評価実験について，今後報告する予定である．

文　献

(1) Bregman A. S.: "Auditory Scene Analysis", MIT Press (1990).

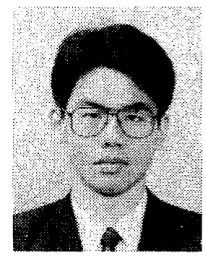
(2) Mitchell O. M. E., Ross C. A. and Yates G. H.: "Signal processing for a cocktail party effect", J. Acoust. Soc. Am., **50**, 2, pp. 656-660 (1971).

(3) 柳田益造，角所　収，植田章彦，野村康雄："一般逆行列を用いたカクテルパーティ効果の知覚的検討"，信学技報，**EA80**-69 (1981).

(4) Flanagan J. L., Johnston J. D., Zahn R. and Elko G. W.: "Computer-steered microphone arrays for sound transduction in large room", J. Acoust. Soc. Am., **78**, 5, pp.

1508-1516(1985).
(5) 永田 仁, 安倍正人, 城戸健一：“多数センサによる音源波形の推定”, 日本音響学会誌, **47**, 4, pp. 268-273(1991).
(6) 長束哲郎, 才脇直樹, 井口征士：“異種楽器を対象とした採譜システム”, 信学'92春大, D-499.
(7) Parsons T. W.: “Separation of speech from interfering speech by means of harmonic selection”, J. Acoust.Soc. Am., **60**, 4, pp. 911-918(1976).
(8) Nehorai A. and Porat B.: “Adaptive Comb Filtering for Harmonic Signal Enhancement”, IEEE Trans. on ASSP, **34**, 5, pp. 1124-1138(1986).
(9) Brown G. J. and Cooke M. P.: “A Computational Model of Auditory Scene Analysis”, In Proceedings of ICSLP 92, pp. 523-526(1992).
(10) Moore B. C. J.: “An Introduction to the Psychology of Hearing, Third Ed.”, Academic Press(1989).
(11) Moore B. C. J., Peters R. W. and Glasberg B. R.: “Thresholds for the detection of inharmonicity in complex tones”, J. Acoust. Soc. Am., **77**, 5, pp. 1861-1867(1985).
(12) Moore B. C. J. and Glasberg B. R.:“Thresholds for hearing mistuned partials as separate tones in harmonic complexes”, J. Acoust. Soc. Am., **80**, 2, pp. 479-483(1986).
(13) Hartmann W. M., McAdams S. and Smith B. K.: “Hearing a mistuned harmonic in an otherwise periodic complex tone”, J. Acoust. Soc. Am., 88, 4, pp. 1712-1724(1990).
(14) Darwin C. J. and Ciocca V.: “Grouping in pitch perception: Effects of onset asychrony and ear of presentation of a mistuned component”, J. Acoust. Soc. Am., **91**, 6, pp. 3381-3390(1992).
(15) Ciocca V. and Darwin C. J.: “Effects of onset asynchrony on pitch perception: Adaptation or grouping?”, J. Acoust. Soc. Am., **93**, 5, pp. 2870-2878(1993).
(16) McAdams S.: “Segregation of Concurrent sounds. I: Effects of frequency modulation coherence”, J. Acoust. Soc. Am., **86**, 6, pp. 2148-2159(1989).
(17) Bregman A. S., Levitan R. and Liao C.: “Fusion of auditory components: Effects of the frequency of amplitude modulation”, Perception and Psychophysics, **47**, 1, pp. 68-73(1990).
(18) Massaro D. W.: “Speech Perception by Ear and Eye: A Paradigm for Psycholgical Inquiry”, Lawrence Erlbaum Associates(1987).
(19) 松山隆志：“Dempster-Shaferの確率モデルに基づくパターン分類—観測情報からの信念の形成と仮想信念空間を用いた信念の統合”, 信学論(D-II), **J76-D-II**, 4, pp. 843-853(1993-04).
(20) 石塚 満：“Dempster & Shaferの確率理論”, 信学誌, **66**, 9, pp. 900-903(1983-09).

(平成5年8月13日受付, 12月9日再受付)



柏野 邦夫

平2東大・工・電子卒. 平4同大大学院修士課程了. 現在同大学院博士課程在学中. 音響的情報を対象とする信号処理および知識処理に興味をもつ. 情報処理学会, 人工知能学会, 日本音響学会, IEEE各会員.

田中 英彦

昭40東大・工・電子卒. 昭45同大大学院博士課程了. 同年東大・工・講師. 昭46同大助教授, 昭62同大教授, 現在に至る. この間昭53～54ニューヨーク市立大客員教授. 工博. 計算機アーキテクチャ, 並列推論マシン, 帰納推論, オブジェクト指向計算システム, 分散処理, CAD等の研究に従事. 著書「非ノイマンコンピュータ」,「情報通信システム」, 共著書「計算機アーキテクチャ」,「VLSIコンピュータI, II」,「ソフトウェア指向アーキテクチャ」. 情報処理学会, 人工知能学会, 日本ソフトウェア科学会, IEEE, ACM各会員.