音楽の音響空間情報論

Spatiotemporal Information Theories on Music

伊東　乾*
Ken ITO

「視聴覚情報」という言葉が頻繁に用いられる。これは一見すると「視覚」と「聴覚」という人間の感官に即したもののように見える。だが実のところオーディオ-ヴィジュアルという観点はユーザというよりハードウエア、ソフトウエアなどシステムの側から考えられた面が小さくない。

いま仮に、レーシングゲームのような形で視覚のみに仮想現実環境が与えられたとしよう。左右への揺れなどがあったとしても、観測者が受容する空間感覚、例えば定位感が受ける影響には限界がある。

生物の内耳は進化の過程で平衡感覚器、前庭感覚器など定位器として始まっており、いわゆる可聴音の聴取よりも感覚の起源は古い。遊園地にされる体感式の遊具を想起すれば明らかなように、こうした器官に刺激を受けると人は容易に目を回す。また視覚だけに頼った仮想現実感の内耳に与える影響は限られている。

翻って「聴覚」念頭で与えられる響きであっても、そのほんの少しの変化が、定位感に影響を与える場合がある。高速で至近距離を走り抜けるバイクの音などが提示されると骨格筋を含む反射的な身体運動の惹起が普通に見られる。つまり体をよけようとする反応など誘発されることがしばしばである。空間と身体を念頭に置くとき、必ずしも「視聴覚メディア」は「視聴覚」だけに関係するわけではない。

本稿では、もっぱら音楽の観点から、聴覚と空間感覚を横断するべく私たちの研究室で構築した、情報認知の基礎的な数理と実測、解析の枠組みを概説する。これらはまた、時間の感覚や意識が諸知覚を束ねる「バインディング問題」など、認知のより深い問題への導入にもなっている。

* 東京大学大学院情報学環

キーワード：音楽、言語認知、正弦波素分解、脳機能可視化、時間に依存する相関解析、視聴覚コンテンツ

導入1　情報エントロピーの定義

あるイヴェントEが生起する確率が$P(E)$であるとしよう。観測者Observerがその事象が起きた事を観測measureしたとき、そこで受け取る情報エントロピーあるいは情報量$I(E)$を

$$(1)\quad I(E) = \log\frac{1}{P(E)} = -\log P(E)$$

と定義する。例えば表裏が等確率で出るコインをカップの中に投じ、よく振ったあと机の上に伏せたとしよう。一回のコイン投げで表が出る確率も裏が出る確率も50%=0.5とする。カップを伏せた状態のままでは中のコインがどのような状態にあるか観測者は知ることが出来ない。その状態では観測者が得る情報量は定義できない。

ここでカップを取り除き中のコインの状態を確認したとき初めて観測者は情報を得ることになる。対数の低を10として（1）を計算すれば

$$I(E) = \log\frac{1}{0.5} = -\log_{10}0.5 = 0.30103\ldots$$

ということになる。

さて、いまカップを取り除いたとして、中をろくろく見もせずに「そんなの見なくたってカップの中はどうせコインは表か裏かのどっちかしかないさ」と嘘吹く人がいたとしよう。裏の確率は0.5、表の確率も0.5、つまり裏か表かどちらかである確率は1（コインが立っているとか回転しているといった場合は考えない）になる。この状態で （1）を計算してみると

$$I(E) = \log 1 = 0$$

となり、この人の言う話に情報量～情報としての価値がないことが定量的に示される。この定義は中々悪くないかもしれない。ここで対数の底を2（以下のように添字2をつける）とすると

$$\mathrm{I}_2(E) = \log_2\frac{1}{0.5} = \log_2 2 = 1$$

となる。オン・オフ二つの状態だけがあるスイッチのように確率2分の1で生起する現象を観測者が受けとったとき、そこで受け取られる情報は2進数binary digit で表現すれば1 binary digit となる。binary digitはしばしば略され1bit と呼ばれる。これがクロード・シャノンによる情報エントロピーあるいは情報量1ビットの定義である[1]。離散的なdigit 度数で問題を取り扱うシステムをdigital systemと呼ぶことにしよう。

シャノンの定義で重要な事は　1　この定義に従う限り「情報」は本質的に確率量で「生起するかもしれない現象」を対象とする事、そして　2　その事実を**観測者が受容～認知**したとき初めて定義されるという2点である。

導入2　人間の聴覚路のあらまし

そこで、音楽や音響の「情報」を観測者が「受容～認知」するプロセスとして聴覚のあらましを整理し直しておこう。

空気中を伝播する空気の縦波は人の耳たぶpinna等で反射され、外耳道ear canalを通って鼓膜ear drum, tympanic membraneを押す。鼓膜の振動は中耳で三つの耳小骨ossiclesを通じて調波されて内耳の前庭vestibule窓oval windowを押す。内耳は中が隔てられた袋でリンパ液lymphが詰まっており、外界の空気の振動は内耳内で液体の振動に変換され蝸牛管cochlearへと導かれる。蝸牛管は円錐状の管で中に張られた基底膜basilar membrane 上に有毛細胞hair cells と呼ばれる知覚細胞が分布し、入り口近くが可聴域上限の高周波、奥に行くにつれて周波数が下がり、最深部が可聴域下限の低周波を探知して、その振幅に応じた神経発火のインパルスを、蝸牛神経を通じてより上位の中枢神経系へと送りだしている。

蝸牛は鼓膜から中耳を介して内耳に齎された物理的な振動を、周波数成分ごとに分解する一種のスペクトル・アナライザーとして機能しており、蝸牛神経発火のインパルスとなることで、物理的な刺激信号がある種のアナログーデジタル変換を経てニューラルなデジタル情報になることに注意する必要がある。

空気や水など媒質の振動としての物理的音波と、聴き手である人間が意識し聴取する認知的音像とは、さまざまな点で大きくことなる。シャノンの情報エントロピーの定義を思い出すなら、物理的な音波が担うシグナルはあくまで聴き取られる可能性のある通牒messageの列であって、それが聴き手＝観測者によって受容されて、始めて認知・聴取の情報過程の基本サイクルが閉じ、議論が始まることに注意する必要がある。平たくいうなら「馬の耳」に向かって唱えられた念仏は、messageとしては念仏であっても、観測者（「観測馬」）にとってはありがたい念仏としての意味を持たないことになる。

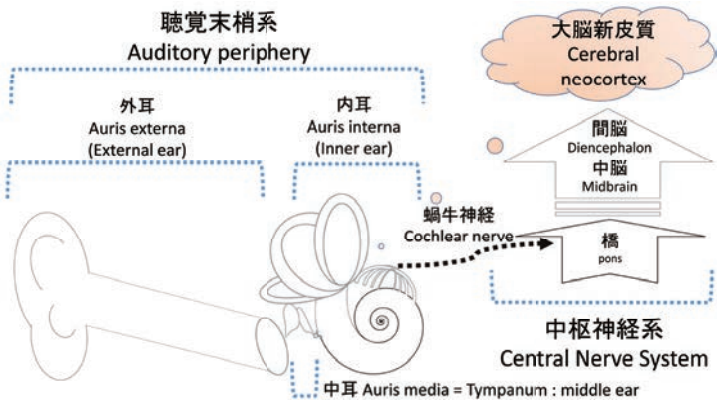


Fig.1 ヒト聴覚末梢と中枢での認知観測のあらまし

§1　認知的音像としての音声言語の脳内創発

ここまで準備した上で、物理的音波によって齎されるメッセージが認知受容されるまでのあらましをシンプルに示す実例を示す。

ヒト内耳では、鼓膜から齎された振動は円錐状の蝸牛管内に導入されると、管内で異なる深さの場所に配位された有毛細胞が同時並行的に周波数成分に対応する神経発火のシグナルを中枢側に送り出すことが出来る。こうしたメカニズムは20世紀前半、フォン・ベケシーらによって明らかにされた[2]。

これに対してフーリエ級数では、短い時間長さT_0ごとのフレームにサンプルを区切り、これを周期関数と見立て、T_0の逆数ω_0を基音とする整数次倍音を基底としてフレームサンプルを展開し、高調波を得るため、内耳と同じ意味で同時並行的な演算を行っているわけではない。すなわち、短時間音源信号サンプル列を時間tの関数$s_n(t)$と書くなら、個々のフレームサンプルについて

(2)　$$s_n(t) = \sum_{k=0}^{\infty} a_k \cos 2\pi k\omega_0 t$$

のように短時間正弦波素にスペクトル分解することが可能である。ここでnはフレーム番号を表す。この短時間の断片をMcAuleyらに倣ってsinusoidと呼ぶことにする[3]。ここでフレーム長T_0はフーリエ級数の計算のために人為的に導入されたもので、生物の内耳には原理的に無関係な量であることを確認しておこう。

さて、仮にこの単一の短時間サンプルが5ミリ秒、10ミリ秒以下の短い長さであるなら、ヒトの聴覚はそれを一つながりの音として認識することは出来ず、私たちの耳はこれらを瞬間的なノイズ以上のものとして知覚することができない。短時間sinusoidそれ自身を私たちは明確な音像として聴き取ることが出来ない。

そこで単一フレームの音を要素の一部として、言語音声の一部を構成する音素、あるいは音楽的な音響としてこれを認識できる音を人為的に合成するためには、多数のフレームを繋げて数百ミリ秒以上の連続した可聴断片audible fragment $Ti(t)$を、係数a_kをa_{nk}と書き直すことにして

(3)　$$T_i(t) = \sum_{n=1}^{N} s_n(t) = \sum_{n=1}^{N}\sum_{k=0}^{\infty} a_{nk} \cos 2\pi k\omega_0 t$$

のように繋げなければならない。この連結演算を考えよう。いま上式の右辺でkに関する和を求めず、第nフレームの特定のk番目の正弦波素と第n+1フレームのj番目の正弦波素に注目して、これら二つの正弦波素を繋げることを考えよう。二つの正弦波の連結をT(t)と書くことにし、時間変数tに関する煩瑣を避けるため記号∪を用いて

(4)　$$T(t) = a_{n,k} \cos 2\pi k\omega_0 t \quad \cup \quad a_{n+1,j} \cos 2\pi j\omega_0 t$$

と書くことにする。

§1-1　超スペクトル弦による音声・音響の分解

いま単純に上記のようにsinusoidを併置すると、二つのsinusoid片の位相は不連続であるため、計算に起因するアーティファクトとして耳障りなノイズの発生が避けられない。内耳内では有毛細胞による並列処理のため短時間フーリエ変換に伴うノイズなどは一切発生しないので位相を連結する演算処理が必要となる。関数fにこの平滑演算をJ [f]と記し、T(t)に平滑演算を施した結果を

(5)　$$J\ [T(t)] = Th(t)$$

と記すことにすれば、Th(t)はsinusoid二つが滑らかに連結された、同時に一つの振動数のみをもつ連結正弦波素となる。この操作を繰り返し、数百ミリ秒以上など十分な長さを持てば、ヒトの聴覚はこれを連続的に変化する、個々の時間では単一の音程をもつ刺激として認知することが出来る。このTh(t)を「糸thread」と呼ぶことにする。

thread 「糸」は、モデル的にヒト蝸牛内の特定の有毛細胞群をターゲットとした、完全に非侵襲的な刺激源の系統だった分解になっていることに注意しよう。私たちは蝸牛機能を前提とする音声・音楽信号の分解をsinusoidal decompositionと呼ぶことにする。

この「糸」は、短時間フーリエ変換で得られたあるフレームnでk番目の単一の周波数を持つ正弦波素$a_{n,k}\ \cos 2\pi k\omega_0 t$から出発してsinusoidを連結していったが、現実の音声や音楽音響はこれら個別の「正弦波素を連結した糸」を多数合わせた、数百ミリ秒程度以上の時間持続する、滑らかに繋げられた複数の糸Thi(t)を縒り合わせた

(6)　$$Str(t) = \sum_{i=1}^{I} Th_i(t)$$

のように再合成することが出来る。このようにして、連結sinusoidで造られたthread の集合体として再構成された音源信号を私たちは弦stringと呼ぶことにする。このstringは個別の短時間フーリエ変換で得られるスペクトルに依存するのではなく、非周期的に時間発展する連結sinusoidを要素とするので超スペクトル的である。そこで以下ではこのような音源信号の再合成を超スペクトル弦合成super-spectral string synthesis と呼ぶことにする[4][5]。

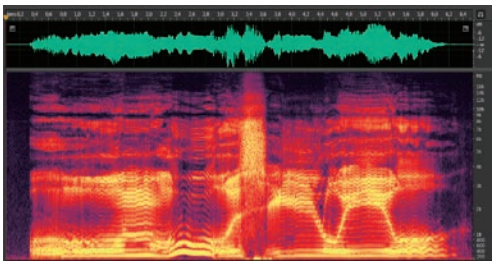


Fig. 2-1音源A：日本の伝統技芸「能」の謡（観世流「邯鄲」）の音声サンプル

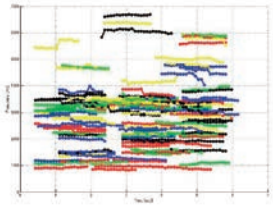


Fig. 2-2音源Aの弦分解
(N=95、横軸は時間、縦軸は周波数

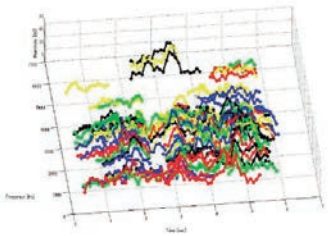


Fig. 2-3

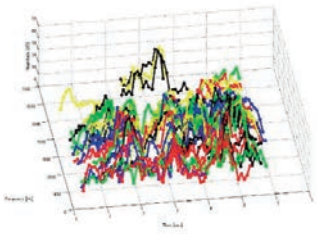


Fig. 2-4

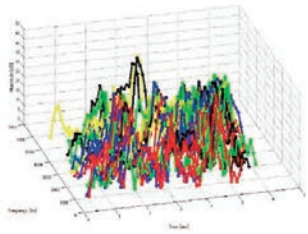


Fig. 2-5

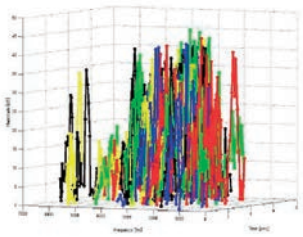


Fig. 2-6

Fig.2-3,2-4,2-5,2-6 音源Aを弦分解の＜時間・周波数・振幅＞3次元表現。周波数、振幅、持続時間の各々の観点から、蝸牛管内の有毛細胞クラスターに与えるインパクトを3次元超スペクトルデータ上で観察することが出来る。

§1-2 超スペクトル弦による音声言語のニューラル系統脱構築

いま音源信号*S*(t)がsinusoidal decompositionによってthreadの集合体に変換され、そこから弦*Str*(t)が再合成されるプロセスを

(7) $$S(t) \quad \fallingdotseq \quad Str(t)$$

と書くことにする。ここで次に、この弦*Str*(t)を個別の糸*Thi*(t)の部分集合に分解してゆくことを考えよう。すなわち

(8) $$Str(t) = \sum_{i=1}^{I} Th_i(t) = \sum_A Th_A(t) + \sum_B Th_B(t) + \sum_C Th_C(t) \cdots$$

のように糸threadの部分集合$\Sigma_A Th_A$(t)などを考えてゆく。仮に部分集合Aとして、たった一つだけの糸threadだけを選んだとすれば、$\Sigma_A Th_A$(t)を再生するときヒトの耳は連続する正弦波様の響きを聴くことになる。元来の音源信号*S*(t)が持っていた情報は失われ、仮に*S*(t)が音声言語としての意味を担っていたとしても$\Sigma_A Th_A$(t)が与える刺激だけからは、聴取する主体のヒト脳は言語の伝える意味を認知することができない。ここでさらに*Str*(t)の部分集合を足し加えて

(9) $$Str\ (t)_{A+B} \quad = \quad \sum_A Th_A(t) \quad + \sum_B Th_B(t)$$

のような部分弦partial stringを構成したとしよう。仮に$\Sigma_A Th_A$(t)+$\Sigma_B Th_B$(t)が元の音源信号*S*(t)がもつ音素の情報をすべて持っていたと

すれば、$Str(t)_{A+B}$は聴き手に元来の音声言語の意味を伝達できる可能性がある。だが実際には$\Sigma_B Th_B(t)$として$Str(t)$のさまざまな部分集合を選ぶことが可能であり、もし$\Sigma_B Th_B(t)$として単一のthreadを選ぶなら、$Str(t)_{A+B}$はたった二本のthreadだけで構成されることとなり、音声言語としての意味を伝達することはほぼ不可能と思われる。

このように、音源信号$S(t)$をsinusoidal decompositionした弦$Str(t)$から任意の糸threadの部分集合$\Sigma Th(t)$を選ぶことで、元来の信号の持つ構造structure例えば音素の構造phonetic structureや音楽的な構造 musical structureを系統だって崩すことが出来る。このような音源信号の崩し方を超スペクトル弦の脱構築super-spectral string deconstruction と呼ぶ。弦による脱構築の方法は、仮にヒト内耳の蝸牛神経が同時に受容すれば、意味を持った音声言語や音楽要素として認識されるものを、聴覚末梢の有毛細胞群ごとにON/OFFするような形で、任意の様態に分解し、あるいは再構成することができる。以下では超スペクトル弦の脱構築と音源情報の再構築を具体的に検討してみよう。

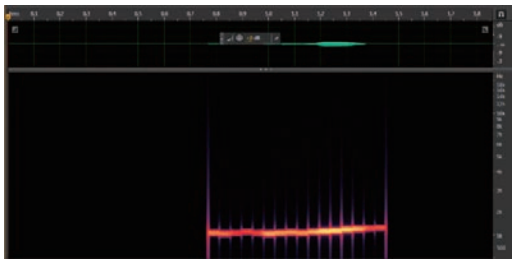


Fig. 3-1 N=1

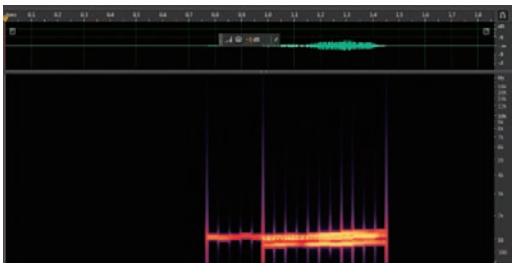


Fig. 3-2 N=2

Fig.3－1 同時に一つだけの周波数を含む、音程が変化する純音の線がひとつだけ聞こえる。器楽音と同様、明確な音程が聴き取れ、音声言語の特徴は一切持っていない。

Fig.3－2 N=1と同様の線が2つ聞こえる。2本の独立した声部が聴き取られ、ポリフォニーの聴取に近い

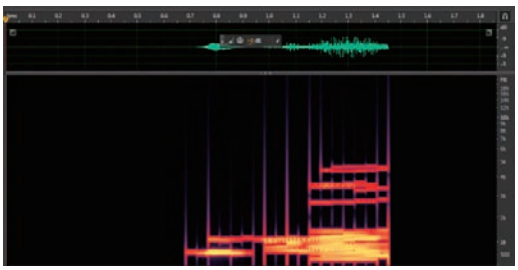


Fig. 3-3 N= 49

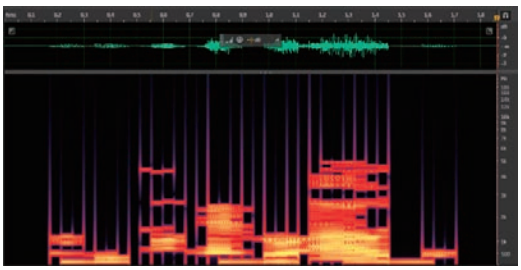


Fig. 3-4 N=81

Fig.3−3 多数の正弦波素が重なりあい、動物の鳴き声や鳥の声のような響きが得られるが、人間の声とは判別が付かず、音声言語としての意味も一切伝達されない。

Fig.3−4 人間の声のシラブルのように聴き取られ始めるが、言語としての意味はまったく聴き取ることができない。母国語のボキャブラリーでない場合はそのまま理解されずに終わる。また、不自然なイントネーションの場合にも聴取は著しく困難である。

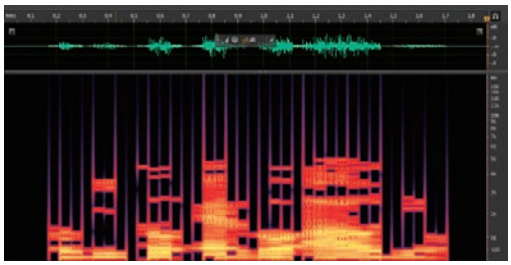


Fig.3-5　N=165

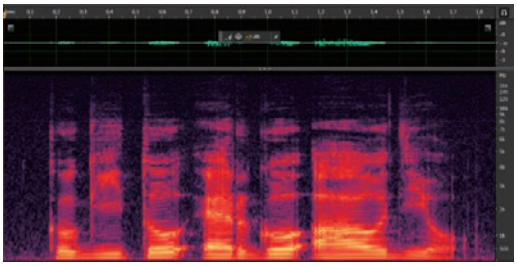


Fig.3-6

Fig.3−5 その言語を理解する者にとっては、音声言語として意味を聴き取ることが可能。

Fig.3−6 元来の音声：言語音声として聴取可能

§1-3 実験と議論　弦によるニューラル系統脱構築による、言語認知創発時の大脳新皮質の血中酸素濃度変化部位の特定

近赤外光機能スペクトロスコピーfunctiolnal near infra-red spectroscopy fNIRSは生活状態in vivo のヒトとくにその脳の生理活性を測定・評価する手法として知られる。血流中の酸素化合ヘモグロビンと脱酸素ヘモグロビンはおのおの近赤外光領域に特徴的なエネルギー順位を持つので、頭皮の外部からこれらの波長を持つ光を照射し、透過、反射、吸収を計測することで、測定部位の呼吸～生理的な活動レベルを評価することが出来る[6]。

脳イメージングの方法としては他に脳波計測EEG、MRI、脳磁計測MEG、陽電子分光PETなどの諸手法が知られるがNIRSは測定に伴う雑音が少なく、また測定環境を選ばないので楽器演奏などパフォーマンスの最中の脳血流を評価可能であり、私たちは2004年以来、株式会社島津製作所の協力のもとNIRS様々なデータ測定を行ってきた。

§1-3-1　NIRSの実験条件と解析方法

前記の方法で神経認知超弦に分解し、系列的に言語脱構築した音声データ刺激列を用いて、ヒト脳内での音声言語の創発をNIRSを用いて計測評価した。

測定は左右の1次聴覚野および運動性言語野に各17チャネル、ならびに前頭前野連合野に17チャネルの合計51チャネルで行った。

「刺激源としてはラテン語の短文3テキスト

を弦分解し、7段階に脱構築した系統刺激源と原音を用いた。7段階の脱構築音源刺激の糸の数、ならびに刺激提示条件は下表の通りである。

Step number	1	2	3	4	5	6	7	original
Threads	1	2	17	49		81	165	#

刺激呈示条件（単位：秒）

Rest	Task	Rest
30	153	30

§1-3-2　音声言語の脳内認知＝観測と認知受容成立の直接測定

以下に示す画像データは、脳波測定で広く採用される国際式10-20法にもとづき、頭頂CZの位置情報をホルダの基準位置にあわせ、被験者の位置情報を磁気式デジタイザを用いて3次元空間情報で検出、代替脳上にfNIRSデータを重ねてマッピングしたものである。前額部および左右両側部に各17ch、合計51chの光ファイバープローブを圧着して測定を実施した。

測定は酸素化型ヘモグロビンOxy-Hb、脱酸素化型ヘモグロビンDoxy-Hb 全ヘモグロビンTota-Hb各々に対応する近赤外光3波長を用いて行った。前頭部の17チャンネルに関する素データを図7－1に示す（赤：Oxy-Hb、青：Doxy-Hb 緑Tota-Hb）。

神経活動の負荷に伴い変化するOxy-Hb変化量を目的変数として多重検定（比較補正なし）を実行した。酸素化型ヘモグロビン量が多ければ活発な神経活動が可能であり、大脳新皮質の当該部位での認知演算の活発化指標として解釈が可能である。

検定は、有為水準P＜0.05としてコンテンツの問題箇所呈示前後の10秒間を比較した。得られたT値を±30でスケーリングして代替脳ポリゴン上にマッピングしたものが図7－2（側頭葉）7－3（前頭葉前頭前野）7－4（右側頭葉）である。

このデータは被験者3のものである。個々人の脳の機能モジュール分化は個別的であり、複数の人間の脳マッピングを加算することなどは出来ない。標準脳モデルへのマッピングにより統計的な比較も不可能ではないが、個体ごとに異なる画像診断の読影結果であり、ここでは臨床医学の症例研究同様、個別の例を示しておく。定性的にはすべての被験者で同様の結果が得られた。

可視化で用いた色彩は各々＋30（酸素型）：赤色　＋15：黄色　±0：緑色　－15：水色　－30（脱酸素型）：青色　のT値に対応している。

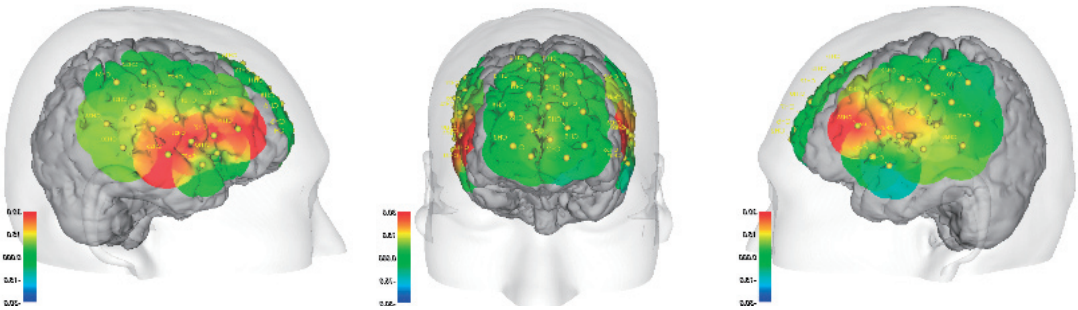
少数の「糸」だけの響きは器楽音のポリフォニーのように聞こえる。そこに「糸」が加わってゆくことで、響きは生物由来のものと認識さ

れ、やがてシラブルが認識され最終的に音声言語としての意味が獲得する。

被験者は、この意味を得るタイミングで指動作で指示を行い、それをマーキングしてイベント編集を行った。こうした被験者のすべての動きすべてfNIRSと同期したビデオを用いて撮影・記録している。

計測位置は国際式10-20法にもとづきCZの位置情報を全頭用ホルダの基準位置にあわせて装着した。さらに被験者の位置情報を磁気式デジタイザにより、3次元の空間情報で検出し、代替脳の上にfNIRSのデータを重ねた。部位前額および左右両側に各17ch計測し、合計51chの計測を行っている。実験は二人の被験者で行った。FNIRSは臨床的な脳機能可視化装置であり、まずその「読影」から議論を始める。多数の事例に基づく統計的な取り扱いは別論としょう。

すべての被験者で、正弦波の本数が少ない器楽音的な音刺激の認知と、より複雑で音声的な響きを持つ音刺激の双方で、感覚言語野相当部位に反応、変化が見られた。また意味が不明確な音刺激に「糸」が順次追加され、言語の意味内容が明確に把握されたタイミングで、前頭前野に特徴的なシグナルが見出された。



段階1　「糸」の本数が少ない「器楽音的」な刺激では音に対する注意や集中に伴う右の一次聴覚野に賦活が見られる。

段階2　意味が解らない声の響きから　はっきり音声言語として認識された際には、前頭前野腹内側および言語野における領域の明確な賦活が見られた。

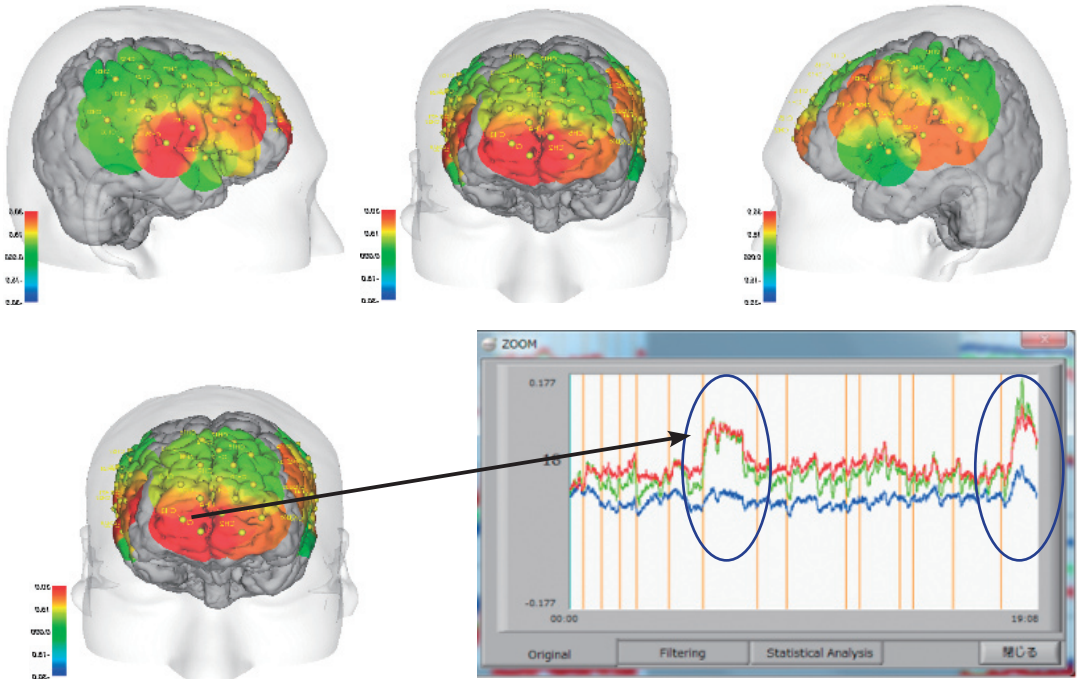


Fig.4 fNIRSによる、ヒト脳内での音声言語に受容認知と言語理解創発の直接測定

器楽的な音響であるか、言語音声であるかを問わず、音刺激により一次聴覚野に賦活が見られるが、言語の意味が獲得される段階で腹内側前頭皮質領域に大きなHb変化がみられる。腹内側前頭皮質領域は情動との関わりが深い。このため、このシグナルは言語の意味を認識したことによる驚きなど、情動の動きを反映している可能性がある。この方法による言語そのものの認識に関わる皮質部位の特定には、より慎重な検討が加えられるべきと考えられる。また言語認識においては、二人の被験者とも左背外側におけるHb変化が大きく、意味を獲得したタイミングではブローカ野を中心とする賦活が見られた。

このようにして言語音声と楽器音色を隔てることなく、内耳の周波数分解メカニズムを模したSinusoidal Decomposition のシステムを構築することができ、それにより音声言語など構造を持つ音源信号を系統だって脱構築した刺激音列を生成し、脳内での言語聴の創発部位をNIRSによって直接測定し、音声と音色を一元的に取り扱う一般的な方法を構成することが出来る[4][5]。

§2　歌うことと語ることの動力学：　第二相関解析と音楽諸量の導出

§1では音声言語や音楽音響を単振動に分解し再合成する、いわば「線形側」からのアプローチを紹介したが、同じ問題に別の面から取り組むことが出来る。フーリエ変換を用いる解析は基本周波数を与える短時間窓長によって解析可能な時間長さや周波数が規定されてしまうが、現実の人間の聴取は、認知分解能の限界内で連続的で、そこでの微細なずれ、例えば同一音源から左右の耳に到着する音の時間差micro delayや、側壁に反射して生まれた微細な木霊lateral reflection が決定的な役割を果たす。

以下ではまず、1チャネル信号の短時間自己相関関数を定義し、その長時間発展を聴覚の特質を生かすように考えて作ったフレームワークを紹介しよう。

§2-1　自己相関関数

モノラルの音源信号sを時間の関数$s(t)$として扱おう。この信号の自己相関関数$\Phi_s(\tau)$は

$$(10)\quad \Phi_s(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} s(t)s(t+\tau)dt$$

と解析的に定義される。τは信号の時間変数tと独立に変化する相関のタイムシフトである。τ=0での自己相関関数の値$\Phi_s(0)$を信号sの相関のノルム、ノルム$\Phi_s(0)$で割って正規化された自己相関関数$\phi_s(\tau)$を

$$(11)\quad \varphi_s(\tau) = \frac{\Phi_s(\tau)}{\Phi_s(0)}$$

とする。現実の音源信号は有限の信号長を持つ。そこでその自己相関関数もT→∞の極限を用いるのではなく有限時間長2Tをもつ相関関数$\Phi_s(\tau;t,T)$を

$$(12)\quad \Phi_s(\tau;t,T) = \int_{t-T}^{t+T} s(\tilde{t})s(\tilde{t}+\tau)d\tilde{t}$$

として、以下では有限時間長の正規化自己相関関数

$$(13)\quad \varphi_s(\tau;t,T) = \frac{\Phi_s(;t.T)}{\sqrt{\Phi_s(0;t,T)\Phi_s(0;t+\tau,T)}}$$

を考える。

相関解析は音響分析とりわけ建築音響の分野で長く用いられ有産な結果を多く得ており[7]、Tとしては従来ミリ秒単位の短時間サンプルを用いて「両耳間相関」「両耳相関幅」など各種のパラメータが計算されている。以下に例を示す。

Fig.5に私たちがバイロイト祝祭劇場で収録したヴァーグナーの楽劇「トリスタンとイゾルデ」冒頭の「水夫の歌」の音源信号のスペクトログラムを示す。この中から10ミリ秒ほどの長さの短時間サンプルを取り出して自己相関を見たのがFig. 6-1 1000ミリ秒ほどの長さの短時間サンプルを取り出して見たのがFig. 6-2である。自己相関解析はサンプル内の時間的に反復する構造をピークとして示す。Fig.6-1は数ミリ秒＝数百～1000ヘルツの周期性すなわち可聴域の反復構造を、またfig.6-2は数百ミリ秒＝数ヘルツの周期性すなわちこの場合は劇場内での側壁反射などより長い時間単位での反響の構造を反映している。

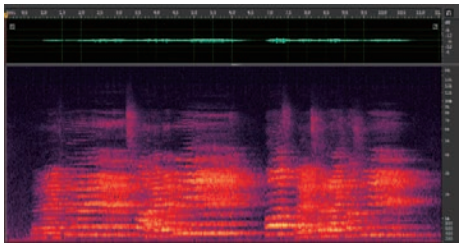


Fig. 5

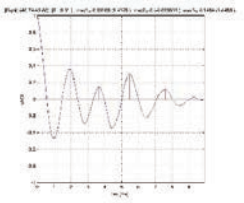


Fig. 6-1

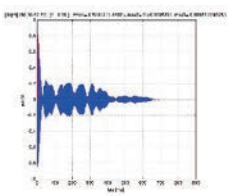


Fig. 6-2

Fig.5 楽劇「トリスタンとイゾルデ」冒頭部「水夫の歌」のスペクトログラム

Fig.6-1 10ミリ秒オーダーの自己相関　　Fig.6-2 1000ミリ秒オーダーの自己相関の振舞い

§2-2　相関関数の時間発展

この短時間信号の相関は、準定常な状態を仮定すれば長時間の考察にも意味を持つ。建築音響に量子力学で導入されたディラックのδ関数[8]を応用したインパルス応答が室内空間の音響指標として普及したのはその一例で、我々のグループも上記バイロイト祝祭劇場に於いて詳細なインパルス応答測定を実施している[9]。しかし上記の「トリスタンとイゾルデ」など、私たち音楽家が取り扱う響きは多様で微細な変化に価値を見出すものであって、1秒ないし10ミリ秒の短時間サンプルだけによる指標は殆ど意味をなさない。そこで準定常な状態から出発し、その時間発展を考えることで動的な問題を解決する理論的なシナリオを考えよう。ちなみに筆者の念頭にあったのは水素原子の静的なスペクトル[10]を散乱理論に拡張したマックス・ボルンの取り扱いである[11]。ボルンによる波動関数の確率解釈の援用については後に記す。

いま、ヒトが耳で聴き取ることが出来る時間長さLの長時間音源信号$\zeta(t)$を考える。この$\zeta(t)$をt=0から時間幅2Tごとに合計M個の短時間フレーム音源サンプル片s_m(t)に分割することを考える。

$$\zeta(t) = \sum_{m=1}^{M} s_m(t) \tag{14}$$

定義からL=2T×Mである。このように切り出された短時間フレーム音源サンプル片s_m(t)に対してフレーム自己相関関数$\Phi_{s,m}(\tau;t,\mathrm{T})$を上の議論と同様にフレーム毎に

$$\Phi_{s,m}(\tau;t,\mathrm{T}) = \int_{t-T}^{t+T} s_m(\tilde{t})s_m(\tilde{t}+\tau)d\tilde{t} \tag{15}$$

として定義する。$\Phi_{s,m}(\tau;t,\mathrm{T})$は、もとの長時間音源サンプル$\zeta(t)$のm番目の短時間フレーム・サンプルについての自己相関関数を与える。

ここで自己相関関数$\Phi_{s,m}(\tau;t,\mathrm{T})$、相関のずれ時間$\tau$の双方と直交する向きにランニングタイムtの軸を設定し、フレーム番号mの順に自己相関関数の値を並べることを考える。

自己相関関数はランニングタイムt軸上で

$$t_m = 2m\mathrm{T} \quad (m = 1,2,3\cdots) \tag{16}$$

だけで値を持つものとする。このように考えると、ランニングタイム軸上で離散化時刻t_mにのみ値を持つ自己相関関数列$\Phi_{s,m}(\tau;t,\mathrm{T})$

の3次元的なconfigurationを得ることができる。ここでおのおのの離散化時刻t_mで見れば$\Phi_{s,m}(\tau;t,T)$は従来どおりの短時間フレーム内での信号の自己相関を表す。またこれをランニングタイムtの関数と考えるなら、$\Phi_{s,m}(\tau;t,T)$ (m=1,2,3…)は自己相関の時間発展を表すと見る事ができる。上記のようにランニングタイムt軸上に整列された$\Phi_{s,m}(\tau;t,T)$の全体を

$$\overrightarrow{\Phi_s}(\tau;t,T) = \bigcup_{m=1}^{M} \Phi_{s,m}(\tau;t,T) \quad (17)$$

と書くことにする。$\overrightarrow{\Phi_s}$をランニングタイムtに依存する音源信号$\zeta(t)$の時間に依存する伝播相関関数Time-dependent traveling autocorrelation functionと呼ぶことにしよう。

Fig. 7

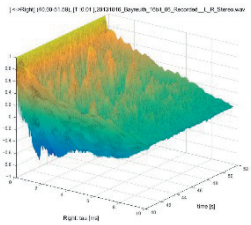


Fig. 8-1

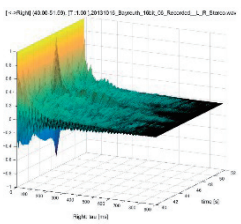


Fig. 8-2

Fig.7　バイロイト祝祭劇場での演奏収録の模様

Fig.8 Traveling Autocorrelation funcitons　Fig.8-1 T=10msec　　Fig.8-2 1000msec

§2−3　聴覚的認知のタイム・スケーリング

式（17）は有限時間長さTをもつ音源信号sの内部構造の情報が反映する。とりわけτの関数として$\phi_s(\tau;t,T)$は信号sの周期性を反映するピークを示す。そこでヒト聴覚の認知特性をもとに、Tの値の違いにより$\phi_s(\tau;t,T)$が示す特徴を以下のように分類しよう。端的に言うなら、Tは相関による音の「顕微鏡」の倍率を与えるものである。

場合α：可聴域sが比較的短く有限時間長 T≤10ミリ秒程度のtime scale

T～10ミリ秒以下の特徴的時間幅での音源信号sの周期性を反映した構造がϕ_sに見られる。fc=1/Tとして特徴的な周波数を評価すればfc>100Hzとなるのでfcは可聴域の範囲に入り、この信号周期性は信号の音程＝ピッチを反映するものと考えられる。

場合β：弁別閾上sが比較的に長く有限時間長 T≥100ミリ秒程度のtime scale

T～100ミリ秒以上の特徴的時間幅での音源信号sの周期性を反映した構造がϕ_sに見られる。上と同様に考えるとfc≤10Hzとなるのでfcはメトロノーム・テンポの範囲に入り、この信号周期性はヒトが時間的前後関係の弁別が可能な構造、つまりリズム的な特徴や、音声言語であればシラブルを反映するものと考えられる。

場合γ：含サブリミナル領域sが上記の中間の値で有限時間長 10≤T≤100ミリ秒程度のtime scale

この場合10Hz≤fc≤100Hzとなるが、特に16Hz≤fc≤50Hz程度の周波数帯域はヒトにとって可聴域の下限、ないし前後関係が独立して弁別しにくい刺激提示頻度となるため、場合1や場合2のように「音程」「リズム」といった画然とした特徴は得られない。だが音楽音響であればアタックやタッチ、音声であれば子音の立ち上がりなどに関係する極めてデリケートな特徴を検出可能することが可能である。

そこで以下では、これら各々のスケーリングタイムscaling timeにおける音楽・音声信号の相関を具体的に考え、とりわけその時間に依存じた発展time-dependent developmentを検討してみよう。

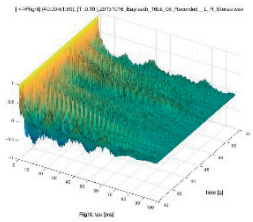


Fig. 9 時間に依存する自己相関関数の推移

T = 100msec ヒトの聴覚的時間分解能の臨界近辺での音の振る舞いを確認できる。

§2−4　聴覚的相関の時間発展解析

音源信号の動力学的自己相関$\overrightarrow{\Phi_s}(\tau;t,T)$は観測可能な物理量から計算されるが、ここから様々なタイムスケールでの認知的諸量を導くことが出来る。

タイムスケール長時間Tがパラメータとなることからヒト認知の異なる認知的特徴量が得られることから、サンプルを切り出すフレーム長は顕微鏡の倍率のような役割を果たす。$\overrightarrow{\Phi_s}(\tau;t,T)$を一種の通時的なポテンシャルと考えて以下の考察を進めたい。

対象となるあるフレーム長Tを与えて得られた動力学的自己相関は$\overrightarrow{\Phi_s}(\tau;t,T)$ $tm=2mT(m=1,2,3\cdots)$のように、とびとびにτ-Φ_s平面［ずれ時間τと自己相関Φ_sの張る平面］上に相関曲線が点在する空間曲線の集合となる。以下では従来の短時間サンプルを用いた相関解析では不可能だったTraveling Correlation Function の時間tに添った$\overrightarrow{\Phi_s}$の変化を取り出すことを考えよう。

いま正規化された$\overrightarrow{\Phi_s}$を相関ϕが一定の値cでの平面で切断した断面を考えよう。このとき断面$\overrightarrow{\Phi_s}\cap$ Cは$\overrightarrow{\Phi_s}$等高線Cの集合を示す。cは0<c<1の値を取る実数とする。これを用いてずれ時間τ上で相関が残存する「有効なずれ持続時間」を指標化してみよう。

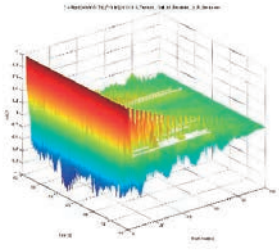


Fig. 10-1 Impression view

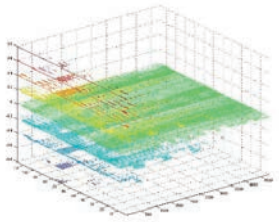


Fig. 10-2 Contour of
"$\Phi_s \cap C$." c = +0.8.+0.6.+0.4.+0.2.0

一定の間隔⊿cで$\overrightarrow{\Phi_s}$を切断した等高面による$\overrightarrow{\Phi_s}$の断面$\overrightarrow{\Phi_s}\cap$ Cは画家Piet Mondrianの作品を想起させる。私達の研究室内ではこのような階層的離散化をImpression viewの”モンドリアン格子による階層化”と呼んでいる。仮にc=0.2で相関の値が十分小さくなったと考えるなら、そこに相当する階層面に現れるτの値の包絡線から、相関が有効である限界時間長さを評価出来る。相関のシフト時間の限界は量子力学における電子の波動関数の相互作用の距離的限界と似ており、そこからマックス・ボルンは波動関数を「電子の存在確率密度を示す量」と考える量子力学の確率解釈を提出し、およそ多様な物理量計算を可能とした[11]。これを一つのモデルとして、相関が有意に働き得る限界範囲を見積もって認知諸量を計算してみよう。

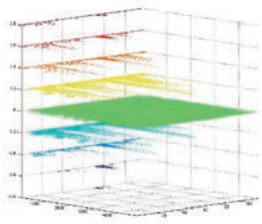


Fig. 10-3 相関の格子化＝離散指標化

いま$\overrightarrow{\Phi_s}\cap$ Cのずれ時間τ方向での最大値を結んだ包絡線をランニング時間tの関数Ψ(t)として

(18) $$\Psi(t) = \mathrm{Max}\left[\overrightarrow{\Phi_s} \cap C\right] = \mathrm{Max}\ \tau\ (t)\big|_{c=r,\tau max=T}$$

と定義する。$\Psi(t)|_{c,T}$は単一フレーム内での相関の有効持続時間の変化を示す。

以下に幾つかの具体的な有効相関時間関数Ψの実例を示す。以下ではc=0.2として計算を実行している。音源はすべて私達がバイロイト祝祭劇場で演奏したトリスタン冒頭の水夫の歌である。

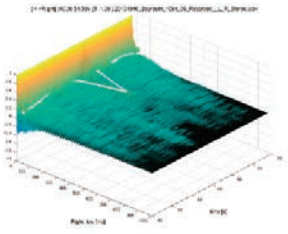


Fig. 11-1 $\Psi(t)|_{c=0.2,\,T=1000msec}$

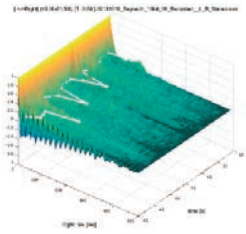


Fig. 11-2 $\Psi(t)|_{c=0.2,\,T=500msec}$

Figure 9-1はT=1000ミリ秒の粗い格子で相関を見るため、同じサンプルの中でも最も長く伸ばした母音一つを反映するピークを持つΨ(t)が得られる。Figure 9-2はT=500ミリ秒としたため、Figure 9-1と異なりサンプルに含まれる複数の母音シラブルの有効な持続を反映するΨ(t)が得られる。Tとして100ミリ秒を与えると、その中に含まれる数十ミリ秒の時間スケールにヒトの時間認知分解能限界Δ_{KL}が含まれるため、Ψ(t)は音程をもって聞き取れる「周波数」とリズムとして聴き取られるパルスの双方を反映する（Figure 9-3）。

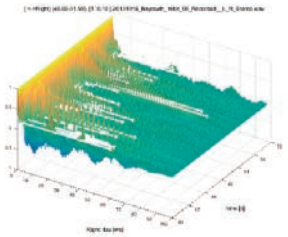


Fig. 11-3 $\Psi(t)|_{c=0.2,\,T=100msec}$

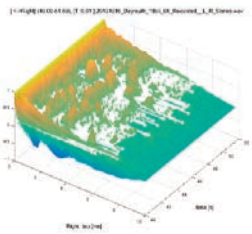


Fig. 11-4 $\Psi(t)|_{c=0.2,\,T=10msec}$

T＝10ミリ秒ではこの時間長さの内部をヒト聴覚は弁別出来ないため、Figure 9-4に示すように細かな反復構造として可聴域の周波数のほか、サブリミナルな変動を表すΨ(t)が得られる。階層化平面内の$\Psi(t)|_{c,T}$の振る舞いをFigure 12に示す。

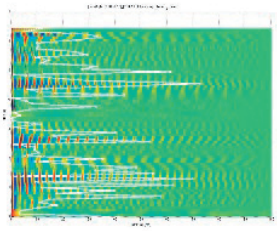


Fig. 12 $\Psi(t)|_{c,T}$

§2−5　時間に依存して変化する音楽量を導く：第二相関解析と有効相関時間関数

以下ではc=0.2断面内で、Tとしてサブリミナルな領域を含む時間長さを取って$\Psi(t)|_{c,T}$を考えよう。Fig.13はFig.12と同一のテータを整理したものである。

これをランニング時間tの関数として、ヒト聴覚が聴き取りうる限界近くの時間粒度で短時間音源信号の有効相関持続時間の変化として評価しよう。

いま、ここから可聴域に相当する高周波の変動を取り除き、低周波成分を与える包絡を得たものがFig. 14である。

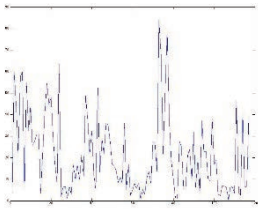


Fig. 13

Fig. 14

Fig.12 の$\Psi(t)|_{c,T}$と同じもの（Fig.13）に低域通過処理を施したもの （Fig.14）

Fig. 13の信号Ψ(t)を対象に、これ自身の自己相関関数Dを

(19)　$$D \equiv \langle \Psi(t) | \Psi(t) \rangle$$

として定義すると、Dは相関の有効持続時間の周期成分を反映する構造を示す事になる。

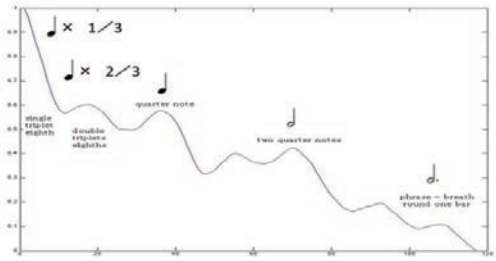


Fig. 15 $\langle \tilde{\Psi} | \tilde{\Psi} \rangle$ 音価の動力学的再構成

ローパスフィルタを用い可聴域のノイズを取り除いたΨの有効持続時間の周期的反復構造はサンプルに含まれる音価つまり音の持続を反映する。

このようにして、実際におのおののピークは元サンプルで歌われるシラブルの長さに対応する＜持続の相関スペクトル＞構造が見え、これにより音価が動力学的に再構成することが出来た。

ここで示したサンプルは定量的に譜面に記すことが可能なヴァーグナー楽劇の抜粋であるが、そのような記譜が容易でない能楽や雅楽、声明といった対象、さらには合理的な音価システムを持たない「梵鐘の一打」のような響きに対しても、その周期構造を（無理やり音符に押し込める事無く）量化することが出来るメリットがある。

また相関時間関数Ψ(t)を時間で偏微分して得られる時間の関数

(20) $$\frac{\partial}{\partial t}\Psi(t) = A(t)$$

から、聴覚的に認識される時間に沿った振る舞いの変化が記述出来る（Fig.16）

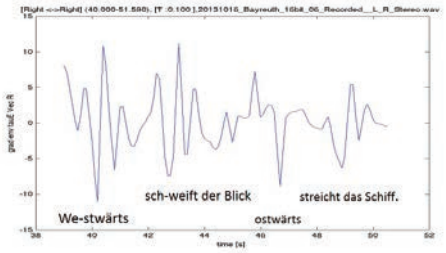


Fig. 16 $\frac{\partial}{\partial t}\Psi(t)$アタックの動力学的再構成

これは音の立ち上がりやソノリティを直接反映する、やはり動力学的なアタックの再構成になっており、歌唱の場合はシラブルの母音や子音の現実の響きを明示するものになっている。

§3　時間と空間のサブリミナル認知：聴覚的複素平面とラプラス解析

前章で1チャネル音響情報の自己相関関数解析に適用した手続きを、今度は2チャンネル情報のステレオ情報に対して、相互相関関数解析を用いて実施してみよう。

ほぼ同様の論理的な手順を踏むため、以下では詳細を略しながら概説したい。

§3−1　相互相関関数

ある連続した一定時間のあいだ、聴き手が両耳で聴取する音の像や、そこから認知される聴覚的な空間像などは、次式で定義される両耳間相互相関関数$\Phi_{lr}(\tau)$で評価することが出来る。ここでs_l(t)は左チャネルに齎される音源信号、s_r(t)は右チャネルに齎される音源信号を現す。すなわち

(21)　$\Phi_{lr}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} s_l(t)s_r(t+\tau)dt$

の様に解析的に定義される。τは信号の時間変数tと独立に変化する相関のタイムシフトでτ=0 での同一チャネル音源信号の自己相関関数の値$\Phi_{ll}(0)$、$\Phi_{rr}(0)$を各チャネル信号の自己相関のノルム、これらを用いて正規化された両耳間相互相関関数$\phi_{lr}(\tau)$を

(22)　$$\varphi_{lr}(\boldsymbol{\tau}) = \frac{\Phi_{lr}(\boldsymbol{\tau})}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}}$$

と定義しておく。現実の各チャネル音源信号は有限の信号長を持つので、信号長2Tをもつ相互相関関数$\Phi_{lr}(\tau;t,T)$を

(23)　$\Phi_s(\tau;t,\mathrm{T}) = \int_{t-T}^{t+T} s_l(\tilde{t})s_r(\tilde{t}+\tau)d\tilde{t}$

として相関関数を評価する。相関関数の値は積分区間Tのタイム・スケールによって全く異なる指標を表す。Tが1000ミリ秒単位の長さを持つなら、相互相関関数は100ミリ秒単位での2チャンネル音響現象の特徴的な構造、例えば側壁反射によるホールの響きの音風景 echoic soundscape を反映するものになるだろう。

建築音響への応用において、Tを一秒程度に設定している背景は、この特徴時間長さの音源信号の情報構造を抽出したいためである。1秒内外の音のずれを私たちの耳は「こだま」エコーとして知覚する。

下にT=10msec ならびに1000msecの、二つの短時間2チャンネル間自己相関解析のデータ例を以下に示す。

音源S_{1lr}(t)は私たちがバイロイト祝祭劇場で演奏した「トリスタンとイゾルデ」冒頭、水夫の歌の最初の4小節“Westwärtsschweift der Blick ostwärtsstreicht das Schiff.”の6チャンネル音場収録のうちのLR2チャンネルである。両者の基線距離は100cmである。

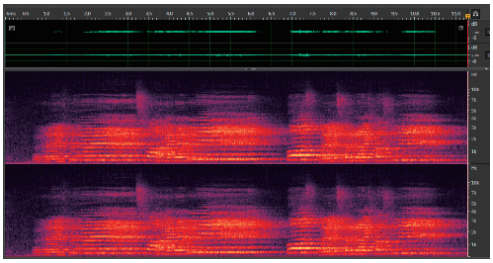


Fig. 17 $S_{1/r}(t)$ トリスタンとイゾルデ冒頭の2チャンネルデータ

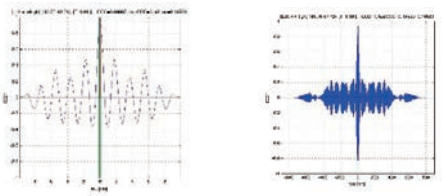


Figure 18　短時間相互相関関数
18−1 T=5msec, 18-2 T=1000msec

§3−2　相互相関関数の時間発展の取り扱い

§2で示したのと同様の手続きによって、時間に依存するTraveling Correlation Functionとして2チャンネルの相関を取り扱う事ができる。紙幅の制限からここでは結果のみを示す。詳細は原著を参照されたい。

Fig. 19 $S_{1/r}(t)$ のTraveling Correlation function.

Fig. 19-1 T=5msec　　Fig. 19-2 T=1000msec

ここから、聴覚における超越的統覚のバインディング問題を取り扱うことが可能となる。紙幅の限界から本稿では省略し、原著ならびに別論に譲ることとする[12]。

§3−3　指標の導出

§2と同様の手順によって私たちは多チャンネルの相関（例えば両耳）から様々な聴覚的、言語的、そして音楽的な指標を取り出すことが出来る。ここでは結果のみを示す。

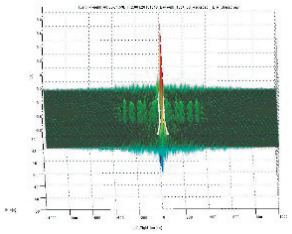


Fig. 20-1 $\Psi_{/r}(t)$ T=1000msec,

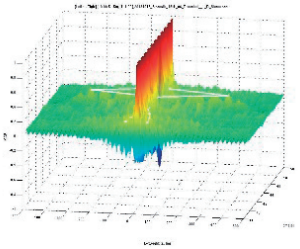


Fig. 20-2 T=500msec

積分区間Tを変化させることで私たちは聴取される時空間の異なる「倍率」での構造を知ることが出来る。端的にはFig.20-1は音源の広がり＝位置、20-2は大きなブレス、20－3はフレーズ、20-4はシラブルのアタックを反映している、

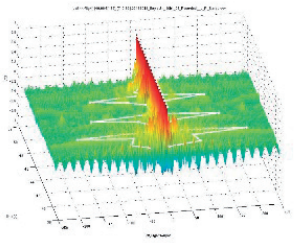
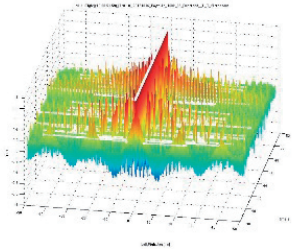


Fig. 20-3 $\Psi_{/r}(t)$ T=250msec,



20-4 T=50msec

Fig.20-4 から相関の値が0.2となる断面を取り出したものがFig.21である。

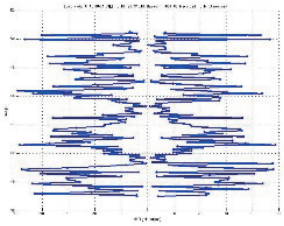


Fig. 21 $\Psi_{/r}(t)$ T=50msec

§3−5　音響空間認知の静的側面：

§2と同様、相関の時間発展Fig. 21から幾つかの指標を定性的、定量的に取り出すことを考えよう。

まず、先ほど述べた「主観的音源広がり」など、$\Psi_{lr}(t)$の「幅」が重要な役割を担う要素を考えてみたい。ランニングタイムtごとにこの幅を取り出したものがFig.22-1である。 いま観測者が不動、また音源も移動しない静的な状態で緩やかに変化する時間指標に注目するべくFig.22-1に低周波通過、すなわちローパスフィルタの処理を施したのがFig.24-2 である。

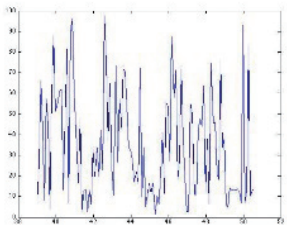


Fig. 22-1 $\overline{\overline{\Psi_{lr}(t)}}=|\Psi_{l}(t)|+|\Psi_{r}(t)|$

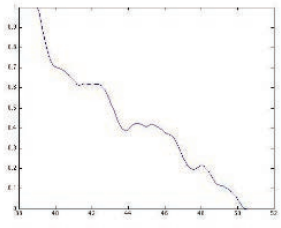


Fig. 22-2 $\widetilde{\Psi_{lr}(t)}$

Fig.22-2は、より緩やかな時間変化成分を抜き出したトラベリング相関時間幅$\widetilde{\Psi_{lr}(t)}$をランニング時間の関数として取り出したことになる。まず先ほど同様、それ自身の自己相関w_{lr}を計算してみよう。

(24)　$$w_{lr}=\left\langle \widetilde{\Psi_{lr}(t)} \middle| \widetilde{\Psi_{lr}(t)} \right\rangle$$

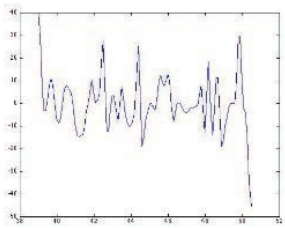


Figure 23 w_{lr}の振る舞い：2チャンネルで評価する時空の周期的構造を示す。

Fig.23にw_{lr}を示す。w_{lr}は先に自己相関で示したFig.16と似たように見えるが、舞台上で時間空間的に広がる音の拡散repercussionのもつ周期性を示しており、別の指標を与えている。Fig.16が1チャネル情報を扱い音の持続の周期性を示すのに対して、Fig.23は2チャンネルの相互相関時間幅は時空間双方の情報を含むことによる。

またトラベリング相関時間幅$\widetilde{\Psi_{lr}}(t)$をランニング時間で偏微分して得られる関数

(25) $$\frac{\partial}{\partial t}\widetilde{\Psi_{lr}}(t) = \pi(t)$$

は劇場内で時空間に広がるアタックやソノリティ、シラブルの分離などを反映するものになる。Fir.24に同じサンプルからの演算例を示す。

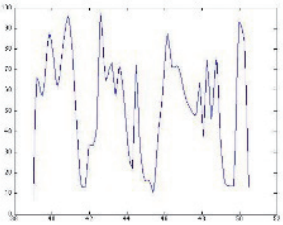


Figure 24 π(t)劇場内に実際に広がるアタックの振る舞い

オペラハウスやコンサートホールでの実際の音楽作りではこうした響きの質を瞬時に聴き取り、反射的に修正してゆくソルフェージュの能力が必要不可欠である。我々音楽家同士の会話では「音の伸び」「切れ」とか「アタックの強さ」「音の硬さ」「音の厚み」といった言葉が普通に使われる。Fig.24はそうした属性を示す。従来の素朴な音楽理論には、これらをカバーする数理枠組みが存在しなかった。

とはいえ別段複雑なことではなく、これらは時間に依存し空間的に広がった響きの変化をヒトの耳が捉えているのだから、原理に即して整理して客観指標化、定量化したものである。

§3−6　一般化のアウトライン：聴覚的複素平面とラプラス解析

ちなみにτとtは共に「時間」の次元を持つ変数だが、上記の枠組みではグラフ上で直交し、平面を張っているので、これを奇異に感じた人があるかもしれない。

いま可聴域の周波数成分をω、これと直交するもう一つの時間方向の周波数成分を$i\sigma$（iは虚数単位）と書くことにすれば、波動W（t）として

(26) $$W(t) = A(t)e^{-i(\omega - i\sigma)t} = A(t)e^{(\sigma - i\omega)t}$$

を考えれば（12）はよく知られたラプラス変換の積分核と同じ形をしているのがわかるだろう。

複素周波数ω-iσの虚数成分σは過渡現象の時間に依存して変化する成分を表現していることが解る。ここから、本論文で示した枠組みがラプラス解析の非定常状態ダイナミクスと親和性を持つ事が知られる。小研究グループは可聴域のあらゆる線スペクトルと帯域雑音を超格子にモデル化するプレトダイナミクスの方法を開発し、定常的な超狭域雑音が聴覚認知に誘起する、音色のKolmogorov Compoment等の基礎

的な事実を初めて見出した。

東京大学大学院情報学環・作曲＝指揮研究室はこうした基礎的な数理枠組みの整備を並行することで、同様の数学的形式の社会科学など他分野への応用も進めている。

§4　結語に代えて～「悟性は情動に遅れる」　音響空間情報倫理

Fig.25-1に、10年前の拙著「さよなら、サイレント・ネイビー」（2005-2006）[13]で用いた実測データを示す。インターネット上に不特定多数に向けテロリストが公開した音声動画コンテンツを視聴し、前頭前野連合野全体の血中酸素濃度が低下し、十分な思考が困難な状態になった被験者の脳血流可視化画像である。2004年イラクで身柄を拘束された日本人青年、香田証生氏を被写体とする残虐画像を刺激源とし、視聴の結果、前頭前野の酸素濃度が極端に低下した状態を示している。被験者へのインタビューでも思考しにくかった内観を確認している。

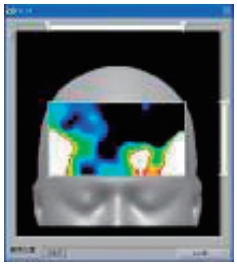


Figure 25-1
ネットワーク・テロコンテンツの視聴で虚血する前頭前野

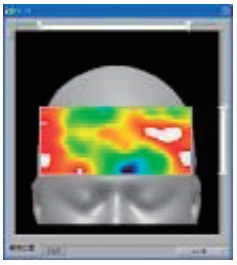


Figure 25-2
お笑い視聴時の皮質同部位

人間は、強い情動に支配されるとき、同時に液性の情報が血中に齎され、その状況で反射的な行動をとったり、意思を決定してしまう。意思決定については、さらにそれが永続するという生理的事情がある。同様に**Fig.25-2**は音声動画コンテンツとして示されたお笑い番組を見て大笑している状況での同じく前頭前野新皮質の脳血流酸素濃度の可視化画像である。

激しい情動の動きがあるとき、概して前頭前野は虚血に近い状態を示す。これは、新皮質など進んだ部位を進化的に持たなかった時期から大半の情動を生物が獲得していた経緯を反映するものと思われる。

耳からもたらされた刺激は直接情動に働きかけ、多くの情動の発動は極めて早く50ミリ秒以内に発動する。翻って新皮質での高度に知的な演算には数百ミリ秒以上を要するものが少なくなく、熟慮するような場合には数十秒、数分という事も珍しくない。この時間的な逆転から認知的な逆転が起きる。

一般に**情動は悟性に先立ち、反射的な行為を誘引し、悟性が意識するより前に意思を決定してしまう。**

上記の測定を行った2004年、ネットワークはいまだナロウバンドの状況で音声動画コンテンツのトラフィックは今日とは比較にならない少なさであった。

それから10年を経た2014年、ISIS「イスラム国」によるネットワーク上への系統だった残虐音声動画のリリースが、国際社会に重篤な影響を及ぼし、日本国内でもさまざまな波及が懸念されるようになった。これを刺激源とする視聴時の脳血流可視化結果を以下に示そう。

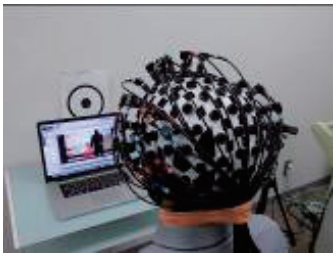


Fig. 26 測定の模様

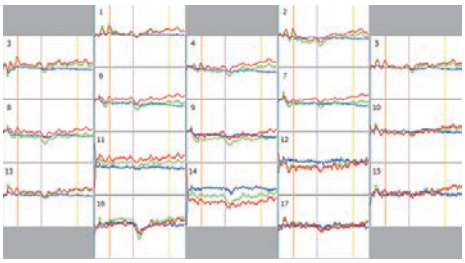


Fig. 27 前頭前野各チャネルでの経時変化

以下の測定では2015年1月から2月に掛けて武装勢力ISIL（イラクとレヴァントの「イスラム国」）が日本人を誘拐し、人質を被写体として撮影・編集しインターネット上に世界公開した音声動画コンテンツを直接刺激源として提示し、3人の被験者の脳血流の酸素濃度変化を測定した。

測定は通常のネットワーク音声動画視聴を念頭に、電灯が点いた室内環境で行った。

Fig.26に刺激の提示とfNIRSによる測定の模様を示す。光ファイバーの測定端子が前頭前野ならびに左右の側頭葉をターゲットとして合計51チャネル準備された。3人の被験者に対して予め準備された30秒のコンテンツが提示され、その間の上記当該領域の大脳新皮質の血中酸素の酸素結合度を測定した。セットアップや測定原理、評価の定量は§1と同一である。結果をFig.27ならびにFig.28に示そう。

Fig.27で各チャネル中央付近でカーソルによって示されているのが、問題になりうるシーンの提示で、直後から酸素化ヘモグロビンの指標が下がってゆくのがはっきりと確認できる。統計処理を施したマッピングをFig.28に示す。

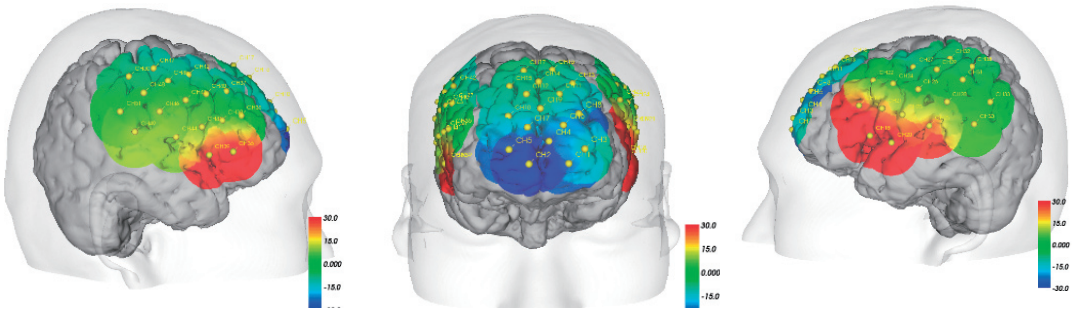


Fig. 28 ISILがネットワーク公開した残虐画像視聴時の前頭前野虚血のマッピング

本稿の冒頭にも記した通り、音声動画がもたらす仮想環境の体験は実空間の定位感を丸ごと欠く。しかしその状態での経験や記憶も、悟性の作動に先んじて情動を動かし、身体は反射的な行動を取り、動作主体は十分な悟性を働かせる以前に意思決定を下してしまうリスクを免れない。

私たちが進めるのは音楽の基礎を物理的音波の振る舞いとヒト聴覚と脳による認知・観測から記述、解析する極めて地道な基礎研究であるが、ネットワークコンテンツの脳認知評価による予防公衆情報衛生といった喫緊の応用分野にも密接に関係していることを付記しておく[14]。

基礎研究は基礎的であればあるほど応用範囲が広い。軍事研究が最たるもので、応用によって両刃の剣ともなり得ることを銘記しておくべきだろう。

本稿で取り上げた私の理論は、一方で音楽の創出・演奏の道具でもあり、また超越的統覚の死角を探る基礎的な探求も可能である。と同時に、時代が直面するこのような喫緊の課題にも答える倫理的な課題を負うものと常々考えている。

尊敬する哲学者の一ノ瀬正樹教授は、倫理の定量を扱う可能性の哲学を模索しておられるが、そうした観点に音響空間認知の数理的な考察が、ネットワークコンテンツの倫理的考察にも資する事を広く共有したく、本稿をまとめた。紙幅の限界から省略した幾つかの詳細については原著をご参照頂ければ幸いである。[15][16]。

本稿で取り上げたシステムの構築、測定そして解析を共に進めてくれた佐藤貢士、武井明則の両君に感謝する。何らかの瑕疵があればすべて筆者一人の責任である。

References

1 Shannon, C. “Mathematical Theory of Communication ” University of Illinois Press（1963）

2 von Békésy, G, “Experiments in Hearing”, McGrau-Hill, New York, （1960）

3 McAuleyR.J. and Quatieri, T.F., Speech Analysis/Synthesis Based on a Sinusoidal Representation” IEEE Transaction on Acoustics, Speech and Signal Processing. Vol. ASSP-34, No.4. （1986）

[4] Ito, K. et al. “Analysis of Noh and Composition of Noh Opera using a Sinusoidal Model”, Music Information Science 34-13, p.79 - p.82. Information Processing Society Japan. (1999)
[5] Ito, K, Takei, A. Sato, K. and Toyoda, T. “SEVERAL APPROACHES FOR INHARMONIC COMPONENTS IN SINGING VOICE AND SPEECH” ICSV-22 # 739 (2015)
[6] Siesler, H. W. ed. “Near Infrared Spectroscopy”, Wiley-VCH (2002)
[7] Ando, Y. “Auditory and Visual Sensations”, Springer (2010)
[8] Dirac, P.A.M. "The Principals of Quantumu Mechanics" Oxford University Press (1930)
[9] Garai, M. Ito, K et al. THE ACOUSTICS OF THE BAYREUTH FESTSPIELHAUS, Proc. ICSV 22 #651 (2015)
[10] Schrödinger, E. “Quantisierung als Eigenwertproblem” Annalen der Physik, 384-4, pp.361–376 (1926)
[11] Born, M. “Zur Quantenmechanik der Stoßvorgänge". Zeitschrift für Physik 37. (12) 863-867 (1926)
[12] Ito, K, Takei, Yamauchi, H.A. Sato, K. and Toyoda, T. “Correlation function analysis for 3-dimensional opera performance” ICSV-22 # 737 (2015)
[13] 伊東　乾　「さよなら、サイレント・ネイビー」集英社　(2006第四回開高健ノンフィクション賞)
[14] 伊東　乾　井上正雄　武井明則「ネットワーク視聴覚コンテンツ受容の脳血流評価と認知的死角」情報社会学会誌（2015　情報社会学会誌Vol.10 pp.35-44）：
[15] Sato, K. Ito, K. Takei, K.“A New Method of Analysis & Visualization over the Cognitive Real/Virtual SPACE-TIME Using Correlation Function” Proc. ICSV-22 # 734 (2015)
[16] Ito, K, Takei, A. Sato, K. and Toyoda, T. “Supra-Spectral methods for cognitive musicology“ ICSV-22 # 738 (2015)

伊東　乾 (いとう・けん)

1965 年 1 月　東京生。東京大学理学部物理学科、同大学院理学系研究科物理学専攻修士課程修了、博士課程単位取得退学、同大学院総合文化研究科超域文化科学専攻博士課程修了。
[**専攻領域**] 作曲、指揮、音楽の自然哲学
[**主要作品、演奏等**]
作品：3 群の管弦楽のための Dynamorphia (1978-98)　他
演奏：John CAGE 遺作“OCEAN”世界初演 New York Merce Cunningham Dance Company (1998)　他
著書：「さよなら、サイレント・ネイビー」(第四回開高健賞) 集英社 (2006)　他多数
[**所属**] 東京大学大学院情報学環・作曲指揮研究室　ベルリン空間音楽コレギウム芸術監督
[**所属学会**] 国際時空間設計学会 ISTD (2013-15 Presidential Chair)、情報社会学会　他。

Spatiotemporal Information Theories on Music

Ken ITO*

New frameworks of spatio-temporal analysis on music cognition are introduced and results are shown. First, the author enlarges conventional linear approaches for musical-verbal cognition of sound into non-linearity. Following the mechanism of human cochlea, sinusoidal decomposition of sound sources is defined. With this decomposition, series of decomposed sound fragments could be produced and by use of this, emergence of verbal cognition within human brain is clearly observed and quantitatively evaluated.

The author also enlarges non-linear approaches for acoustics sufficient for musical aims; short-term correlation function analysis is expanded into time-dependent region and new parameters which can explain dynamical characteristics of sound and voice are obtained. Using such basic methods of plethodynamics, we can also construct countermeasures against various kinds of network-media abuse of audio-visual contents.

Interfaculty Initiative in Information Studies The University of Tokyo

Key Words：music, verbal cognition, sinusoidal decomposition, brain imaging, time-dependent analysis of correlation functions, audio-visual contents