

群遅延情報を利用した音声の駆動情報の多重解像度分析について

河原 英紀¹、 Parham Zolfaghari²

¹和歌山大学システム工学部/ATR/CREST, ²CIAIR/名古屋大学

¹〒640-8510 和歌山市栄谷930

kawahara@sys.wakayama-u.ac.jp

あらまし 音声波形の群遅延特性を利用して、声門閉止等の音声を駆動する主要なイベント生起時刻とイベント属性を定量的に高精度に抽出する新しい方法を前報において提案した。この方法では信号を分析する時間窓のスケールのみをパラメタとしており、窓の中心時刻から平均時刻への写像の不動点としてイベントが求められる。本報告では、こうして求められた不動点が画像の多重解像度分析における零交差によるエッジ検出に相当することに注目し、スケール依存性を利用して音声の駆動情報の性質を調べる方法を提案する。また、主要な特異性と背景雑音とのP/N比を観測された持続時間から求める方法を導く。本報告では、実音声とEGGを同時収録したデータベースを用いた具体例を通じてこれらの方法の性質を紹介している。

キーワード 群遅延、基本周期、声門閉止、最小位相、多重解像度、EGG

Multi-resolution analysis of group delay based event detection for excitation attributes of speech sounds

Hideki Kawahara¹ and Parham Zolfaghari²

¹Faculty of Systems Engineering, Wakayama University/ATR/CREST, ²CIAIR/Nagoya University

¹930 Sakaedani, Wakayama, Wakayama, 640-8510 Japan

kawahara@sys.wakayama-u.ac.jp

Abstract A new procedure for event detection and characterization was proposed by one of the author based on a group delay and fixed point analysis. The proposed method enabled to detect precise timing and spread of speech event like a vocal fold closure. The proposed method consists of only one parameter on scale of events under study. Events are represented as fixed points of mapping from the window center location to the mean time. In this report, a structural analysis of fixed points in the time-scale plane was proposed taking advantages of the mathematical similarity of the proposed event detection methos with the multi-resolution zero-crossing analysis which are commonly found in visual scene analysis. A pulse plus noise model was also introduced to analyse an event attribute, duration. Demonstrations of the proposed method are given using real speech examples and a speech database with EGG information.

key words group delay, fundamental period, vocal fold closure, minimum phase, multi-resolution, EGG

1 はじめに

著者は、channel VOCODER[3]の原理に基づき、最近の計算機の性能の急速な向上を生かした高品質な音声分析変換合成方法STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram)を提案してきた[5, 6, 8]。この方法では、瞬時周波数と周波数領域での不動点の性質に基づく基本周波数抽出方法[7]、基本周波数に適応したスペクトログラムからの周期性の影響の除去[8]、合成音源の群遅延特性の操作[9, 8]を組み合わせることにより、分析合成方式の特長であるパラメタの柔軟な操作性を保ったまま、高品質の音声の合成を可能にしている。しかし、音源情報の表現と抽出に関する検討は、他の要素と比較すると遅れており多くの問題をかかえていた。特に、合成音源の群遅延特性に関しては、適切な数理モデルに基づく分析方法を欠いていたため、試行錯誤によって求められた固定値が用いられていた。また、破裂子音や摩擦子音のような非周期的信号についても、適切なモデル化を欠いていることによる品質上の問題が指摘されていた[11, 13]。

これらの問題を解決すべく、著者は信号をイベントの集まりとして表現することを狙って、群遅延特性の最小位相成分の補償に基づく分析方法を提案した[14]。この方法では、本報告の付録に示すように、分析のための時間窓の中心時刻から窓に含まれるエネルギーの重心として求められる平均時刻への写像の不動点としてイベントが求められ、最小位相システムの群遅延特性を用いてイベント位置のずれの補償が行われる。また、イベント周辺でのエネルギーの拡がりを持続時間として表現することで、音声の駆動の種類を表現することを提案した。様々な種類の合成音声をういた評価や、合計28人の男女が発声した音声とEGG(ElectroGlottoGraph)信号を同時に収録したデータベース[12]を用いた評価により、声門閉止に対応するイベント位置の推定については、提案した方法による推定誤差の標準偏差が $100\mu\text{s}$ 程度という非常に高い精度であることが示された。しかし、この分析方法をSTRAIGHTの音源情報の抽出部分に適切に組込むためには、更に詳しく様々な性質を明らかにして行かなければならない。

本報告では、まず、提案したイベント抽出方法が画像の解析でよく用いられる零交差[10, 15]による特徴抽出に対応することを示し、多重解像度での分析の必要性について説明する。次いで、その性質を時間-スケール領域で調べることを容易にするため、持続時間をより物理的意味の明確なP/N比(パルスのエネルギーと背景雑音のエネルギーの比)に対応させる方法を明らかにする。また実音声の分析例を用いて、音源情報の抽出法としてどのような実装が望ましいか、どのような応用があるかについて議論するとともに、聴覚系における情報処理との関連についても論じる。

2 イベント抽出と多重解像度分析

時刻 u に中心がある時間窓 $w(t-u)$ を信号 $s(t)$ に掛けてできる信号(以下では観測信号と呼ぶ)を $x(t, u)$ と表すこととする。 $x(t, u)$ のエネルギーの重心位置として定義される平均時刻 $\langle t(u) \rangle$ を用いると、イベントの時刻 t_e は、以下の条件を満たす時刻として求められる¹。

$$\{t_e\} = \{u | \langle t(u) \rangle = u, \frac{d\langle t(u) \rangle}{du} < 1\}. \quad (1)$$

ここで、窓関数 $w(t)$ として付録の式7で示すガウス型のものを用いると、平均時刻を次のように書き換えることができる。

$$\begin{aligned} \langle t(u) \rangle &= \int \tau |x(\tau, u)|^2 d\tau \\ &= - \int s^2(\tau) \frac{dw(\tau-u)}{du} d\tau + u \end{aligned} \quad (2)$$

式2の第一項を $\lambda(u)$ と表して式1に代入して整理すると、以下を得る。

$$\{t_e\} = \{u | \lambda(u) = 0, \frac{d\lambda(u)}{du} > 0\}. \quad (3)$$

すなわち、写像の不動点として求められるイベントは、窓関数の時間微分をインパルス応答とするフィルタに自乗した信号を通した出力 $\lambda(u)$ の零交差点でもある。このように、 $\lambda(u)$ の零交差点を抽出することは、信号のパワーをある時間スケールで眺めた場合のエッジを検出していることに相当することが分かる。画像の場合と同様に、時間スケールを変えた場合に零交差点がどのように変化するかを調べれば、信号についての豊かな情報が得られる。

2.1 因果律に起因する偏りの補償

観測される音声信号は、例えば声門における体積流の不連続のようなイベントによって声道が駆動されて生ずる。このようなシステムでは、通常は出力は入力に先行しないという因果律が成立していなければならない。この拘束条件のため、零交差の位置は駆動源のイベントの時刻から遅れたところに生ずる。このため、音声信号に対する多重解像度分析では、画像では必要の無かった因果律に起因する偏りの補償を行わなければならない。ここでは、付録に示すように、因果律のみを要請することで得られる振幅スペクトルと位相との関連を利用して最小位相システムの群遅延特性を求め、この偏り(遅延)を補償する。

図1に示すように、実際の音声を例として、この補償について説明する。資料は男性の発声した「アイウエオ」という連続母音である。ここでは、開始部分にある零交差点の一つを例として取り上げた。図1の左側のグラフの実線は、観測信号から求められた群遅延である。同じ

¹議論が煩雑になるのを避けるため、前回の報告で提案した定式化の骨子を付録に収録した。

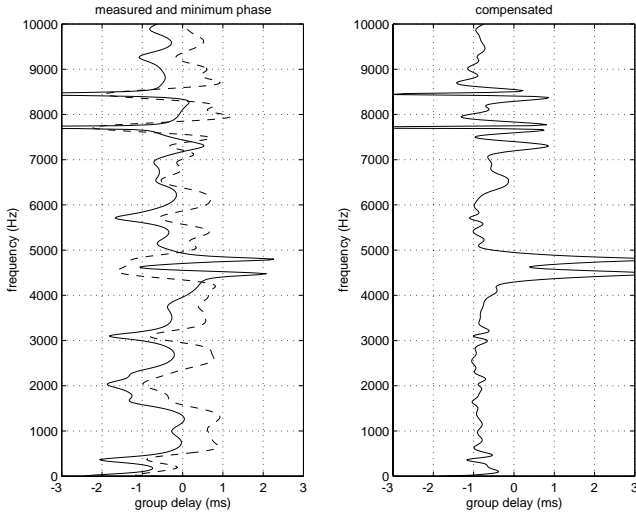


図 1: Measured group delay and compensated group delay. Left plot shows the measured group delay (solid line) and the minimum phase group delay (dashed line). Right plot shows the compensated group delay.

グラフの中の破線は、観測信号の振幅スペクトルから因果律を要請して求められた最小位相システムのインパルス応答に対応する群遅延である。1 kHz から 4 kHz の範囲では、両者は非常に類似している。図 1 の右側のグラフは、観測信号から求めた群遅延から因果律による成分を取り除いて得られる（仮想的な）駆動信号の群遅延を示す。全体に複雑な凹凸が取り除かれ、特に、1 kHz から 4 kHz の範囲では、 $-850\mu\text{s}$ の固定的な遅延量を示す直線にほぼ一致している²。ここで、先に求めた例交差の位置を $-850\mu\text{s}$ 移動させると、ちょうど、声門閉止の位置に対応する。このように、最小位相システムのインパルス応答を介することで、音声信号のように音源とフィルタによるモデルで表される信号の場合に特有の因果律に起因する偏りを補償することができる。

2.2 P/N 比による特徴付け

求められるイベント全てが同じ重要性を有するわけではない。あるイベントは非常にはっきりとしており、別のイベントはぼやけていることがある。このようなイベントの性質を表すものとして前報では持続時間という量を提案した。しかし、観測信号の持続時間は、観測に用いた時間窓のパラメタ σ_w に依存しており、多重解像度分析の指標としては使い難い。ここでは、背景雑音の存在を仮定して、イベントのエネルギーと背景雑音のエネルギーの比として定義される P/N 比を用いる方法を提案する。詳細は、付録に示すが、P/N 比 Q は、観測信号から求められた持続時間 σ_o と時間窓のパラメタ σ_w を用いて

²結局、坂野らの指摘した群遅延特性の形状の特徴 [1] の多くは、振幅周波数特性から求められる最小位相システムの群遅延によるものであったことが分かる。

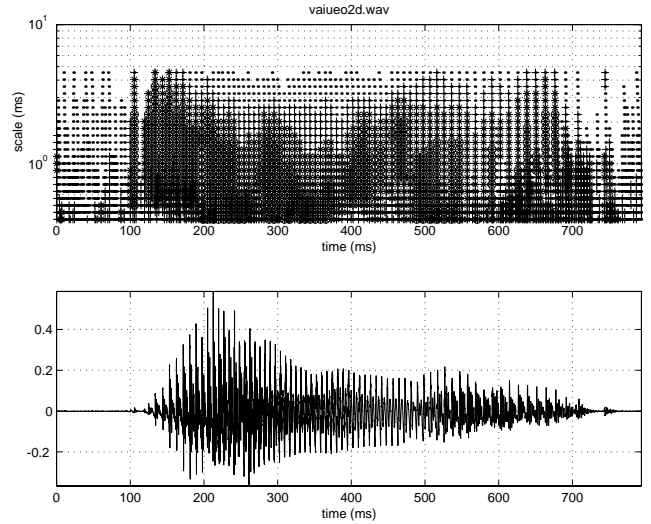


図 2: Scale dependency of the detected event. The upper plot shows extracted event locations for different scale parameter σ_w . The event marks are designed to produce dense symbol when P/N is high. The lower plot shows the corresponding waveform. The utterance is a Japanese vowel sequence /aieuo/.

次式により計算される。

$$Q = \frac{\sigma_w^2 - 2\sigma_o^2}{2\sigma_o^2} \quad (4)$$

3 実音声の多重解像度分析

以上で説明した手法を、実際の音声の分析例を用いて紹介する。音声資料は標準化周波数 22050Hz、16bit で直接パーソナルコンピュータに収録されたものである。以下の分析では、スケールのパラメタである σ_w を $0.4 < \sigma_w < 5$ の範囲にわたって $2^{1/6}$ のステップで変化させた。

連続母音 最初に用いる音声資料は、男性が連続的に発声した日本語母音「アイウエオ」である。図 2 に、抽出されたイベントと音声波形の全体を示す。上段は、イベントを時刻-スケール平面上に表示したものである。縦軸のスケールは、 σ_w により表わされている。それぞれのイベントを表すシンボルは、P/N 比の値が高いほど大きく濃いものを用いている。下段は、音声波形を表す。

図 3 に、図 2 の特徴的な部分の一部を拡大して示す。それぞれイベントと波形を組にして示している。なお、イベントの位置は、最小位相システムの群遅延を補正した後のものである。

図 3 の最上段の組は、発声の開始部分を示す。母音「ア」に対応する部分である。声門閉止に対応するイベントが 0.5 ms から 3 ms の広い範囲にわたって高い P/N 比を示していることが分かる。また、それらの強いイベントは、垂直の直線上にほぼ並んでおり、群遅延の補償がうまく

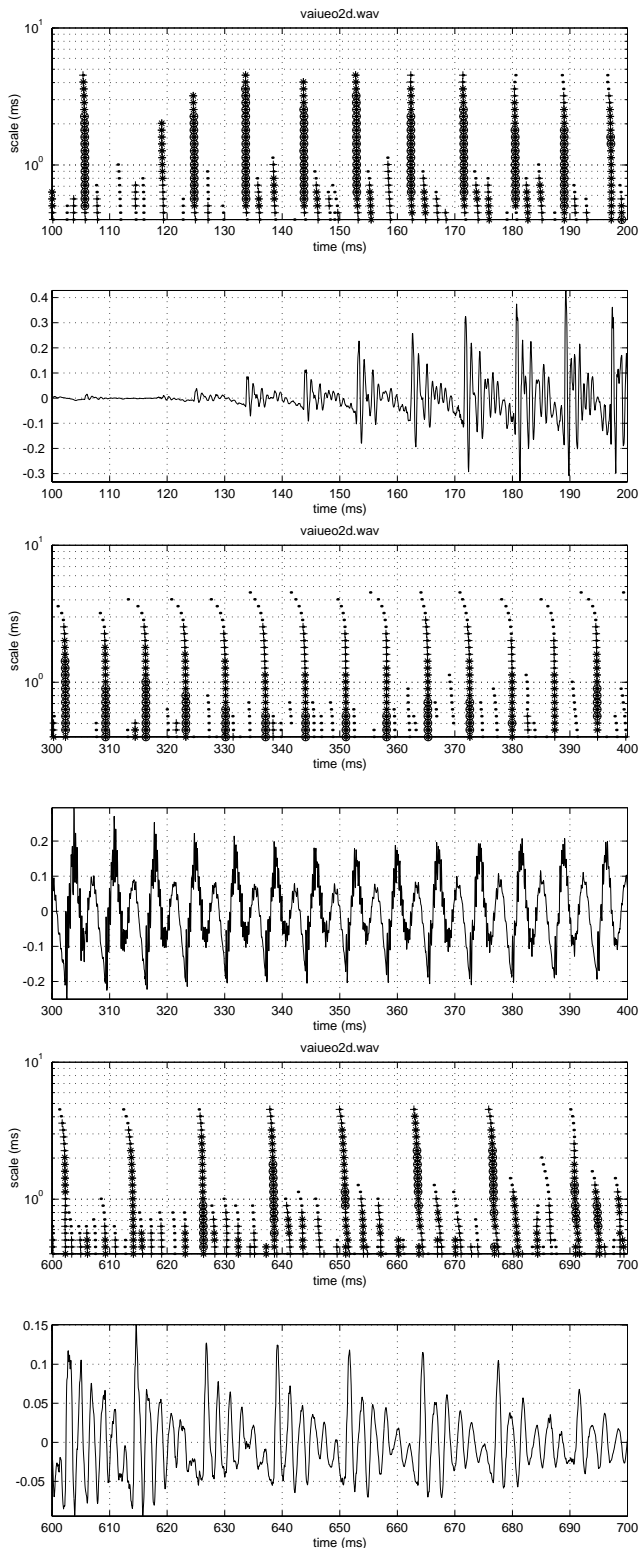


図 3: Scale dependency of the detected event. (Partial view)

働いてスケールの変化に伴う位置の変化が少なくなっていることを示唆している。なお、P/N 比は、音声信号の振幅とは直接的には対応していないことが分かる。

中段の組は、母音「イ」に相当する部分を示す。ここ

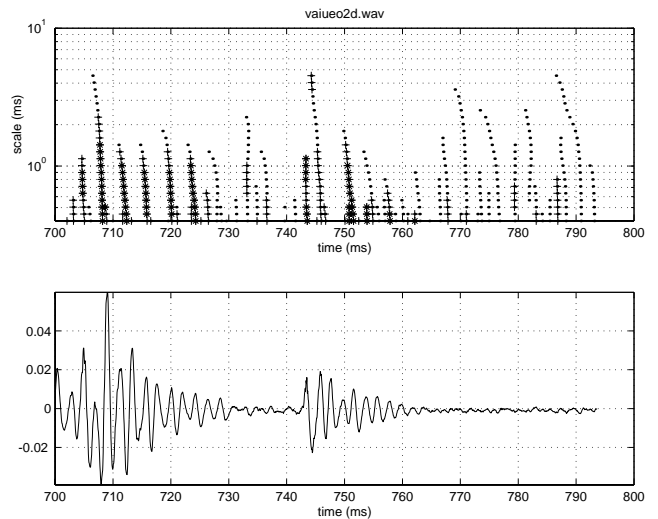


図 4: Scale dependency of the detected event. (Partial view)

でも同様に、強いイベントは直線上に並んでいる。この辺りは基本周波数が最大の部分であるため、直線上の強いイベントが現れるのはより小さなスケールの方になっている。

下段の組は、母音「オ」に相当する部分を示す。声門閉止に対応するイベントが支配的であることは他の部分と同様であるが、680 ms 以降に認められるように、それ以外の特異性に対応するイベントも強くなって来ていることが分かる。

この傾向は、発声の終了部分を示す図 4において更に顕著になる。ここでは、声門の開閉はもはや周期的ではなく、予想できないタイミングで発生している。また、声門閉止のイベントが弱くなったため、雑音に起因するイベントが見えるようになっている。

連続数字音声 図 5は、子音を含む例として取り上げた「ひやくにじゅうご」と発声した音声の全体である。発声者および収録条件は、図 2と同様である。

図 6に、特徴的な部分を幾つか拡大して示す。同様に、イベントの時間-スケール平面での表示と対応する音声波形を、時間軸をそろえて表示している。

図 6の上段の組は、「ひゃ」の部分に相当する。摩擦音の部分と半母音の部分は無声化しており、見えている母音の波形は「あ」に類似している。ここでも、有声音の部分では声門閉止に対応するイベントが支配的であることが分かる。無声部分と摩擦音の部分では、イベントの分布はスケールの縮小に応じて密になっている。

中段の組は、「く」の部分に相当する。ここでは、子音の破裂位置が非常に強いイベントとして分析に用いた全スケールで抽出されている。グラフの後半は鼻子音に移行している。鼻子音では声門閉止に対応するイベントは弱くなる。鼻子音では口腔を通る経路と鼻腔を通る二つ

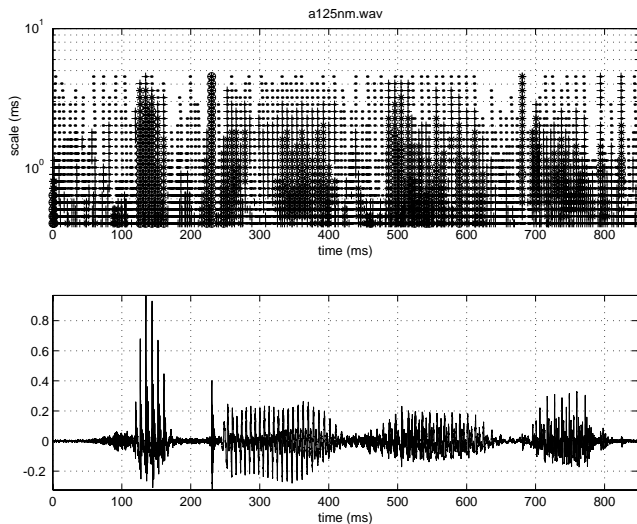


図 5: Scale dependency of the detected event. The upper plot shows extracted event locations for different scale parameter σ_w . The event marks are designed to produce dense symbol when P/N is high. The lower plot shows the corresponding waveform. The utterance is a Japanese number sequence /hyakunijyugo/.

の伝播経路がある。一般にこれらの経路長は異なるので、観測されるイベントが二重になり、P/N比は低下する。また、これらの子音における高域成分の低下も相対的な雑音レベルの増加を招き、P/N比の低下の要因となる。

下段の組は、「じゅ」の部分に対応する。有声摩擦音では、声門の開閉と乱流による雑音が同時に存在するため、イベントのP/N比は低下し声門閉止に対応するイベントが消失しているところも認められる。イベントの分布のスケール依存性は、無声部分や無声の摩擦子音と同様である。

図 7は、「ご」の子音部に対応する。680 ms 付近にある有声破裂音の破裂部分は、無声破裂音と同様に強いイベントとして抽出されている。それに先行する母音の部分では有声摩擦の影響が消滅していないためか、声門閉止は顕著ではなく、イベントの分布は無声部分や無声の摩擦子音と同様の傾向を示している。

このように多重解像度分析で得られるパターンは、駆動源のタイプに応じた特徴的なものとなる。基本周波数や単一のスケールの分析では駆動源の性質を特徴付けられない場合には多重解像度分析が有用であろう。

4 EGG データベースの分析

多様な発声に対する振舞いを調べるため、阿竹らによって作成されたEGGと音声を同時収録したデータベース [12] を用いて分析実験を行った。データベースには、唇前方 20cm の位置で収録された音声と、EGG 信号、視察

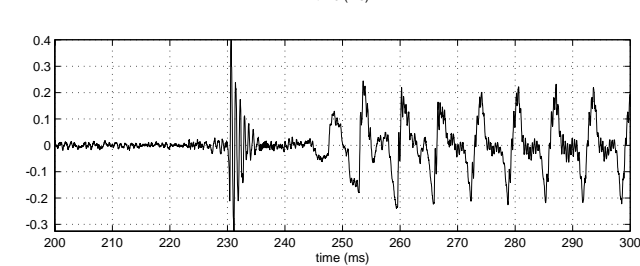
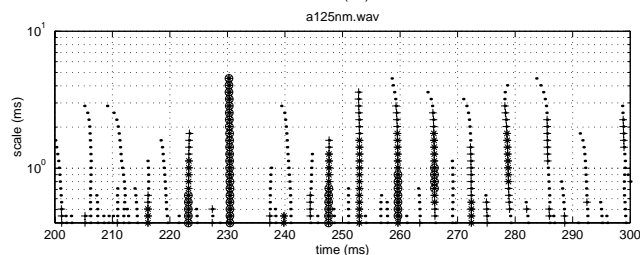
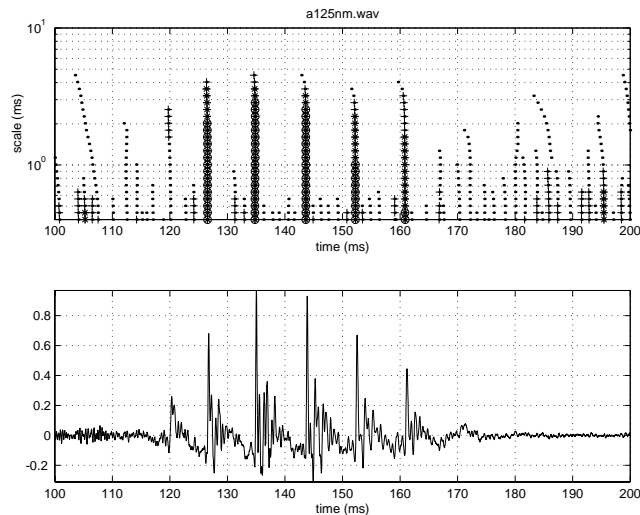


図 6: Scale dependency of the detected event. (Partial view)

による有声/無声判定情報が時間軸を揃えて収録されている。(唇からマイクまでの距離 20 cm で生ずる遅延は補正されて、唇位置での時間軸になっている。) 予備実験の結果に基づき、EGG 信号は、係数が [-0.5 1 -0.5] である

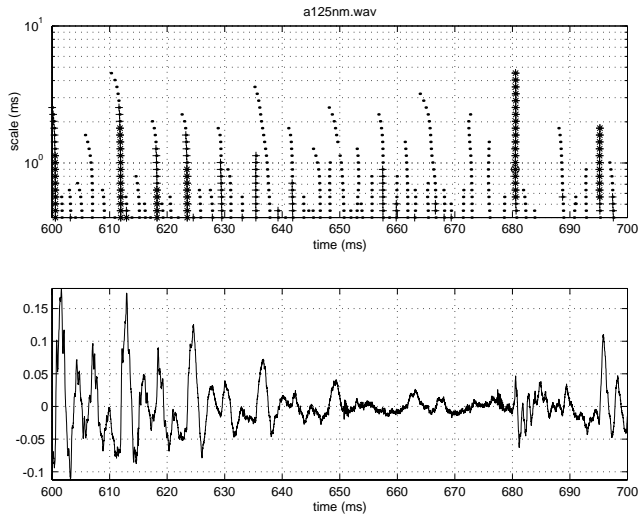


図 7: Scale dependency of the detected event. (Partial view).

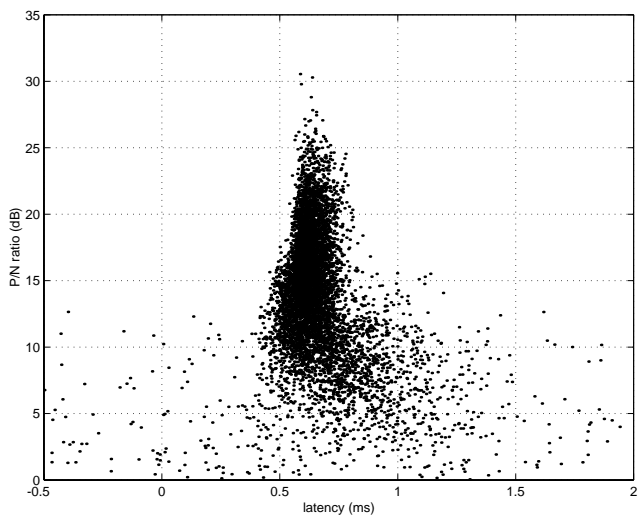


図 8: Distribution of event location and P/N ratio for a male speaker (M04).

FIR フィルタを通したものを分析対象とした。

ここでは、P/N 比とイベント位置の分布を中心に紹介する。図 8 は、ある男性話者について、EGG から求められるイベント位置と最も近くにある音声から求められるイベント位置との時間間隔の分布を示す。 σ_w としては、多重解像度分析の結果に基づき、1 ms, 1.4 ms, 2 ms の 3 種類について調べた。ここでは $\sigma_w = 1$ ms の場合の結果を示す。また、周波数加重として 500 Hz から 4000 Hz の範囲を選択した。

図 8 のたて軸は P/N 比である。P/N 比が高いほど分布が鋭くなる傾向が認められる。

この傾向を、P/N 比をイベント選択の基準とした場合の EGG イベントと音声のイベントの時間間隔の標準偏

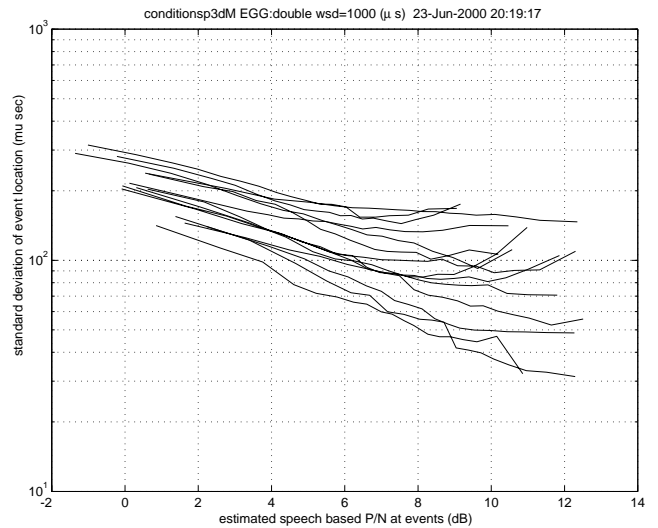


図 9: Standard deviations of location estimation errors in terms of minimum P/N for event detection. Each line represents the result of a single male speaker.

差を指標として調べた結果を図 9 に示す。図中の 14 本の実線は、それぞれの被験者に対応している。散布図で認められた傾向は、全話者で認められており、P/N 比がイベント位置の推定精度の指標として有用であることを示している。例えば、P/N 比が 7 dB 以上の部分をイベントとして選択すれば、イベント推定の標準偏差が $60 \mu\text{s}$ から $170 \mu\text{s}$ になることが分かる。女性の場合には、一般的に P/N はより低く標準偏差も 2 割程度大きくなるものの同様な傾向を示す。ただし、無声部分においても P/N が 7 dB を超えることは稀ではなく、単一のスケールによる分析だけで声門閉止部分だけを選択することは困難である。

5 討論

前報で提案したイベント抽出法が零交差を用いたエッジ検出としても解釈できることを示し、実音声を対象とした多重解像度分析の性質を調べてきた。このようないわばボトムアップの方法により見いだされる処理に対応するものが聴覚においても見いだされるかは大変興味深い問題である。少し思弁的な領域に踏み込んで群遅延と瞬時周波数の双対性を考慮すれば、イベントとピッチはそれぞれ『周波数方向に統合されるオブジェクト』と『時間方向に統合されるオブジェクト』を表していると言えるかも知れない。

本研究には、類似した様々な先行研究がある。例えば、尺度空間での零交差の分岐構造を調べることによる基本周期抽出の試みがある [15]。また、spline wavelet を用いて声門閉止部分を選択する試みも報告されている [4]。本報告での提案は、これらの先行研究に、音響信号に特有

の因果性による偏りの補償と持続時間あるいはP/N比による特徴付けを導入し体系化を図ったものと位置付けられよう。

6 まとめ

前報で提案した群遅延を用いた音声のイベント抽出法が時間窓関数の微分を用いた零交差点の抽出に相当することに注目し、多重解像度分析による音源特徴の分析について紹介した。また、前報で導入した持続時間というパラメタをP/N比として表現できることを示し、上記の多重解像度分析に応用した。本報告では、提案したイベント抽出法の定性的な性質を中心に紹介した。今後は、本方法を高品質音声分析変換合成システム[8]に組み込むと共に、音声・聴覚研究への応用を検討して行きたい。本方法は、これらの他にも、音声生成過程の研究・診断等、様々な方面への応用が可能であろう。

参考文献

- [1] Hideki Banno, Jinlin Lu, Satoshi Nakamura, Kiyohiro Shikano, and Hideki Kawahara. Efficient representation of short-time phase based on group delay. In *Proc. ICASSP'98*, pp. 861–864, Seattle, 1998.
- [2] L. Cohen. *Time-frequency analysis*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [3] H. Dudley. Remaking speech. *J. Acoust. Soc. Am.*, Vol. 11, No. 2, pp. 169–177, 1939.
- [4] Shubha Kadambe and Faye Boudreaux-Bartels. Application of the wavelet transform for pitch detection of speech signals. *Trans. IEEE Information Theory*, Vol. 38, No. 2, pp. 917–925, 1992.
- [5] 河原英紀, 増田郁代. 時間周波数領域での補間を用いた音声の変換について. 信学技報, Vol. EA96-28, , August 1996.8.
- [6] Hideki Kawahara. Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. In *Proceedings of IEEE int. Conf. Acoust., Speech and Signal Processing*, Vol. 2, pp. 1303–1306, Muenich, 1997.
- [7] Hideki Kawahara, Haruhiro Katayose, Alain de Cheveigné, and Roy D. Patterson. Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of f0 and periodicity. *Proc. Eurospeech'99*, Vol. 6, pp. 2781–2784, 1999.
- [8] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain de Cheveigné. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction. *Speech Communication*, Vol. 27, No. 3-4, pp. 187–207, 1999.
- [9] Hideki Kawahara, Minoru Tsuzaki, and Roy D. Patterson. A method to shape a class of all-pass filters and their perceptual correlates. *Tech. Com. Psycho. Physio. the Acoust. Soc. Jpn.*, Vol. H-96-79, pp. 1–8, 1996. [in Japanese].
- [10] David Marr. *Vision: A Computational Investigation into Human Representation and Processing of Visual Information*. Freeman and Co., New York, 1982.
- [11] Parham Zolfaghari, 河原英紀. Subjective evaluation of STRAIGHT. 音響学会秋季講演論文集, pp. 193–194, 1999.
- [12] 阿竹義徳, 入野俊夫, 河原英紀, 陸金林, 中村哲, 鹿野清宏. 調波成分の瞬時周波数を用いたピッチ推定方法の検討. SP99-170, 1999.
- [13] 阿竹義徳, 陸金林, 中村哲, 鹿野清宏, 河原英紀. S-TRAIGHTの分析合成方式パラメタの主観評価による検討. 音響学会春季講演論文集, pp. 205–206, 2000.
- [14] 河原英紀, 阿竹義徳. 音声の群遅延特性に基づく声門閉止等のイベント抽出について. SP99-171, 1999.
- [15] 高松亮, 木村勝己, 佐藤誠. 尺度空間フィルタリングを用いた音声波形のピッチ抽出に関する研究. SP91-3, 1991.

A エネルギー重心としてのイベント

ここでは、前報[14]で提案したイベント抽出方法の骨子を再掲する。本節の議論では、注目している信号のエネルギーが正規化されているものとする。次の節のP/Nに関する議論では、エネルギーを明示的に扱う。

A.1 時間領域での表現

信号 $s(t)$ が時間窓 $w(t-u)$ によって切り出されたものを観測信号 $x(t, u)$ と表すことにする。観測信号のイベントの位置 $\langle t(u) \rangle$ および持続時間 $\sigma_t(u)$ を以下によって定義する[2]。

$$\langle t(u) \rangle = \int t |x(t, u)|^2 dt \quad (5)$$

$$\sigma_t^2(u) = \int (t - \langle t \rangle)^2 |x(t, u)|^2 dt \quad (6)$$

$$x(t, u) = w(t-u)s(t)$$

ここで、 u は時間窓の中心位置を表す。

写像の不動点としてのイベント窓は、以下のようなガウス型のものを用いる。

$$w(t) = e^{-\frac{t^2}{2\sigma_w^2}}. \quad (7)$$

ここで、 σ_w は窓のスケールを定めるパラメタである。また、信号の絶対値の自乗の期待値も同様なガウス型を持つとしてモデル化する。

$$\langle |s(t)|^2 \rangle = e^{-\frac{(t-t_e)^2}{\sigma_s^2}}, \quad (8)$$

ここで、 t_e は実際のイベントの位置を表す。すると若干の計算の後、観測されるイベント位置が次のように窓の位置 u と、実際のイベント位置 t_e の加重平均として求められることが分かる。

$$\langle t(u) \rangle = \frac{\sigma_s^2 u + \sigma_w^2 t_e}{\sigma_s^2 + \sigma_w^2}. \quad (9)$$

ここで、 $\langle t(u) \rangle$ を、同じ時間軸上で表される窓の中心時刻 u から平均時刻 t への写像と見ると、イベントの位置は次

の条件を満たす不動点であることが分かる。

$$\{t_e\} = \{u | \langle t(u) \rangle = u, \frac{d\langle t(u) \rangle}{du} < 1\}. \quad (10)$$

音声信号のように、受動的なシステムが駆動されることで生ずる信号の場合には、求められるイベントの位置は駆動信号から求められるイベントの位置よりも遅れる。この遅れは、以下のような周波数領域の処理により補償することができる。

A.2 周波数領域での表現

観測信号の平均時刻と継続時間は、以下のように周波数領域で表すことができる [2]。

$$\langle t(u) \rangle = - \int \psi'(\omega, u) |S(\omega, u)|^2 d\omega \quad (11)$$

$$\begin{aligned} \sigma_t^2(u) &= \int \left(\frac{B'(\omega, u)}{B(\omega, u)} \right)^2 B^2(\omega, u) d\omega \\ &+ \int (\psi'(\omega, u) + \langle t(u) \rangle)^2 B^2(\omega, u) d\omega \quad (12) \end{aligned}$$

$$\begin{aligned} S(\omega, u) &= \frac{1}{\sqrt{2\pi}} \int x(t, u) e^{-j\omega t} dt \\ &= |S(\omega, u)| e^{j\psi(\omega, u)} = B(\omega, u) e^{j\psi(\omega, u)}. \quad (13) \end{aligned}$$

ここで、 $B(\omega, u)$ は観測信号の振幅スペクトルを表し、 $\tau_g(\omega, u) = -\psi'(\omega, u)$, where $'$ は群遅延を表す。なお、 $'$ は角周波数 ω に関する微分を表す。

最小位相成分による影響の補償 駆動されるシステムが因果律を満たすものである場合、群遅延と振幅周波数特性の間には一対一の対応関係がある。具体的には、ある振幅周波数特性に対応する最小位相系の群遅延 $\tau_\phi(\omega, u)$ は、次のように複素ケプストラム $C(q, u)$ を介して求めることができる。

$$\tau_\phi(\omega, u) = -\frac{d}{d\omega} \left(\text{imag} \left[\frac{1}{\sqrt{2\pi}} \int C(q, u) e^{j\omega q} dq \right] \right) \quad (14)$$

$$\begin{aligned} C(q, u) &= \begin{cases} 2c(q, u) & q > 0 \\ c(q, u) & q = 0 \\ 0 & \text{その他の場合} \end{cases} \\ c(q, u) &= \frac{1}{\sqrt{2\pi}} \int \log B(\omega, u) e^{-j\omega q} d\omega \quad (15) \end{aligned}$$

ここで q は、ケフレンシーである。

この最小位相による群遅延を補償することで、駆動信号のイベント位置 $\langle \tilde{t}(u) \rangle$ と継続時間 $\tilde{\sigma}_P(u)$ を求める。

$$\langle \tilde{t}(u) \rangle = - \int (\psi'(\omega, u) + \tau_\phi(\omega, u)) |S(\omega, u)|^2 d\omega \quad (16)$$

$$\tilde{\sigma}_P^2(u) = \int (\psi'(\omega, u) + \langle \tilde{t}(u) \rangle + \tau_\phi(\omega, u))^2 B^2(\omega, u) d\omega \quad (17)$$

B P/N比としての継続時間

ここでは、継続時間として求められる量のもう一つの表現を導く。まず、時間軸上にデルタ関数状の特異性があるとしてイベントをモデル化する。また、分散が ε^2 であるような無相関な背景雑音が存在していると仮定する。このモデルでは、イベント自体の継続時間は0である。しかし、背景雑音が存在することにより、見かけ上の継続時間は0以上となる。以下では、信号 $s(t)$ が時間窓 $w(t-t_e)$ によって切り出されたものを観測信号 $x(t)$ と表すことにする。まず、仮定より以下が成立する。

$$\langle x^2(t) \rangle = \varepsilon^2 e^{-\frac{(t-t_e)^2}{\sigma_w^2}} + \eta^2 \delta(t-t_e)$$

ここで、 σ_w は、窓のスケールを定めるパラメータであり、 η^2 はイベントのエネルギーである。また、イベントは t_e にあるとした。なお t_e を原点に置いても一般性を失わないので、以下では、 $\tau_e = 0$ とする。

観測信号の継続時間 σ_o は、このイベント周りの2次のモーメントをエネルギー p_w^2 で正規化した正規化分散 $\langle t^2 \rangle$ の平方根として表される。

$$\begin{aligned} \sigma_o^2 = \langle t^2 \rangle &= \frac{1}{p_w^2} \int t^2 \left(\varepsilon^2 e^{-\frac{t^2}{\sigma_w^2}} + \eta^2 \delta(0) \right) dt \\ p_w^2 &= \int \langle x^2(t) \rangle dt \end{aligned}$$

正規化に用いられるエネルギー p_w^2 は、 $\langle x^2(t) \rangle$ を時間で積分して、次のように表される。

$$\begin{aligned} p_w^2 &= \int \varepsilon^2 e^{-\frac{t^2}{\sigma_w^2}} + \eta^2 \delta(0) dt \\ &= \varepsilon^2 \sigma_w \sqrt{\pi} + \eta^2 \quad (18) \end{aligned}$$

これらより、以下が得られる。

$$\begin{aligned} \sigma_o^2 &= \frac{1}{p_w^2} \int t^2 \left(\varepsilon^2 e^{-\frac{t^2}{\sigma_w^2}} + \eta^2 \delta(0) \right) dt \\ &= \frac{1}{p_w^2} \left[\int t^2 \varepsilon^2 e^{-\frac{t^2}{\sigma_w^2}} dt + \int t^2 \eta^2 \delta(0) dt \right] \\ &= \frac{\varepsilon^2 \sigma_w^3 \sqrt{\pi}}{2\varepsilon^2 \sigma_w \sqrt{\pi} + 2\eta^2} \\ &= \frac{\sigma_w^2/2}{1 + \frac{\eta^2}{\varepsilon^2 \sigma_w \sqrt{\pi}}} \quad (19) \end{aligned}$$

ここで、イベントのエネルギーと背景雑音のエネルギーとの比として定義されるP/N比を Q と表すことにする。また、式18の第一項が背景雑音のエネルギーを表し、第二項がイベントのエネルギーを表すのであることに注意し、 Q を用いて式19を書き換え、 Q について解く。

$$\begin{aligned} \sigma_o^2 &= \frac{\sigma_w^2/2}{1+Q} \\ Q &= \frac{\sigma_w^2 - 2\sigma_o^2}{2\sigma_o^2} \quad (20) \end{aligned}$$

これで、イベントの属性としてP/Nを観測値から求めることができる。