

零交差波による音声信号の復元について ——零交差に基づく音声信号処理の基礎的検討——*

荒井 隆行・吉田 裕一**

(上智大学理工学部)

(平成 3 年 8 月 21 日受付)

内容梗概 音声認識において古くから行われている 2 値化した情報を用いた音声信号処理に対し、それを発展させた零交差に関する音声信号処理の基礎的検討について述べた。2 値化された音声信号がある程度の明瞭度を持つことは以前から報告されている。しかし、音声認識や他の応用を考えた場合、それだけでは情報として不十分である。そこで本論文では、音声信号に対する 2 値化の理論的背景を考察し、その結果もとの零交差情報から得られる 2 値化信号において欠落した情報を補うため、スペクトル反転信号の零交差波を補助情報として用いた。実際には、 π rad シフトした信号を線形補間してスペクトル反転の近似とした。この両者の零交差情報を用いて復元された音声信号の評価を行い、音声情報抽出のための処理方式として応用の可能性を調べた。復元信号に対してはスペクトル評価を行うと共に、補助情報に対して表現するビット数を削減して音質の劣化について主観評価を行った。そして、これらの情報が音声信号の特徴を表していることを確認した。

キーワード 零交差, 解析信号, 零点, スペクトル反転

Zero-crossings, Analytic signal, Zeros, Spectrum-reversal

1. はじめに

音声信号処理の分野において零交差情報 (zero-crossings) を用いる試みは古くから行われており、例えば 1 ビット量子化に相当する 2 値化された音声信号をそのまま再生するとかなりの明瞭度が得られることがよく知られている^{1)~3)}。その後、音声信号伝送あるいは合成の目的などでかなり研究が進められた^{4)~15)}。信号の零交差に関する理論的な基礎研究は、主として米国の H. B. Voelcker^{16)~21)}, A. A. G. Requicha²²⁾, L. R. Morris²³⁾, A. Sekey²⁴⁾, R. J. Niederjohn^{11)~13)} らなどによって行われている。これらの研究者によって正弦波重畳法あるいは多重微分法などが提案されているが、これは結局時間軸上の A/D (analog-to-digital) 変換に外ならない。国内においては、堂下氏²⁴⁾ や平松氏^{25), 26)} などをはじめとする多くの研究者によって、解析信号理論や SSB (single-sideband) 音声信号などを通じて零交差に関する研究が行われてきた。しかし現在の音声信号処理全般から見れば、零交差波を用いる方法はごく一部にすぎない。これは現在 DSP (digital signal processor) 装置、

A/D 変換器などハードウェアが十分に高性能となり、大規模で複雑な信号処理も高速に行われることが、最大の理由と思われる。

このように実時間信号処理は A/D 変換器と DSP 装置を用いて行われるのが主流になりつつあるが、より一層の高速処理を行うためには階層的な処理を行うことが有効であると考えられる。そのためには音声処理装置に入力する前段階において補助的に簡単なアルゴリズムによる可能な限りのハードウェア処理を行い、その結果を補助的な情報として後段の処理装置に受け渡すというようなシステムが考えられる。ここで示す前段階の補助的な音声処理としては、有音検出、子音検出、音声区分などが考えられ、これらが既に前段階において高速に処理されていることは後段の処理系に対して相当な演算量の負担が軽減されると予測される。そしてハードウェアにおける実現が容易であるという観点から、零交差による信号処理が重要な意味を持つてくる。一方、音声特徴量の一つとして零交差情報は有力であり、音声区分や子音認識の補助情報として幾つかの試みが行われている^{27)~29)}。零交差波による信号処理は、現在の高速信号処理技術のもとでも原波形処理に比較してはるかに高速処理に適しかつ小規模計算機処理に向いており、実現されれば実用上の意義は大きい。例えば、次世代の補聴器を開発するにあたり子音検出への応用などがあげられる。補聴器においては各々の子音に対してそれぞれに適

* A restoration of speech signal from zero-crossings—A study of speech signal processing based on zero-crossings—.

** Takayuki Arai and Yuichi Yoshida (Faculty of Science and Technology, Sophia University, Tokyo, 102)

した効果的な信号処理をする必要があり³³⁾, また実時間で処理をする必要があるため簡単なアルゴリズムでなければならない。その点, 零交差による信号処理は子音検出への応用に期待される。

著者らは, 音声の小規模計算機による高速信号処理, 特に将来の実時間音声認識への応用に向けて零交差音声信号処理の研究を行っており, 零交差とパワーの情報をを用いた音声信号の音声識別においてはかなりの識別率を得ていることから³⁰⁾⁻³²⁾, 零交差情報が音声認識上かなり重要なパラメータとなりうる点に着目している。また, 著者らは, 音声区分において単なる呼気音(風圧)と音声の識別においては零交差情報が有効であることを見出し, このような零交差情報が子音検出に対しても有効であるという点に注目している。

このように音声認識などの音声処理においては階層的な処理が必要であり, その前段階である補助的なハードウェア処理において零交差を用いた有音部検出, 子音検出, 音声区分の分類などが有効であると考えられる。著者らは, 零交差情報の音声情報としての不完全性をスペクトル反転信号の零交差波が補完することを見出し, これら2種の零交差波を用いれば音声情報として十分であるとの見通しを得るに至った。本論文の目的は零交差理論を詳細に検討して上記の方式を導き出し, 零交差波からの音声復元によって実験的に零交差情報利用の可能性を検証しようとするものである。

2. 信号の零交差に関する理論

零交差に関する基礎的な理論として, 解析信号理論¹⁷⁾及び整関数理論²²⁾が重要である。そこで, ここではこれらの理論に基づき解析信号や実信号の零点とその配置について考察する。

2.1 解析信号の変数変換

実信号 $s(t)$ に対して, 解析信号 $m(t)$ を次式のように定義する。

$$m(t) = s(t) + j\hat{s}(t) \quad (1)$$

ここで, $\hat{s}(t)$ は $s(t)$ のヒルベルト変換 (Hilbert transform) で, 次式で示される。

$$\hat{s}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau \quad (2)$$

ここで, P はコーシーの主値 (Cauchy principal value) を表す。解析信号 $m(t)$ のフーリエ変換 $M(f)$ は, 負の周波数で0になる。

$$M(f) = \begin{cases} 0 & f < 0 \\ S(0) & f = 0 \\ 2S(f) & f > 0 \end{cases} \quad (3)$$

いま, $m(t)$ の帯域が有限であり, かつ周期信号であると仮定すると, 次式のように有限のフーリエ級数展開によって表現される。

$$m(t) = \sum_{k=0}^N c_k e^{j\omega_k t} \quad (4)$$

ここで, 時間 t を複素変数 $\zeta = \tau + j\sigma$ に置き換えた関数 $m(\zeta)$ を考えると, この $m(\zeta)$ は整関数 (entire function) となり零点 ζ_i のみによる表現が可能になる。

つまり,

$$m(\zeta) = c_0 \prod_{i=1}^N [1 - e^{j\omega_0(\zeta - \zeta_i)}] \quad (5)$$

従来, 解析信号 $m(\zeta)$ に対しては ζ -平面で論じてきたが, 周期信号を扱う上では,

$$z = e^{j\omega_0 \zeta} \quad (6)$$

と変数変換をした方が合理的で扱い易い。本論文では以後, 変数 z を用いて話を進めて行くことにする。この変数変換により, ζ -平面上の実軸, LHP (lower half plane), UHP (upper half plane) はそれぞれ z -平面上の単位円, 単位円の内側, 単位円の外側にそれぞれ写像される。

2.2 解析信号の零点と零交差の関係

前述のとおり, 有限帯域を持つ周期性の解析信号 $m(\zeta)$ は整関数でその零点 ζ_i のみによって表現されたが, $\exp(j\omega_0 \zeta_i)$ が多価関数であることから ζ_i は ζ -平面上に無限に存在した。ところが, 前節の変数変換を施すと z -平面上における零点 z_i を次のように定義することにより, 零点の数は有限個となる。

$$z_i = e^{j\omega_0 \zeta_i} \quad (7)$$

このとき, 解析信号 $m(z)$ は z -平面上の零点 z_i によって,

$$m(z) = c_0 \prod_{i=1}^N (1 - z_i^{-1} z) \quad (8)$$

と表される。

実信号 $s(t)$ が与えられると, その解析信号 $m(z)$ の零点 z_i の配置は一意に決定する。逆に ω_0 と c_0 が既知であれば, 零点 z_i の配置が与えられるとそれに対する実信号 $s(t)$ が一意に決まる。結局, 信号と零点は1対1対応の関係にあることになる。周期信号においては, z -平面上の単位円内に存在する零点に対して1周期内に二つの零交差が生じる。ただ単位円外の零点のみでも零交差を生じることもあるので, 実信号の零交差が必ず単位円内の零点のみによって生じるとは限らない³⁰⁾。しかし音声信号零交差に関しては, 単位円内の零点が情報として与えられていると考えると基本的には差し支えないと考えられる。

ここで, 代表的な音声信号に対する例を Fig. 1, 2 に図示する。音声信号において有音声の定常部を周期信号と見なし, その1周期における原波形とそれぞれに対する解析信号の z -平面上における零点を示した。Fig. 1 が母音 /a/, Fig. 2 が母音 /i/ にそれぞれ対応し, 各図において (a) が原波形, (b) が z -平面上の零点配置

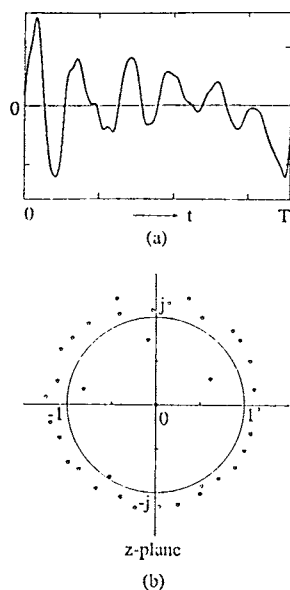


Fig. 1 Original signal of a Japanese vowel /a/.
(a) Waveform, (b) zeros.

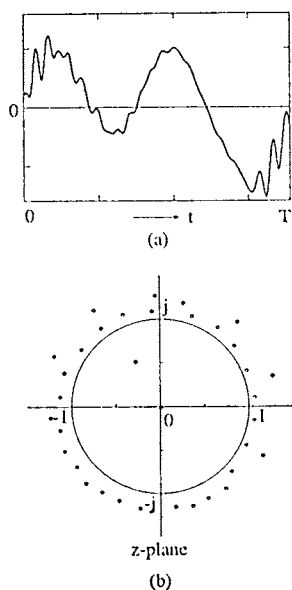


Fig. 2 Original signal of a Japanese vowel /i/.
(a) Waveform, (b) zeros.

を表している。これらの図を見ると、上記の性質が確認される。しかし、単位円外の零点の影響で場合によっては2回以上の零交差を生じることもありうる（ヒルベルト変換信号の零交差と照合すれば単位円外零点による零交差か否かは見出しうるが、計算処理が複雑になる。実用上、このようなケースは少ない）。

2.3 スペクトル反転と零点の鏡像写像

前節において、 z -平面における単位円内の零点によって実信号が零交差を生じることを説明した。すなわち、

実信号の零交差は単位円内の零点情報を与えていると考えることができるが、この零交差に含まれる情報だけでは、実信号を表現するのに明らかに情報が不足している。そこで、逆に零交差には現れてこない単位円外の零点情報を付加する必要がある。そのためにスペクトル反転による零点の鏡像写像を行うことを考える。

いま、解析信号 $m(z)$ に対して、そのスペクトルを反転させた信号 $\bar{m}(z)$ を考える。すなわち、

$$\bar{m}(z) = \sum_{k=0}^N c_k^* N^{-1} e^{j k \omega} \quad (9)$$

となる。この式を因数分解すると、

$$\bar{m}(z) = c_0 \prod_{i=1}^N (z - 1/z_i^*) \quad (10)$$

ここで、 z_i は式(8)で与えたものと同じであり、これは零点を鏡像写像したことに等価である。すなわち、 $\bar{m}(z)$ の零点 z_i' は、

$$z_i' = \frac{1}{z_i^*} \quad (11)$$

と表される。

ここで、先に示した音声信号に対するスペクトル反転の例を Fig. 3, 4 に図示する。Fig. 3 が母音 /a/, Fig. 4 が母音 /i/ にそれぞれ対応し、各図において (a) がスペクトル反転信号波形、(b) が z -平面上の零点配置を表している。Fig. 1 及び Fig. 2 の (b) と、この Fig. 3 及び Fig. 4 の (b) を比べると明らかなように、各零点は互いに鏡像写像の関係になっている。

2.4 鏡像関係にある零点による波形に関する二、三の考察

ここでは、実信号の波形の様子と解析信号の零点の関

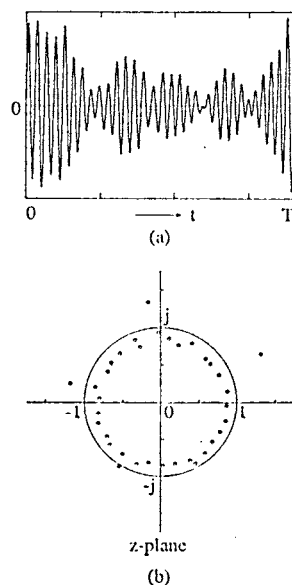


Fig. 3 Spectrum-reversal signal of a Japanese vowel /a/. (a) Waveform, (b) zeros.

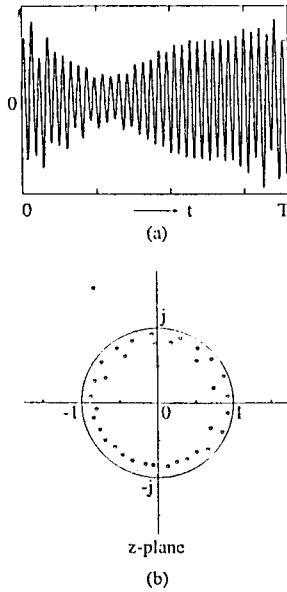


Fig. 4 Spectrum-reversal signal of a Japanese vowel /i/. (a) Waveform, (b) zeros.

係を具体的な例を通して見てみることにする。

1) 零点が1個存在する場合

次式のように、解析信号の零点が単位円内に1個だけ存在する場合を考えてみる。

$$z_1 = a^{-1} \tag{12}$$

ここで、 a は実数で $0 < a < 1$ とする。このようにすべての零点が単位円内に存在するときは、最小位相 (minimum phase; MP) の場合に相当する。このとき、解析信号 $m(z)$ は、

$$m(z) = 1 - az \tag{13}$$

と表されることから、実信号 $s(t)$ は、

$$s(t) = 1 - a \cos \omega_0 t \tag{14}$$

となる。 a は $0 < a < 1$ を満たすことから、 $s(t)$ は零交差を生じない。

逆に解析信号 $m(z)$ に対するスペクトル反転信号 $\bar{m}(z)$ は、定義から次式のように表される。

$$\bar{m}(z) = z - a \tag{15}$$

この $\bar{m}(z)$ の零点は、

$$z_1' = a \tag{16}$$

となり、 z -平面上で $m(z)$ の零点を単位円に関して鏡像写像したものになっており、このとき零点は単位円外に存在する。このようにすべての零点が単位円外に存在するときは、最大位相 (maximum phase; MaxP) の場合に相当する。

$\bar{m}(z)$ に対する実信号 $s(t)$ は、

$$s(t) = \cos \omega_0 t - a \tag{17}$$

となる。 a は $0 < a < 1$ を満たすことから、 $s(t)$ は1周期内で2回の零交差を生じることが分かる。

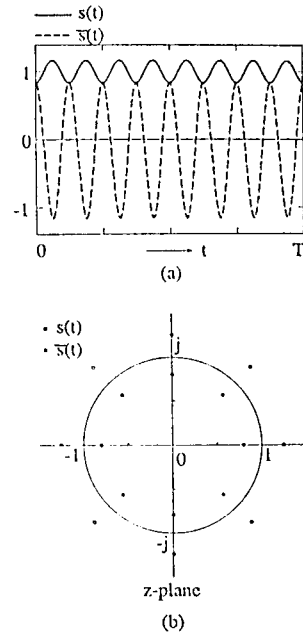


Fig. 5 Waveforms and zeros of $s(t) = 1 - a^N \cos N\omega_0 t$ with $a = 0.8$, $N = 8$, $\omega_0 = 2\pi/T$ (minimum phase) and the inversion (maximum phase). (a) Waveform, (b) zeros.

2) 零点が円周上に一樣に配置している場合

解析信号の N 個の零点が単位円内の同一円周上に一樣に配置している場合を考えてみる。例えば $N = 8$ の場合、Fig. 5 (b) に示す位置に零点が存在するものとする。すなわち、

$$z_i = a^{-1} e^{j2\pi i/N} \quad (i = 0, 1, \dots, N-1) \tag{18}$$

ここで、 $0 < a < 1$ であり、最小位相の場合に相当する。

このとき、解析信号 $m(z)$ は、

$$m(z) = c_0 \prod_{i=1}^N (1 - z_i z) \tag{19}$$

$$= c_0 \prod_{i=1}^N (1 - a e^{-j2\pi i/N} \cdot z) \tag{20}$$

$$= c_0 (1 - a^N z^N) \tag{21}$$

と変形されるので、実信号 $s(t)$ は、

$$s(t) = c_0 (1 - a^N \cos N\omega_0 t) \tag{22}$$

となる。 a は $0 < a < 1$ を満たすことから、Fig. 5 (a) に示すように $s(t)$ は零交差を生じない。

逆に解析信号 $m(z)$ に対するスペクトル反転信号 $\bar{m}(z)$ は、定義から次式のように表される。

$$\bar{m}(z) = c_0 (z^N - a^N) \tag{23}$$

$$= c_0 \prod_{i=1}^N (z - a e^{j2\pi i/N}) \tag{24}$$

この $\bar{m}(z)$ の零点は、

$$z_i' = a e^{j2\pi i/N} \tag{25}$$

となり、Fig. 5 (b) に示すように z -平面上で $m(z)$ の零点を単位円に関して鏡像写像したものになっている。

この場合は、最大位相に相当する。 $m(z)$ に対する実信号 $s(t)$ は、

$$s(t) = c_0 \cos(N\omega_0 t - a^N) \quad (26)$$

となる。 a は $0 < a < 1$ を満たすことから、Fig. 5 (a) に示すように $s(t)$ は 1 周期内で 16 回の零交差を生じることが分かる。

3. 信号及びそのスペクトル反転信号の各零交差波による信号の復元

3.1 実信号の実零点 (RZ) と複素零点 (CZ) による表現と各零交差波との対応

実際の周期信号 $s(t)$ を実零点 τ_i と共役複素零点对 $(\tau_i, \pm\sigma_i)$ で表現すると、次式のようになる²⁹⁾。

$$s(t) = A \prod_{RZ} \sin \frac{\omega_0}{2}(t - \tau_i) \cdot \prod_{CR} [\cosh \omega_0 \sigma_i - \cos \omega_0(t - \tau_i)] \quad (27)$$

ここで、 A は実定数である。また、RZ で示した因子を RZ (real zero) 信号 $s_{RZ}(t)$ と呼び、CZ で示した因子を CZ (complex zero) 信号 $s_{CZ}(t)$ と呼ぶことにする。もし複素零点 CZ が一つも存在しないとき、すなわち $s(t)$ が RZ 信号のときは最大位相である。言い換えると、最大位相のときは実零点 RZ のみで信号 $s(t)$ を表現できる。信号の零交差は RZ 信号の実零点 RZ と一致することは明らかであり、容易に求めることができる。それに対し、CZ 信号は零交差しないリップルを表しているが、これを正確に求めるには式全体を因数分解する必要があり、簡単に求めることは一般に困難である。そこでこの CZ 信号を簡単に推定することを考える。ところで 2 章で考察したように、もし $s_{RZ}(t)$ の零点は存在せず $s_{CZ}(t)$ の零点だけが z -平面の円周上に一様に配置されていれば、その円周半径によりリップル振幅が決定するので、逆にリップル振幅から $s_{CZ}(t)$ を求めることができる。この際スペクトル反転を行えば鏡像写像によりリップルが零交差を生じるようになり、その零交差位置から $s_{CZ}(t)$ を求めることができる。ただしこの方法によって求められた $s_{CZ}(t)$ は式 (27) において実零点が存在しないと仮定した場合であるので、実際波形においては $s_{RZ}(t)$ の乗算効果を補償する必要がある。ただ実際波形においては、ある隣接する二つの零交差間隔内ではほぼ正弦波の半波上にリップルが重畳していると見なせる。そこで波形を正弦半波とリップルの積であると仮定し、そのスペクトル反転信号を求めることにする。

3.2 正弦半波による波形近似

実信号の零交差情報から RZ 信号を復元することを考える。前述のとおり、零交差が検出されればその位置から正弦波の積によって RZ 信号が復元されるが、実用上は正弦波を幾つも掛け合わせることを考えると計算

量が多くなることが考えられる。また、零交差を生じない区間が長くなるとその区間において RZ 信号の振幅が著しく大きくなるなどの問題が生じる。そこで、ここでは RZ 信号に対して零交差間隔に比例した振幅を持つ正弦半波の連結によって近似することを考える。このとき、連結部において微係数の連続性が保たれている。この近似によって得られた信号を、近似 RZ 信号 $s_{\text{sin}}(t)$ と呼ぶことにする。この近似 RZ 信号は RZ 信号に比べ遮断特性があまり良くないが、低周波部のスペクトルピークはよく保たれている³⁰⁾。

3.3 線形補間による疑似スペクトル反転

前述のように信号波形の CZ 信号を直接求めることは困難であり、実際の計算においては信号が正弦半波とリップルの積であると仮定する。そしてそのスペクトル反転信号の零交差情報から CZ 信号を推定することを考える。実用性を考えるとスペクトル反転信号を正確に得るためには複雑な計算を要するので、より簡単に求める方法として疑似スペクトル反転信号を考える。

ある離散信号系列 $s(n)$ に対して、次のような各サンプルの極性を交互に反転させた信号 $s_{\pm}(n)$ を考える。

$$s_{\pm}(n) = (-1)^n s(n) \quad (28)$$

$s_{\pm}(n)$ の z 変換 $S_{\pm}(z)$ は、

$$S_{\pm}(z) = S(-z) \quad (29)$$

と表され、これは $z = \exp(j\omega)$ とした場合、 $s_{\pm}(n)$ のスペクトルが $s(n)$ のスペクトルを周波数軸上で π rad シフトしたものになっている。そこで、ここでは $s_{\pm}(n)$ を $s(n)$ に対する π rad スペクトルシフト信号と呼ぶことにする。

もしこの π rad スペクトルシフト信号について、角周波数が $|\omega| > \pi$ [rad] の範囲で $S_{\pm}(z) = 0$ にできれば、 $s(t)$ のスペクトル反転信号を得ることができる。しかし、スペクトル反転信号を正確に得るには計算量が多くなるので、 π rad スペクトルシフト信号 $s_{\pm}(n)$ のサンプル間を線形補間することによってスペクトル反転を近似することを考える。

隣接するサンプル間を線形補間するということは、三角波とのたたみ込みに相当する。これは周波数領域で考えると、sinc 関数の 2 乗の周波数特性を乗じることと等価である。Fig. 6 にその周波数特性を示す。この図から分かるように、角周波数が $|\omega| > \pi$ [rad] の範囲で完全には減衰しきっていないことから、正確なスペクトル反転にはなっていない。このように線形補間を用いて得られた信号を、疑似スペクトル反転信号と呼ぶことにする。音声信号に関しては、後述の実験からこの疑似スペクトル反転信号が十分に良い近似となっていることが確かめられている。

3.4 CZ 信号の推定

前節までで CZ 信号を求めるために信号が正弦半波

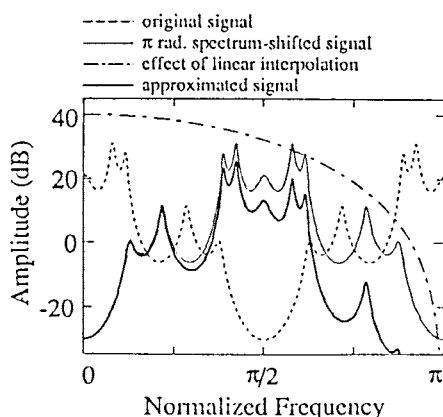


Fig. 6 Approximation of spectrum-reversal.

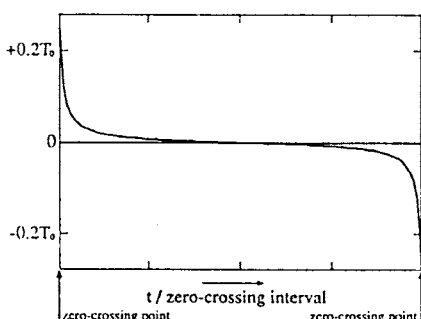


Fig. 7 Correction of zero-crossing for supplementary signal.

Table 1 Conditions for the experiment.

Sampling	10 kHz
Quantization	12 bit
Low pass filter	3 kHz
Speech sounds	five Japanese vowels
Data length	204.8 ms

とリップルの積であると仮定し、その π rad スペクトルシフト信号の零交差を求めるまでを述べたが、更にこの零交差を用いて CZ 信号を求めることを考える。正弦半波を基準と考えれば一つの零交差間隔において信号は最小位相信号となり、そのスペクトル反転信号は最大位相信号となる。そこで理想的には最大位相信号の零交差位置からリップルである CZ 信号を復元することが可能である。しかし仮定や近似を行った結果として得られた零交差位置は誤差を含んでいる。そこでその誤差を補正するため、正弦波に対する疑似スペクトル反転信号の零交差を考える。正弦波に対してはスペクトル反転信号の零交差は理想的に等間隔に生じるが、疑似スペクトル反転信号の零交差は等間隔にはならず、ずれを生じる。そのずれを補正する曲線を Fig. 7 に示す。この図は、正弦波に対する疑似スペクトル反転信号の零交差が等間隔に配置するためのずれを表した曲線である。図において、横軸は 1 零交差間隔で正規化した時間を表して

いるのに対し、縦軸は実際の時間軸に対応しており、 $T_0[s]$ はサンプリング周期を表している。ある零交差間隔内における零交差位置をこの曲線の示すずれだけ時間軸の正又は負の方向へずらすことによって、零交差が等間隔に生じるように補正される。これにより任意の信号に対しても零交差位置に対してこの補正を施し、その零交差位置から最大位相信号を求め、それをさらに極性反転することで CZ 信号を求めることにする。このようにして得られたリップルを補助信号と呼ぶことにする。零交差から得られる RZ 信号とこの補助信号を掛算することによって、もとの信号が復元されることになる。

3.5 信号復元の検証

信号の零交差情報と疑似スペクトル反転信号の零交差情報からもとの信号が復元されることを述べたが、この両零交差情報のみを用いて実際の音声信号に対する復元実験を行った。Table 1 に示すように音声試料は成人男性による日本語 5 母音の定常部を用いた。各母音に対して、まず正弦半波によって近似 RZ 信号 $s_{\text{sin}}(t)$ を求めた。そして更に、疑似スペクトル反転信号の零交差位置を求め補正をした後、その零交差位置から最大位相信号を求め更に極性反転を行うことによって補助信号 $scz(t)$ を計算した。そして、両者の積から求められた復元信号ともとの信号とを比較した。/i/ と /u/ に関しては、それらの近似 RZ 信号が非常に似た波形になっており、両者を区別するのは困難である。それに対して補助信号をそれぞれに付加することによって、復元波形は /i/ と /u/ の違いが明確に現れてくると考えられる。このことから、両者についての復元結果を 1 ピッチ周期に関して Fig. 8, 9 に示す。また、復元信号の周波数情報を論じる上で聴覚的な評価が重要となることから、原波形と復元波形のスペクトルにおける比較を行った。その結果を Fig. 10, 11 に示す。/i/ に関しては、ホルマント周波数を含め両者のスペクトルがよく類似しているのが分かる。これは特に /i/ の場合、第 1 と第 2 ホルマント周波数が比較的離れていることに起因すると考えられる。この二つのホルマント周波数が離れていることにより、低い周波数成分に高周波が重畳されたような波形と見なせるので、低い周波数成分はもとの信号の零交差位置でほぼ決定され、また高周波成分は疑似スペクトル反転信号の零交差位置によって表現されていると考えられる。

3.6 補助信号に必要なビット数

前節の実験では音声信号を零交差情報によって復元したが、その際表現するビット数については制限を加えなかった。そこで、復元する際に補助信号を表現するビット数を減らしていった場合についての復元実験を行い、聴覚実験による主観評価を行った。

まず近似 RZ 信号 $s_{\text{sin}}(t)$ については 2 値化信号を保

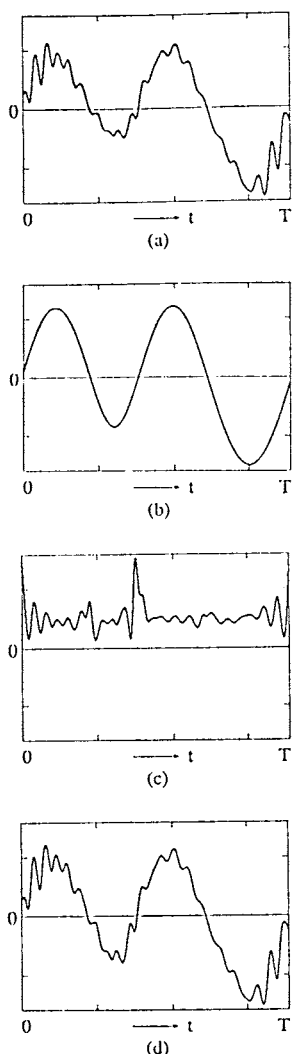


Fig. 8 Restoration of a Japanese vowel /i/. (a) Original signal, (b) approximated RZ signal, (c) supplementary signal from spectrum-reversal, (d) restored signal.

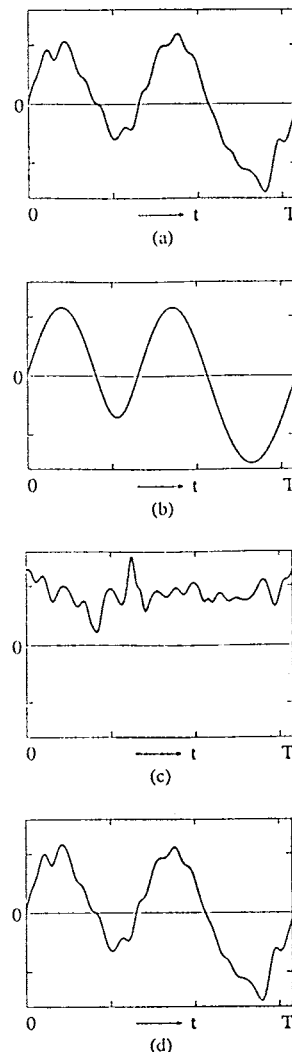


Fig. 9 Restoration of a Japanese vowel /u/. (a) Original signal, (b) approximated RZ signal, (c) supplementary signal from spectrum-reversal, (d) restored signal.

存することに相当するので、1サンプルにつき1ビットを割り当てる。一方、補助信号 $scz(t)$ に関しては1サンプルに割り当てるビット数を変えて、補助信号を表現した。そして、それらの積によって得られる復元信号の音質について評価した。音声試料は前述の実験と同様に日本語5母音を用い、各母音について204.8msで切り出したものを用いた。

実験は、各母音に対し補助信号に1サンプルにつき0~5ビット割り当てたもののほか、比較のためとの信号とそれを1~6ビットで量子化したものの計13種類について主観評価を行った。評価の方法は、5母音13種類全65パターンの音声をランダムに用意し、11人の被験者に対してそれぞれの音声を3回ずつ提示する

ことによってどの母音であるかを判定させた。そしてその明瞭度を、各割り当てビット数に対する正答率 (articulation score) として求めた。このとき補助信号は、割り当てビット数を制限するためにそのリップルを平均値に対して一定範囲に抑え、その最大リップル率を $\pm 50\%$ とした。その評価実験の結果をTable 2に示す。

Table 2を見ると分かるように、2値化信号については40%弱程度の正答率であった。これは前にも述べたとおり、2値化された音声信号がある程度の明瞭度を保存するが、零交差のみでは情報が不十分であることを裏付けている。近似RZ信号 $s_{in}(t)$ については、含まれる情報としては2値化信号と同等であると考えられる

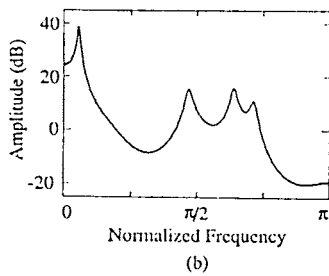
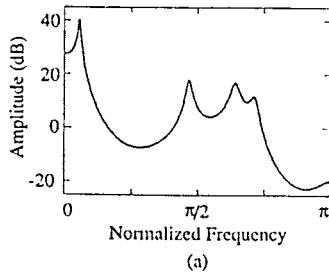


Fig. 10 Spectrum of the Japanese vowel /i/.
(a) Original signal, (b) restored signal.

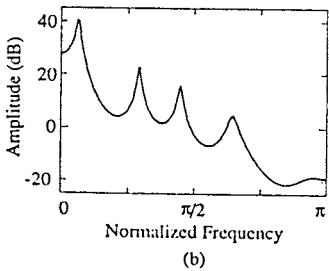
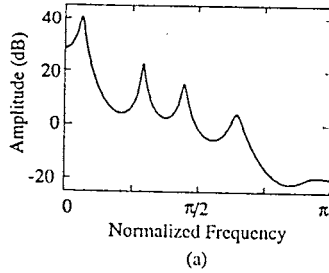


Fig. 11 Spectrum of the Japanese vowel /u/.
(a) Original signal, (b) restored signal.

が、60%程度と多少正答率が上がっていた。更にここで注目すべき点は、補助信号を加えたことにより飛躍的に正答率が良くなったことである。例えば、補助信号に対して1サンプルにつき1ビットを割り当てた場合と同じ情報量を持つ2ビット量子化とを比較すると、その正答率は約16%前後の差が生じている。補助信号に対して1サンプルにつき3ビットを割り当てた場合、同じ情報量を持つ4ビット量子化を行ったものと比較すると、その正答率はほぼ等しくなっていた。更に割り当てビット数を増やしていくと、最大リップル制限のため音質はそ

Table 2 Intelligibility vs. bits/sample.
(a) From zero-crossings (one bit for original zero-crossings).

Bits/sample	Articulation (%)
1	58.2
2	81.8
3	87.3
4	92.7
5	96.4
6	100.0

(b) Direct A/D conversion.

Bits/sample	Articulation (%)
1	36.4
2	65.5
3	83.6
4	92.7
5	100.0
6	98.2
Original signal	98.2

れほど良くならなかったが、正答率は100%近くにまで達した。以上により、特に補助信号を低いビット数で表現したときに零交差情報と補助情報を用いて音声信号を表現する意義が確認された。

4. おわりに

階層的な音声信号処理を考えた場合、その前段階の補助的なハードウェア処理に適している零交差についての基礎的な検討を行った。まず音声信号の零交差に関する理論的な検討を行い、解析信号や整関数表現による基礎的な考え方を述べた。そして、スペクトル反転信号の零交差情報を付加し、もとの信号の零交差情報を補完することによって、この音声信号を零交差情報でのみ表現した。更にこれらの零交差情報に対し、信号の復元実験を行った。そして、聴覚的並びにスペクトル上においても、ほぼ十分な復元信号を得ることができた。本研究で用いた方法は従来のような正弦波重畳法あるいは多重微分法と異なり、原波形の零交差情報を基本とするものである。補助情報となるスペクトル反転零交差情報の1サンプル当たりの割り当てビット数を極めて低くしても良好な明瞭度を保ちえたことから、補助信号は明瞭度の上からは多くの情報量を有していないことが判明し、零交差による音声処理への応用の十分な可能性を示したものと言える。零交差情報は S/N (signal/noise) 比が低い環境において不利であることは本質的な欠点であるので、例えば零交差処理のみによる音声認識は困難であろうが、補助的な処理としては有効であると結論づけられる。本研究においては母音を用いた復元実験を行ったが、子音を含むような音声(例えば「本日は晴天なり」)に対しても良好な復元音声を得ており、これは連続音声

に対しても零交差処理の有効性を示しているものと考えられる。今後はこの研究を基礎として、零交差波を用いて音声認識のための有音部検出、子音検出、音声区分などの応用を行っていく予定である。

文 献

- 1) J. C. R. Licklider and I. Pollack, "Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech," J. Acoust. Soc. Am. 20, 42-51 (1948).
- 2) J. C. R. Licklider, "The intelligibility of amplitude-dichotomized, time-quantized speech waves," J. Acoust. Soc. Am. 22, 820-823 (1950).
- 3) G. D. Ewing and J. F. Taylor, "Computer recognition of speech using zero-crossing information," IEEE Trans. Audio Electroacoust. AU-17, 37-40 (1969).
- 4) N. Bom and B. W. Conoly, "Zero-crossing shift as a detection method," J. Acoust. Soc. Am. 47, 1408-1411 (1970).
- 5) I. J. Good, "The loss of information due to clipping a waveform," Inf. Control 10, 220-222 (1967).
- 6) B. F. Logan, Jr., "Information in the zero crossings of bandpass signals," Bell Syst. Tech. J. 57, 487-510 (1977).
- 7) L. R. Morris, "The role of zero crossings in speech recognition and processing," Ph. D. Dissert., Imperial Col. of Science and Technology, London, U. K. (1970).
- 8) S. M. Kay and R. Sudhakar, "A zero crossing-based spectrum analyzer," IEEE Trans. Acoust. Speech Signal Process. ASSP-34, 96-104 (1986).
- 9) W. Bezdel and H. J. Chandler, "Results of an analysis and recognition of vowels by computer using zero-crossing data," Proc. IEE 112, 2060-2066 (1965).
- 10) R. W. A. Scarr, "Zero-crossings as a means of obtaining spectral information in speech analysis," IEEE Trans. Audio Electroacoust. AU-16, 247-255 (1968).
- 11) R. J. Niederjohn, "A mathematical formulation and comparison of zero-crossing analysis techniques which have been applied to automatic speech recognition," IEEE Trans. Acoust. Speech Signal Process. ASSP-23, 373-380 (1975).
- 12) R. J. Niederjohn and M. Lahat, "A zero-crossing consistency method for formant tracking of voiced speech in high noise levels," IEEE Trans. Acoust. Speech Signal Process. ASSP-33, 349-355 (1985).
- 13) R. J. Niederjohn, M. W. Krutz and B. M. Brown, "An experimental investigation of the perceptual effects of altering the zero-crossings of a speech signal," IEEE Trans. Acoust. Speech Signal Process. ASSP-35, 618-625 (1987).
- 14) 坂井利之, 大谷謙治, "零交差波による音声合成システムについて," 信学会誌 52-C, 704-711 (1969).
- 15) 富田真治, 大谷謙治, 坂井利之, "零交差波を用いた合成音声多重出力システム," 信学会誌 55-D, 149-155 (1972).
- 16) H. B. Voelcker, "Demodulation of single-sideband signals via envelope detection," IEEE Trans. Commun. COM-14, 22-40 (1966).
- 17) H. B. Voelcker, "Toward a unified theory of modulation," Proc. IEEE 54, 340-353, 735-755 (1966).
- 18) H. B. Voelcker, "Generation of digital signalling waveforms," IEEE Trans. Commun. COM-16, 81-93 (1968).
- 19) H. B. Voelcker, "Zero-crossing properties of angle-modulated signals," IEEE Trans. Commun. COM-20, 307-315 (1972).
- 20) H. B. Voelcker and A. A. G. Requicha, "Clipping and signal determinism: Two algorithms requiring validation," IEEE Trans. Commun. COM-21, 738-744 (1973).
- 21) H. B. Voelcker, "Band-limited random-real-zero signals," IEEE Trans. Commun. COM-21, 933-936 (1973).
- 22) A. A. G. Requicha, "The zeros of entire function: Theory and engineering applications," Proc. IEEE 68, 308-328 (1980).
- 23) A. Sekey, "A computer simulation study of real-zero interpolation," IEEE Trans. Audio Electroacoust. AU-18, 43-54 (1970).
- 24) S. Doshita, "Studies on the analysis and recognition of Japanese speech sounds," Ph. D. Thesis, Kyoto Univ. (1965).
- 25) 平松啓二, "音声の S. S. B.-clipping 波に含まれる情報," 音響学会誌 18, 301-308 (1962).
- 26) 平松啓二, 熊川幸伸, "S. S. B.-clipping を利用した音声帯域の圧縮方式—Formac—," 音響学会誌 18, 310-319 (1962).
- 27) 岡部 直, 吉田裕一, "零交差及び差分信号の零交差を用いた無声破裂音の認識," 音講論集, 243-244 (1989.3).
- 28) 椎山弘隆, 吉田裕一, "簡単なラベルを用いた無声子音の認識," 音講論集, 137-138 (1990.3).
- 29) 荒井和博, 鬼山康人, 野村康雄, 山下洋一, 北橋忠宏, 溝口理一郎, "知識処理に基づく音声自動ラベリングシステム," 信学論 74-D-II, 130-141 (1991).
- 30) 荒井隆行, 吉田裕一, "解析信号による音声信号零交差の考察," 音響学会誌 46, 242-244 (1990).
- 31) 荒井隆行, 吉田裕一, "音声信号の零交差波形による復元について," 音講論集, 199-200 (1990.9).
- 32) 荒井隆行, 吉田裕一, "音声信号の零交差波による復元について," 信学技報 SP 90-78 (1991).
- 33) 鈴木良二, 吉住嘉之, 口片強司, 山田義則, 田中 豊, 河野 淳, 船坂宗太郎, "糖時マスキングを補償する音声強調方法の検討," Audiol. Jpn. 34, 335-336 (1991).



荒井 隆行

昭和41年11月12日生。平成元年上智大学理工学部電気電子工学科卒業。平成3年同大学院理工学研究科修了(電気電子工学専攻)。現在、同大学院博士後期課程在学中。信号処理、音声処理の研究に従事。日本音響学会、電子情報通信学会、IEEE 各会員。



吉田 裕一

昭28東大・工・電気卒。同年八坂電機(株)(現、富士通ゼネラル)入社。昭35東大・工・研究生。昭37上智大・理工・電気電子専任講師。現在、同大・理工・電気電子教授。工博。テレビスタジオ機器、トランジスタ回路、電子交換機、通信方式の研究に従事し、現在は音声認識の研究に従事。著書「電子回路」など。電子情報通信学会、電気学会、IEEE 各会員。