

リアルタイム音響構造配信システムの開発 Construction of Local Area Network for transmitting acoustic structure

入場 健仁[†] 吉田 秀樹^{††} 藤原 祥隆^{††} 岡田 信一郎^{††}
Takehito Iriba Hideki Yoshida Yoshitaka Fujiwara Okada Shin-ichirou

1. はじめに

従来、音響信号に含まれる周波数成分の抽出にはフーリエ変換が用いられ、高速な演算を実現するためのアルゴリズムとして高速フーリエ変換 (FFT: Fast Fourier Transform)[1] が知られている。しかしながら、非定常波を含む音響信号の時間周波数分析を考えた場合、時間分解能の点で充分とは云い難い。非定常波に対しても正確な周波数成分の抽出が可能な手法に一般調和解析 (GHA: Generalized Harmonic Analysis)[2] があるが、FFT に比べ桁違いに計算量が多く、やはり充分とは云い難い。そこで我々は、少ない計算量で非定常波も観察することができる周波数分析法として、半周期計時方式 (ITM: Interextremal Timer Methodology)[3] を使用してきた。それまで ITM による音響構造データは周波数情報だけであったが、振幅情報を付加し、原音の近似波形を再構成する試み [4] もなされた。また ITM により生成される音響構造データは、元の音響データに比べ、平均 1/5 程度にまで小さくなる特徴も持っていた。

我々は、インターネットを利用し 24 時間休みなく音響構造の配信を行う、電子聴覚の実現を目指した。そこでまず本稿では、ITM を用いリアルタイムに音響構造を 5 台のクライアントに配信するシステムを試作した。

2. ITM による音響構造データの生成と再構成法

図 1 に音響構造データの作成から構造再構成までの一連の流れを示す。原信号は 1 オクターブ長の帯域通過フィルタ (BPF: Band Pass Filter) を複数個使用したフィルタバンクに通される。フィルタを通過した波形は歪んだ正弦波のような形状をしている。この波形について、極大値から極小値、もしくは極小値から極大値まで

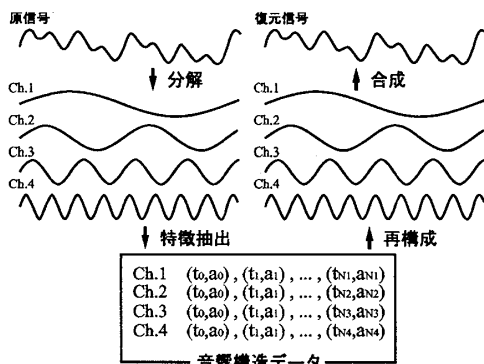


図 1: 音響構造データの生成と再構成の流れ

[†]北見工業大学大学院工学研究科情報システム工学専攻

^{††}北見工業大学工学部サテライト・ベンチャービジネス・ラボラトリー

の時間 t_n を計測し、極値間の波形を半周期分の正弦波と仮定することで周波数を算出する。この際、周波数は $1/2t_n$ で表現することができ、得られる時間分解能は算出された周波数相当成分の周期の 1/2 となる。こうして生成された音響構造データは、極値 a_n と極値間の時間 t_n を要素として持つベクトルの時系列で表現され、ベクトルの成分数はフィルタバンクに使用したフィルタ数 (チャンネル数) と等しくなる。

音響再構成もまた、極値間が正弦波の一部であると仮定して行う。まずは音響構造データの a_n と t_n から振幅変調と周波数変調をして正弦波を造りだす。そしてその一部を使用し極値間を補間し、補間した半周期分の正弦波を時間軸上につなぎ合わせていく。この処理を各チャンネルに対して行い、全てのチャンネルを合成することで原波形の再構成を行う。

この手法の有効性を調べるため原波形と再構成後の波形を比べたところ、その相互相関は 0.9931 ± 0.0041 であった。この際、音響構造の生成には 1MHz で 1 秒間サンプリングした音声 205 個、音楽 200 個の合計 405 個のデータを使用した。比較は 44.1kHz にダウンサンプリングした再生データで行った。

3. 音響構造配信システム

図 2 に試作したシステムを示す。音響構造配信システムは原信号を分割し必要な信号を取り出すフィルタ部と、極値間の計時や A/D 変換を行うサンプリング部、構造データを配信するサーバと受信をするクライアントの 4 つで構成した。ここでは、フィルタ部とサンプリング部からなる構造データ生成部と、サーバとクライアントからなる構造データ配信部について詳しく説明する。

3.1 構造データ生成部

本システムでは電話の周波数帯域を網羅するように、256Hz から 4096Hz の帯域で BPF を 4 チャンネル分用意した。図 3 に構造データ生成部で音響信号から特徴抽出を行う様子を示す。フィルタ部では入力された音響信号を BPF によって帯域ごとに分解し、さらにその 1 次微分波形から矩形波を造った。サンプリング部ではこの 2 つの波形を使用し、音響構造データを生成した。サンプリング部には 1.6MHz で動作する 24bits カウンタを用意した。カウンタの値を使用して矩形波のエッジ間の

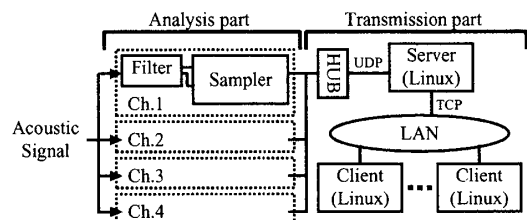


図 2: 音響構造配信システム

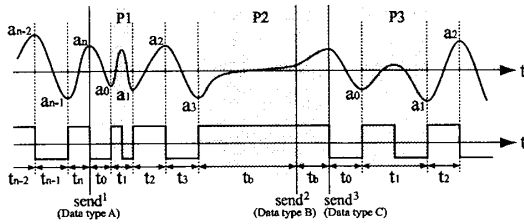


図 3: 構造データ生成部での処理

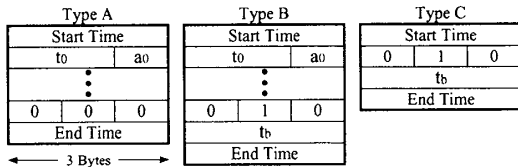


図 4: サーバに送信されるデータの型

Channel	Data Size	Start Time	Blank Time	End Time	Count Data	Amp Data
2bytes	3bytes	8bytes	4bytes	8bytes	2bytes * Data Size	1bytes * Data Size
		Count Data = {t ₀ , t ₁ , ..., t _n }		Amp Data = {a ₀ , a ₁ , ..., a _n }		

図 5: クライアントに配信されるデータの型

計時を行うと同時に、矩形波の変化に合わせて BPF を通過した波形の A/D 変換 (8bits 線形量子化) を行った。これにより、極値間の時間 t_n と極値 a_n を得た。ただし、P1 に示す様に極値間の時間が極めて短い場合、その極値は無視することにした。P2 の様に極値間の時間が極めて長い場合は無音であると考え、次の極値を待たず、途中で強制的にデータを送信することにした。P3 に示す様に振幅が極めて小さい場合、その極値は無視することにした。このようにして得られた音響構造データは図中 send のタイミングでサーバに送信され、そのタイミングによってサーバに送信するデータのフォーマットを変化させた。

データフォーマットを図 4 に示す。サーバに送信する各データには極値に関する情報の他に、ヘッダとフッタを加えた。ヘッダは各データの始まりの時間 (24bits カウンタの値) とし、フッタはデータフォーマットによって変化した。図 3 中 send² の様に強制的にデータが送信される場合は TypeB で、send³ の様に空白しか含まない場合は TypeC で送信する様にした。それ以外の場合は TypeA で送信した。TypeA のフッタは 000 と、各データの送信時間 (24bits カウンタの値、以下各データの終了時間) の 2 つとした。TypeB と TypeC の場合は TypeA と区別するための 010 と 24bits で表現された無音時間、各データの終了時間の 3 つとした。フッタに使用される 000 や 010 は受信側では $t_n = 0, a_n = 0$ 又は $t_n = 1, a_n = 0$ と同じデータとなるが、値の小さい t_n はサンプリング部で無視する様にしたので、極値情報と区別することができる。TypeA で送信されるデータにおける t_n や a_n の個数は最大 487 個とした。これは時間にして 60ms から 480ms ごとにデータが送信されていることになる。これを決定しているのは最大パケットサイズであり、本システムでは 1472bytes (IP パケットのサイズにして 1500bytes) である。伝送効率を良くするため、符号データはこのサイズまで蓄え、まとめて送信することにした。この際、周波数が高い信号ほど極値が多

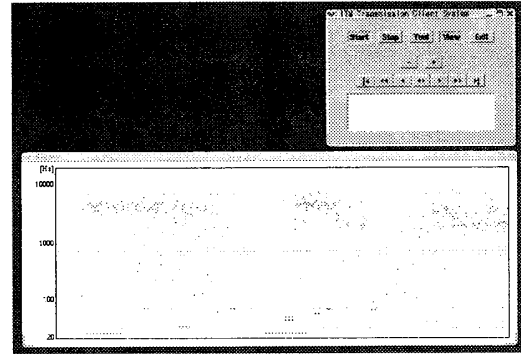


図 6: クライアントでデータを受信した様子

いため、高周波帯域のチャンネルほどデータの送信が頻繁に行われることになる。サーバとの通信は 10BASE-T で UDP を使用し、各チャンネルごとに非同期で行った。

3.2 構造データ配信部

サーバとクライアントはインターネットではなく学内ネットワークという比較的近距离で通信を行った。サーバにも内部でカウンタを持たせ、構造データ生成部から受け取った各データの開始時間と終了時間を 24bits から 64bits に拡張した。サーバがクライアントに配信するデータの型を図 5 に示す。サーバは構造データ生成部から受け取ったデータを図 5 の様に変換し、100BASE-TX で TCP を使用してクライアントに配信する様にした。クライアントでは各データが持つ開始時間を利用してチャンネル間の同期を取り、音響構造の再構成を行った。

図 6 にクライアントで音響構造データを表示した様子を示す。横軸に時間、縦軸に周波数を取り、500ms 分の周波数データを表示させた。

4. まとめ

我々は LAN を利用した音響構造配信システムの設計・構築を行った。ITM を使用することで、可変サンプリングによる音響データを得ることができた。本システムで得られる音響構造データは、元の音響信号の極値に関する情報のみであるため、データサイズは非常に小さいものとなった。今後は現在 4 チャンネルの構造データ生成部を、ヒトの可聴帯域を網羅するよう 20Hz から 20,000Hz の 10 チャンネルに拡張したい。

参考文献

- [1] J.W.Cooley, J.W.Tukey: "An algorithm for the machine computation of complex Fourier series," Math.Comput.19, pp.297-301, 1965.
- [2] N.Wiener: "The Fourier Integral and Certain of Its Application," Dover, NewYork, 1958.
- [3] 入場健仁, 吉田秀樹, 岡田信一郎, 藤原祥隆: "MPI を用いた周波数情報の定量システムの開発," 情報処理北海道シンポジウム 2003 講演論文集, p74, 2003.
- [4] 吉田秀樹, Xie Wei, 藤原祥隆: "感情推定の為の音響構造再構成法の試み," 信学技報 HIP2003-131, pp.29-33, 2004.