Windows対応 音声分析合成ソフトウェア

# 音声録聞見 for Windows (CD-ROMフリーウェア版)

# ユーザーマニュアル

2005年3月 第8版

はじめに

「音声録聞見 for Windows」は、東京大学 大学院 医学系研究科 認知・言語医学講座で開発された音声分析・合成ソフトウェアです。

桐谷 滋先生(現 神戸海星女子学院大学 文学部)と今川 博先生のご指導を頂き、本ソフトウェアの改良作業 に従事することができました。両先生に機会を与えて頂き、ご指導賜りましたことを厚くお礼申し上げます。

本改良作業を始めるに際し、技術的裏付けを確認し習得しておくことが重要でした。長年に渡る広島女子大学 国際文化学部 今石 元久、岩手大学 工学部 三輪 譲二ほか諸先生方のご厚情に感謝申し上げます。

急速に変化するコンピュータの世界にあって、新しい「音声録聞見 for Windows」が言語、医学を始めとする 様々な分野で少しでもお役に立つことを願って止みません。

Microsoft, Windows, Windows NT, Windows XP, Visual Basicは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。 Microsoft Excelは、マイクロソフト株式会社が発売するソフトウェアの名称です。 その他、本書に記載する社名、製品名は、一般にそれぞれの会社の商標または登録商標です。

## 目次

第1章 概要	1
第2章 音声録聞見 for Windowsの組み込み	2
2-1. 動作環境	2
2-2. マイクとスピーカの接続	2
2-3 録音/再生音量の調敷	3
2-3-1. 録音音量の調整手順	3
2-3-2. 再生音量の調整手順	4
2-4. 音声録聞見 for Windowsのインストール	6
2-5. 音声録聞見 for Windowsの削除	8
2-6. 起動	10
第3章 揭作説明	11
3-1. 百尸	11 11
3-1-2. メニュー	
3-1-3. ボタン等	13
3-1-4. 解説	13
3-2. サウンドスペクトログラム	14
3-2-1. 画面	14
3-2-2. メニュー	14
3-2-3. ボタン等	15
3-2-4. 音声波形上での操作	15
3-2-5. サウンドスペクトログラム上での操作	16
3-2-6. スペクトル上での操作	17
3-2-7. 解説	
3-3. ピッチ抽出	19
3-3-1. 画面	19
3-3-2. メニュー	19
3-3-3. ボタン等	20
3-3-4. 音声波形上での操作	20
3-3-5. ピッチカーブ上での操作	21
3-3-6. 解説	23
3-4. ホルマント分析	24
3-4-1. 画面	24
3-4-2. メニュー	24
3-4-3. ボタン等	25
3-4-4. 音声波形上での操作	25
3-4-5. 解析波形上での操作	26
3-4-6. 解説	
第4章 技術解説	
4-1. 用語説明	
4-9 サウンドスペクトログラム	20
4-2-1. Dynamic Measure 2017	

4-2-2. 分析のための予備知識	
第5章 主仕様	35
5-1. 共通仕様	
5-2. 音声波形編集	35
5-3. サウンドスペクトログラム	35
5-4. ピッチ抽出	35
5-5. ホルマント分析	
第6章 初期值	
6-1. 音声波形編集	
6-2. サウンドスペクトログラム	
6-3. ピッチ抽出	
6-4. ホルマント分析	
第7章 参考書·参考文献	
索引	

#### 第1章 概要

このたびは音声分析合成ソフトウェア「音声録聞見 for Windows」をご利用頂きまして誠にありがとうございます。

音声録聞見 for Windowsは、次の4つのプログラムで構成されています。

- ① 音声波形編集 音声の録音・再生、音声データファイルの読み込み・保存、音声の編集(削除、コピー、貼り付け、無音化、 加算、Fade-In、Fade-Out等)を行います。編集画面の印刷を行うこともできます。 また、この音声波形編集プログラムから、直接サウンドスペクトログラムおよびピッチ抽出プログラムを呼び 出し、編集した音声を渡してそのまま解析することもできます。
- ② サウンドスペクトログラム 音声データファイルを読み込み、サウンドスペクトログラム分析を行います。音声の再生、解析結果の印刷 を行うこともできます。
- ③ ピッチ抽出 音声データファイルを読み込み、ピッチ抽出を行います。音声の再生、解析結果のファイルへの保存、印 刷を行うこともできます。
- ④ ホルマント分析 音声データファイルを読み込み、ホルマント分析とピッチ抽出を行います。音声の再生、解析結果のファイ ルへの保存、印刷を行うこともできます。

また、おまけとして音声波形編集プログラムから「遅延エコー」というプログラムを呼び出すことができます。この プログラムは、音声を録音しながら、録音した音声を一定の遅延時間をおいて再生することができるプログラムで す。

#### 第2章 音声録聞見 for Windowsの組み込み

#### 2-1. 動作環境

音声録聞見 for Windowsを動作させるために必要な環境を説明します。

•OS	日本語Windows 95/98/Me 日本語Windows NT 4.0 (SP3以降) (*1), (*2) 日本語Windows 2000 (*2)
	日本語Windows XP Home Edition / Professional (*2), (*3)
・コンピュータ	Pentium 166MHz以上のCPUを搭載した機種を推奨
・メモリ	32MB以上推奨
・ハードディスク	音声録聞見 for Windowsは 10MB程度の容量を使用します
・ディスプレイ	256色以上が表示可能なディスプレイ
	解像度 1024×768以上を推奨、800×600で使用する場合は、「コントロールパネル」→「画
	面」→「設定」の「詳細」または、「ディスプレイの詳細」→「フォントサイズ」を「小さいフォント」 に設定する必要があります
・サウンドボード	Windows対応のサウンドボードを搭載していること
	録音するためにはマイク、試聴するにはスピーカまたはヘッドホン等が必要となります
(*1)	Windows NT 4.0の場合は、Service Pack 3以降が必要です。
(*2)	Windows NT/2000/XPの環境へのインストールには、Administrator(管理者)権限が必要で
	す。
(*3)	ユーザーの簡易切り替え機能には対応しておりません。

### 2-2. マイクとスピーカの接続

音声録聞見 for Windowsで音声の録音・再生をするためには、ご使用になるコンピュータにマイク、スピーカ(またはヘッドホン)を接続する必要があります。

ただし、コンピュータ(特にノートパソコン等)によっては、マイクやスピーカを内蔵した機種があります。この場合は、 これらの内蔵のマイクやスピーカをそのまま使用することができます。詳しくはご使用のコンピュータのマニュアルを ご参照下さい。



#### 2-3.録音/再生音量の調整

ご使用になるコンピュータにマイク、スピーカ(またはヘッドホン)を接続ができましたら、コンピュータの録音/再 生の音量を調整します。ここで説明する手順は、実際に音声録聞見 for Windowsを動作させながら行うと良いでしょう。

#### 2-3-1.録音音量の調整手順

- ① Windows付属の機能である「ボリュームコントロール」を使用して録音音量を調整します。 ここで、説明する画面は Windowsのバージョンやパソコンの機種、サウンドボードの種類や機能によって異なることがあります。
- ② ボリュームコントロールを起動するには、画面左下の「スタート」→「プログラム」→「アクセサリ」→「エンターテ イメント」または「マルチメディア」→「ボリュームコントロール」を選択します。 また、画面右下のタスクバー内のスピーカの形をしたアイコンをダブルクリックしても起動することができます。



③ ボリュームコントロールが起動したら、メニューの「オプション」→「プロパティ」を選択します。



④ 「プロパティ」ウィンドウが開きますので、「音量の調整」の「録音」をチェックし、「表示するコントロール」で録音 に使用する音源をチェックします。



⑤ チェックしたコントロールが表示されますので、使用する音源の「選択」をチェックし、「音量」のつまみを上下にスライドして録音音量を調整します。なお複数のコントロールの「選択」をチェックすると、チェックしたコントロールをミキシングして録音することができます。



⑥ 一度音量を設定すると、その設定は以降ずっと保持されますので、音量調整が終わったらボリュームコントロ ールのウィンドウの右上の ×ボタンを押して終了しても構いません。

#### 2-3-2. 再生音量の調整手順

- Windows付属の機能である「ボリュームコントロール」を使用して再生音量を調整します。
   ここで、説明する画面は Windowsのバージョンやパソコンの機種、サウンドボードの種類や機能によって異なることがあります。
- ② ボリュームコントロールを起動するには、画面左下の「スタート」→「プログラム」→「アクセサリ」→「エンターテ イメント」または「マルチメディア」→「ボリュームコントロール」を選択します。 また、画面右下のタスクバー内のスピーカの形をしたアイコンをダブルクリックしても起動することができます。



③ ボリュームコントロールが起動したら、メニューの「オプション」→「プロパティ」を選択します。



④ 「プロパティ」ウィンドウが開きますので、「音量の調整」の「再生」をチェックし、「表示するコントロール」で最低 限でも「再生コントロール」と「Wave」をチェックします。

プロパティ		? ×
ミキサー デバイス( <u>M</u> ): <mark>SB Audic</mark>	PCI ミキサー	<b>•</b>
一音重の調整		
<ul> <li>再生</li> </ul>		
○録音		
0 その他		7
表示するコントロール:		
図 再生 コントロール		<b>_</b>
I Wave		
■ シンセ □ フィク		
•		
	OK	キャンセル

⑤ チェックしたコントロールが表示されますので、それぞれの「音量」のつまみを上下にスライドして再生音量を 調整します。

なお「再生コントロール」は、チェックしたコントロールの全てをミキシングした後の音量の調節を行います。 また「ミュート」をチェックすると、チェックしたコントロールの音を消す事ができます。



⑥ 一度音量を設定すると、その設定は以降ずっと保持されますので、音量調整が終わったらボリュームコントロ ールのウィンドウの右上の ×ボタンを押して終了しても構いません。

### 2-4. 音声録聞見 for Windowsのインストール

ここでは、音声録聞見 for Windowsをコンピュータにインストールする手順について説明します。本書では Windows 98の画面で説明していますが、Windows 95/Me、Windows NT/2000/XPでもほぼ同じ手順となります。

- ① 使用中のアプリケーションがある場合は全て終了します。さらに、インストール中のトラブルを避けるために Windowsを再起動します。
- ② Windowsが再起動したら、CD-ROMを CD-ROMドライブに挿入します。次に、エクスプローラで CD-ROM 内の音声録聞見 for Windowsのセットアッププログラムが格納されているフォルダを開き、その中の setup.exeをダブルクリックします。
- ③ セットアップが開始して、下図の画面が表示されます。内容を確認して、次へ ボタンを押して下さい。 なお、この画面を含め、キャンセル ボタンをクリックするといつでもセットアップを中止することができます。



④ インストール先を指定する画面が表示されます。表示されているフォルダ(ディレクトリ)で良ければ 次へ ボ タンを押して先に進んで下さい。変更したい場合には、 参照 ボタンを押して希望のフォルダに変更して下 さい。



⑤ スタートメニューに登録するプログラムフォルダを指定するウィンドウが表示されます。
 通常はそのまま 次へ ボタンを押して下さい。



⑥ 今まで設定した内容の一覧が表示されます。この設定で良ければ 次へ ボタンを押します。 変更したい場合は、 戻る ボタンで前の画面へ戻ることができます。



⑦ インストールが開始します。インストールの進行状況が表示されます。



⑧ インストールが完了すると、下記の画面が表示されます。場合によっては、ここで Windowsを再起動する画面が表示される場合もあります。その場合には、画面の指示に従って Windowsを再起動して下さい。

セットアップの完了	
	コンビューリ上へのアプリケーションのインストールを完了しました。
	イバスト-&されたアイコンを選択するとアプリケーフォンを起動できます。
Installibrie(d)	[党了]&'97を炒っりすると、セッド7っ7'を終了します。
	< 戻る(II) <b>完了</b>

以上で、音声録聞見 for Windowsのインストールは完了です。

### 2-5. 音声録聞見 for Windowsの削除

音声録聞見 for Windowsが不要になった場合のコンピュータからの削除(アンインストール)方法を説明します。

 画面左下の「スタート」をクリックしてスタートメニューを開き、「設定」を選択しその中の「コントロールパネル」 をクリックします。



② この中の「アプリケーションの追加と削除」のアイコンをダブルクリックします。



③「アプリケーションの追加と削除のプロパティ」ウィンドウが開き、「インストールと削除」タブが表示されます。 このウィンドウの下の一覧の中から「音声録聞見 for Windows(フリー版)」を選択し、「追加と削除」ボタンを 押します。

アブリケーショ	ョンの追加と削除のプロパティ	×
インストー	ルと削除   Windows ファイル   起動ディスク	
-	フロッピー ディスクまたは CD-ROM から新しいアプリケーション をインストールするには、【インストール】をクリックしてください。	
	インストールロ	ן נ
3	次のソフトウエアは目動的に背側をできます。アプリケーションを 脊側体したり、その側成ファイルを変更するには、一覧から進ん で【追加と脊側を】をりリックしてくだとしく <u>」</u> )	-
Vietnar Visio T Window Window Window Window WinZip マウスウェ 日同し工	rese Language Support controls 30-7 vs 98 G840015 のアップテート vs 96 G890175 のアップテート vs Modia Player 9 シリーズ (システム更新) vs G823559 のアップテート 7 万 満 品 (or Wordstoor(タ))=15) ・	]
	<u> 這力如子們引除(E)</u>	ו
	OK キャンセル 湾用(6)	

④ 確認のメッセージが表示されますので、良ければ はい ボタンを押します。



⑤ 削除が完了すると、「アンインストールが完了しました。」と表示されます。もし、途中で「共有ファイルを削除しますか?」とのウィンドウが表示された場合には、「すべていいえ」を押して下さい。

コンビュータからフロクラムを肯領余	×	共有7ヶイルを削除しますか?	×
	uninstal I Shield Lt、コンピューカからフ*ログ 54、管理時間 見 for #indows (フン-KB)、を利用します。次の名 コンポーネリンが明瞭される間、しけらくお待ちください ・ 共有フ*ログ 54 ファ(4 ・ オ#ジ*切目 ・ フ*ログ 54 フ*(40)り ・ フ*ログ 54 ブ*(40)り ・ フ*ログ 54 芽*(40)り ・ フ*ログ 54 黄*(40)り	9.75人上で、次の共有ファイルさどのプログラムでも使用されていません。 他のブロン゙3ムハビcのファイルを使用する場合。例除をするとプロン゙3ムがほ なる恐れがあります。共有ファイルを利除してもよいですか? このファイルを残して置いても9.27ムを壊すことはありません。あまり値 い場合は、この共有10Å、*3½を得ましないようにして下さい。 ファイル名:	hかなく 実でな E(Q)

以上で、音声録聞見 for Windowsの削除は終了です。

画面左下の「スタート」をクリックして開くスタートメニューの中から、「プログラム」を選択し、その中の「音声録聞見 for Windows (フリー版)」を選択します。

「サウンドスペクトログラム」・「ピッチ抽出」・「ホルマント分析」・「音声波形編集」の中から起動したいプログラムをク リックします。



#### 第3章 操作説明

本章では、音声録聞見 for Windowsの各プログラムの操作方法を説明します。

#### 3-1. 音声波形編集

音声の録音・再生、音声データファイルの読み込み・保存、音声の編集(削除、コピー、貼り付け、無音化、加算、 Fade-In、Fade-Out等)を行います。編集画面の印刷を行うこともできます。

また、この音声波形編集プログラムから、直接サウンドスペクトログラムおよびピッチ抽出プログラムを呼び出し、 編集した音声を渡してそのまま解析することもできます。

なお、音声波形編集で取り扱える音声データ長は、最大で 24秒です。

#### 3-1-1. 画面



Tail Cursor Part波形表示部

: Tailカーソル近辺の時間軸拡大波形

#### 3-1-2. メニュー

File

Open...

音声データファイル (.wav)を読み込みます。

読み込む音声データファイルがステレオ音声の場合、左右どちらのチャネルの音声を読 み込むか指定するウィンドウが表示されますので、読み込むチャネルを指定して下さい。 また、24秒以上の音声データファイルの場合は、音声データを 24秒に切り詰めるメッセ ージが表示されます。

Channel Select
この音声データファイルはステレオです。 読み込む左右チャネルを選択して下さい。
Left CH. Right CH.
ファイル読込
この音声データファイルは 24秒以上の長さです。 音声データを 24秒に切り詰めます。
ОК

Play

音声データファイルがステレオ音声の時は、 読み込むチャネル(左右)を指定するウィンド ウが表示されます。

24秒以上の音声データファイルの場合は、 音声データを 24秒に切り詰めるメッセージ が表示されます。

Save As Print	音声波形上の Head-Tail区間の音声データをファイルに保存します。 画面を印刷します。 プログライナ ゆるします
Exit	ノロクラムを終了します。
Sound	
Play	Head-Tail区間の音声を再生します。
Record	音声を録音します。Settingsで設定した 12秒または 24秒で自動的に停止します。
Pause	再生を一時停止します。
Stop	再生・録音を終了します。
Settings	再生・録音の条件を設定します。

ettings	
Play	
	Normal Repeat
Do Repeat	C Seamless Repeat
Record	
Sampling Rate	11025 <b>T</b> Hz
Wave Length	€ 12sec € 24sec
Safety	🗖 Ctrl Key
OK	Cancel

Do Repeat	チェック時は、音声再生を繰り返すモード
	になります。
Normal Repeat	0.8秒程度の間隔で繰り返し再生。
Seamless Repeat	間隔 0で繰り返し再生。
Record	
Sampling Rate	録音周波数を 8/10/11.025/12/22.05
	/24/44.1/48kHzの中から選択します。
Wave Length	最長録音時間を 12秒または 24秒にしま
	す。
Safety	チェック時は、 Ctrl キーを押しながらで
	ないと録音が開始できません。

#### Edit

Cut	Head-Tail区間の音声を削除します。削除した音声は Paste Bufferにコピーされます。
Сору	Head-Tail区間の音声を Paste Bufferにコピーします。
Paste(H)	Paste Bufferの音声を Headカーソルの前に貼り付けます。
Paste(T)	Paste Bufferの音声を Tailカーソルの後ろに貼り付けます。
Silence	Head-Tail区間の音声を無音にします。
Add	Head-Tail区間の音声に Paste Bufferの音声を加算します。
Sub	Head-Tail区間の音声に Paste Bufferの音声の位相を反転し加算します。
Fade In	Head-Tail区間で Fade-in。
Fade Out	Head-Tail区間で Fade-out。
Sine Wave	Head-Tail区間に正弦波を生成します。
Noise	Head-Tail区間にノイズ(一様乱数)を生成します。
Impulse	Head-Tail区間にインパルス列を生成します。
Amp. Increase	Head-Tail区間の音声の振幅を増大します。
Amp. Decrease	Head-Tail区間の音声の振幅を減少します。
Amp. Normalize	全区間の音声の振幅の正規化(最大振幅値(絶対値)を指定振幅に合わせる。)します。
Waveform	
<b>T</b> ' O I	

Time Scale

x1/4	全体波形表示部の表示時間幅を 24秒にします。
x1/2	全体波形表示部の表示時間幅を 12秒にします。
x1	全体波形表示部の表示時間幅を 6秒にします。
×2	全体波形表示部の表示時間幅を 3秒にします。
x4	全体波形表示部の表示時間幅を 1.5秒にします。
Send to	
Pitch	ピッチ抽出を起動し、Head-Tail区間の音声データを送ります。
Sound Spectrogram	サウンドスペクトログラムを起動し、Head-Tail区間の音声データを送ります。
Omake	
D.A.F.	遅延エコー (Delayed Auditory Feedback)を起動します。
Help	
Version	本プログラムのバージョン情報を表示します。

#### 3-1-3. ボタン等

画面上のボタン	メニューと同じ名前を持つボタンは対応するメニューを同じ機能を持ちます。
Тор	Headカーソルを最初に置きます。(Headカーソルが最初にあるときは水色になります。)
End	Tailカーソルを最後に置きます。(Tailカーソルが最後にあるときは水色になります。)
スクロールバー	音声データの任意の部分に移動して、全体波形表示部に表示します。
$\langle \rangle$	Head/Tail Cursor Part波形表示部の波形をスクロールします。
Head/Tail Part	Head/Tail Cursor Part波形表示部の表示時間幅を 10~500msecの間で指定します。

#### 3-1-4. 解説

- Head、Tailカーソルの指定方法 各波形表示部において、マウス左ボタンクリックで Headカーソルを、マウス右ボタンクリックで Tailカーソル を指定します。
- ② サンプリング周波数について サポートするサンプリング周波数は、8/10/11.025/12/22.05/24/44.1/48kHzの 8種類です。
- ③ ステレオ音声について ステレオ音声データファイルを読み込む時は、左右どちらかのチャネルを指定して読み込みます。
- ④ 音声録音・再生プログラム 音声波形編集および、遅延エコーでの音声再生・録音は、Kentaro Hara氏作成の WAVEオーディオデバイ スインターフェースコントロール Version 2.0を使用しています。 Copyright (C) 1998-1999, Kentaro Hara. All rights reserved. http://www32.ocn.ne.jp/~kentaro/index.html
- ⑤ Delayed Auditory Feedback (遅延エコー)に関して

👪 Del	ayed Aux	litory Feedback			_ = ×
Elle	Help				
		遅延	<u> </u> エコー	-	
		Start		Stop	
		– Delay: abo	ut 150	msec —	
S	nall	De	ay Time		Large
					•

音声を録音しながら再生するプログラムです。録音-再生の遅 延時間は、約 100msecから約 1秒の範囲で指定できます。た だし、最小遅延時間は使用コンピュータにより異なります。

### 3-2. サウンドスペクトログラム

音声データファイルを読み込み、サウンドスペクトログラム分析を行います。音声の再生、解析結果の印刷を行う こともできます。

#### 3-2-1. 画面



#### 3-2-2. メニュー

File	
Open	音声データファイル(.wav)を読み込みます。 読み込む音声データファイルがステレオ音声の場合、左右どちらのチャネルの音 声を読み込むか指定するウィンドウが表示されますので、読み込むチャネルを指 定して下さい。
Print	結果を印刷します。
Exit	プログラムを終了します。
Sound	
Play	音声波形上の Head-Tailカーソル区間の音声を出力します。 読み込んだ音声データファイルがステレオ音声だった場合、実際に読み込んだチャネルに係わらず、ステレオ音声が再生されます。
Waveform	
All	全体の波形を再表示します。
Sound Spectrogram	
Execute	指定の条件で音声波形上の Head-Tail区間の音声のサウンドスペクトログラムを 表示します。
Reset	分析条件を全て初期状態に戻します。
Option	

Dynamic Measure	チェック時は、WLR距離尺度によるスペクトル変化量を表示します。 WLR: Weighted Likelihood Ratio (杉山・鹿野)
Help Version	本プログラムのバージョン情報を表示します。
3-2-3. ボタン等	
画面上のボタン Sound Spectrogram	メニューと同じ名前を持つボタンは対応するメニューを同じ機能を持ちます。
Pre-Emphasis	チェック時は、Pre-Emphasisをかけます。(6dB/Oct)
High–Boost	チェック時は、24dB / 10kHzの傾斜で高域を強調します。
Frame Shift [msec]	分析フレームシフト幅。
Auto frame Shift	チェック時は、サウンドスペクトログラムを画面いっぱいに表示するように自動的に
	Frame Shiftを決定します。
Window Length [msec]	分析窓長。 Narrow ボタンをクリックすると 60が入り、 Wide ボタンをクリックする
	と 6.6が入ります。
	また、指定した値が FFT窓長より長い場合は、警告メッセージが表示され、FFT
	窓長に補止されます。
	FFT窓長を 2048, 1024, 512, 256の中から選択します。
Window Type	時間窓の形を Hamming, Hanning, Blackman, Rectangularの中から選択します。
Freqency [Hz]	サリンドスヘットロクフム表示向波叙範囲。
Auto Detect	アエック時は、衣示取入周波級と取小周波級を音声アータのサンフリンク周波級 かど自動的に決定します。
Linnor	296日期内に伏足しより。 主一是十国油粉
Upper	衣小取八向 <b>似</b> 剱。 表示最小周波粉
Image	式/\`取/1·/H/X 数。
Blackスクロールバー	サウンドスペクトログラム画像の里レベル設定
Whiteスクロールバー	サウンドスペクトログラム画像の白レベル設定。
γ-	チェック時は、サウンドスペクトログラム画像をガンマー補正します。中間濃度値が
	少し暗いときに使用します。
γ+	チェック時は、サウンドスペクトログラム画像をガンマ+補正します。中間濃度値が
·	少し明るいときに使用します。

#### 3-2-4. 音声波形上での操作

マウス左ボタンクリックで Headカーソルを、マウス右ボタンクリックで Tailカーソルを指定します。

Sound Spectro Sound	Waveform	Sound Spectrogram Distion Help				
-	• •	* ***		<b>p</b>		Image Black 240
5 k	2		1 1			=
4.8	11	3 44 3, 44 4	1 1000			
3 k	11	1 N A	1 180			<u> </u>
2 k	11	1 Dime III	1 193	· · · · ·		White
1.k	88	K PR. MAR	A MEN	8		
900		← 4179	16 and			417996 sec
Divide		Sound Spectrogram			Cursor Inform	tion
noling Rate	1025 Hz	Execute Reset	C 2048	C Hamming		Head Tail
ound	Wave form	P Pre-Emphasis IT High-Boost	© 1024 © 512	Hanning Elackman	Sample No	
6 K		Frame Shift (meec)	C 256	C Rectang	Time [sec]	
Play	Al	I wuto rrame shart p	Frequency (Ht	1	Freq [Hz]	

#### 3-2-5. サウンドスペクトログラム上での操作

 マウスの左ボタンをクリックすると、マウスをクリックした位置の時刻(音声データファイルのサンプル番号とファ イル先頭からの時間)と周波数を表示します。



② マウス右ボタンのクリックまたはドラッグにより、ポップアップメニューが表示されます。

#### ポップアップメニュー [Head and Tail]

Headと Tailカーソルを指定します。Headと Tailカーソルの時刻と Headから Tailまでの時間を表示します。



ポップアップメニュー [H and T (both Audio)] Headと Tailカーソルを指定します。音声波形の Headと Tailカーソルも同時に変更されます。



ポップアップメニュー [Spectrum (at Head)]

Headカーソル時点のスペクトルを表示します。上部には FFT窓内の音声波形と時間窓形状を表示します。



**ポップアップメニュー** [ Cancel ] ポップアップメニューの終了。

#### 3-2-6. スペクトル上での操作

マウス右ボタンクリックによりポップアップメニューが表示されます。



ポップアップメニュー [Next Frame] 次フレームのスペクトルを表示します。

**ポップアップメニュー** [Prev. Frame ] 前フレームのスペクトルを表示します。

#### ポップアップメニュー [ReCalculation]

現在のスペクトル表示の分析条件を変更し、スペクトルを再計算して表示しなおします。

変更できる分析条件は、Pre-Emphasisのオン/オフ、High-Boostのオン/オフ、分析窓長、FFT窓長、 Window Typeです。

変更したい分析条件を変更後、 OK ボタンを押すと、スペクトルを再計算して表示します。

なお、分析窓長 (Window Length)に指定した値が FFT窓長より長い場合は、警告メッセージが表示され、 FFT窓長に補正されます。

Recalculation				
Pre-Emphasis T High-Boost				
Window Length [msec]				
FFT	-Window Type			
C 2048	C Hamming			
1024	Hanning			
C 512	C Blackman			
C 256	C Rectang			
ОК	Cancel			

下記の例は、スペクトルの分析窓長を Wide (6.6msec)で表示後、Narrow (60msec)に変更して再計算し表示 したものです。



ポップアップメニュー [Quit] スペクトル表示の終了。

#### 3-2-7. 解説

- ① 本プログラムでは、48, 24, 12, 10, 8, 44.1, 22.05, 11.025kHzのサンプリング周波数のモノラルおよび、ステレオ音声データファイルを読み込むことができます。
- ② 本プログラムの解析手法等の詳細に関しては、巻末の技術解説をご参照下さい。
- ③ サウンドスペクトログラム画像について サウンドスペクトログラム表示窓の垂直方向の画素数が、画面サイズの制約上間引いた状態で表示されます。 プログラム処理過程でオリジナルのサウンドスペクトログラム画像を、アプリケーションをインストールしたフォ ルダに「Sonagraph.bmp」という名前で作成します。このビットマップファイルの垂直方向の画素数は 512となっています。このため、場合によっては「Sonagraph.bmp」も有用である場合があります。

### 3-3. ピッチ抽出

音声データファイルを読み込み、ピッチ抽出を行います。音声の再生、解析結果のファイルへの保存、印刷を行うこともできます。

ピッチ抽出方法:「柔軟性のあるピッチ抽出法」川浦、粕谷 宇都宮大学工学部 電子情報通信学会技術研究報告 SP87-3 1987.4.23

#### 3-3-1. 画面

ina Pito	sh Extract	ion								_ 🗆 X
<u>F</u> ile	<u>S</u> ound	<u>W</u> aveform	Pitch <u>E</u> xtractio	n <u>H</u> elp						
	音声	<b>沙</b> 波形								Wave Head D Tail B1999 0.3 Hz
	400 300 200 100 ピッラ	F-J-J	<u> </u>		Manana P		 			
0. s	ec		( <u>\</u> , <u>\</u>	[	← 3.1999	sec →		Ά.	3.1999	sec
Mem	D:									
File:	C:¥Void	ce Data¥Englisł	n_01		Pitch Extractio	)n	Cu	irsor Inform	ation	T - 11
Samp ⊢Sou	oling Rate: nd	1 0000 Hz Waveform			Execute	Frame Shift [msec] 5	S	ample No	Head 1 0228	16232
		Zoom			Pacat	Silent Level  30	оо ті	ime [sec]	1.0228	1.6232
						F0 Maximum [Hz] 40	00 FC	) [Hz]	213	263
	nay	Draw Pov	wer			FO Minimum [Hz] 60	D Po	ower (dB)	0.2	3.41
		☑ Log Scal	e ngth Fix [sec] 🛛	3.0	Voiced	1.0	d	eriod [sec]	<mark>0.60</mark>	04

解析波形の説明

赤の点 : ピッチ周波数 [Hz] 水色の線 : 音声パワー [dB] (0 [dB]は Silent Levelを示します。)

#### 3-3-2. メニュー

File

Open	音声データファイル(.wav)を読み込みます。
	読み込む音声データファイルがステレオ音声の場合、左右どちらのチャネルの音声を読
	み込むか指定するウィンドウが表示されますので、読み込むチャネルを指定して下さい。
Save As	指定区間の音声パワーとピッチ周波数データを CSVファイル形式で保存します。
Print	結果を印刷します。
Exit	プログラムを終了します。
Sound	
Play	音声波形上の Head-Tailカーソル区間の音声を出力します。
	読み込んだ音声データファイルがステレオ音声だった場合、実際に読み込んだチャネル
	に係わらず、ステレオ音声が再生されます。

Waveform

Zoom All	音声波形上の Head-Tailカーソル区間を拡大表示します。 全体の波形を再表示します
Pitch Extraction	
Execute	指定の条件でピッチ抽出を実行し、結果を表示します。
Reset	分析条件を全て初期状態に戻します。
Help	
Version	本プログラムのバージョン情報を表示します。

3-3-3. ボタン等

画面上のボタン	メニューと同じ名前を持つボタンは対応するメニューを同じ機能を持ちます。
Waveform	
Draw Power	チェック時は、音声パワーカーブをピッチカーブと同じ画面に表示します。
Log Scale	チェック時は、ピッチカーブをログスケールで表示します。
	なお、ログスケール時はピッチカーブ表示範囲は F0 Minimum [Hz]~F0 Maximun [Hz]
	となり、リニアスケール時(チェック無時)は 0 [Hz]~F0 Maximun [Hz]となります。
Zoom Length Fix	チェック時は、拡大表示時に画面表示を指定時間に固定します。
Pitch Extraction	
Frame Shift [msec]	分析フレームシフト幅。通常は 5msec程度を設定します。
Silent Level	本プログラムが無音と判断する閾値。この値はフレーム内平均振幅です。通常 300程度
	を設定します。
F0 Maximum [Hz]	ピッチ探索最大値。
F0 Minimum [Hz]	ピッチ探索最小値。
Voiced / Unvoiced	このスクロールバーは、自己相関関数のピーク検出時、有声/無声の判定に用いる閾
	値を調整します。既定値は 1.0です。
	Voiced側にすると判定が有声傾向になり、UnVoiced側にすると判定が無声傾向になりま
	す。上記 Silent Levelとこのスクロールバーを調整しながらより正確な抽出を行います。

#### 3-3-4. 音声波形上での操作

マウス左ボタンクリックで Headカーソルを、マウス右ボタンクリックで Tailカーソルを指定します。

Nich Extraction Sound Waveform Pitch Extraction	Rep	- Wave
+ + * *		Fight           Tal           E324           Fight           Fight
150 150 150	- 31000 MHC -	31999 sec
ID Wholes Data¥English,Ot reling Rate 10000 Hz und Waveform	Pitch Estraction Execute Prame Shift (meec) 5 Shirt Level 500	Cursor Information Head Tail Sample No Tail

 一 音声波形上で、マウスの左ボタンをクリックした位置(Headカーソル)のサンプル 番号が表示されます。

 音声波形上で、マウスの右ボタンをクリッ クした位置(Tailカーソル)のサンプル番 号が表示されます。

#### 3-3-5. ピッチカーブ上での操作

 マウスの左ボタンをクリックすると、マウスをクリックした位置の時刻(音声データファイルのサンプル番号とファ イル先頭からの時間)とピッチ周波数、音声パワーを表示します。



音声パワーの 0 [dB]は Silent Levelを示します。

② マウスの右ボタンのクリックまたはドラッグにより、ポップアップメニューが表示されます。

#### ポップアップメニュー [Head and Tail]

Headと Tailカーソルを指定します。Headと Tailカーソルの時刻、ピッチ周波数と Headから Tailまでの時間 を表示します。



#### ポップアップメニュー [ UnVoice ]

マウス右ボタンのドラッグで指定した区間のピッチ周波数を無声にします。

#### ポップアップメニュー [ x2 ]

マウス右ボタンのドラッグで指定した区間のピッチ周波数を 2倍の値に修正します。

#### ポップアップメニュー [x1/2]

マウス右ボタンのドラッグで指定した区間のピッチ周波数を半分の値に修正します。

#### ポップアップメニュー [Interpolate]

マウス右ボタンのドラッグで指定した区間の始点と終点のピッチ周波数を用いて、その間を内挿します。



#### ポップアップメニュー [ Change a F0 Value ]

クリックした時点のピッチ周波数を数値で変更します。

(音声波形の Headと Tailを 1周期に合わせるとその周波数が表示されるので参考にして下さい。)



······		ssociaespectacitocito	p=p=p=p=p=p=p=======	10000 1000000
80 30 38	****	Head and Tail Linvoice × 2 × 1/2	••••••	
ORCC SHE Mercol Pare EVAlues Grant English (Of Exercise State Sound Weinefforms Pare Power Pare Discourses	+ 02519.9	Change a (D Volue. <u>Cancel</u> Frame Shift (mec) β Silent Level 500 F0 Maximum (ht) 400 F0 Minimum (ht) 60	Cursor Information Head Sample No 17903 Tane (sec.) (2,7913 FO (Hz) (7 Power (att)) (7,84	p 204



New F0 Value New F0 Value: 172 Hz

**ポップアップメニュー**[Cancel] ポップアップメニューの終了。

#### 3-3-6. 解説

- ① 波形拡大手順
  - (1) 音声波形上にマウスカーソルを置き、左ボタンで開始時点(Headカーソル)を、右ボタンで終了時点(Tail カーソル)を指定します
  - (2) Zoomボタンをクリックすると波形が拡大されます。



② 本プログラムで読める音声データ (.wav)ファイル

下表のサンプリング周波数のモノラルまたは、ステレオ音声データファイルを読み込むことができます。 ステレオの場合は、読み込み時に読み込むチャネル(右または左)を指定します。

サンプリング周波数 (kHz)	48	24	12	10	8	44.1	22.05	11.025
Main分析窓長 (msec)	30	30	30	36	35	32.6	32.6	32.6
全分析窓長 (msec)	45	45	45	54	52.5	48.9	48.9	48.9

 
 ③ 解析結果の保存ファイル例(CSVファイル形式) 保存した解析結果を Microsoft Excelに読み込んだときの例を示します。

🔀 Mi	crosoft Exce	l – test.csv		_ [	×	
副コ  ッー)   D @	ファイル(E) 編 WI) データ(I 全日 ④ C	集(E)表示() 2) ウィンドウ() 1 <b>4 8 2</b>	⊻ 挿入① № へいづ(H) 【】 <mark>そ</mark> 10	售式(Q)  0%、▼	X × ×	
MS	Pゴシック	• 18 ·		כעם פין אייע ב	카 🎇	
	F3		= /			
	A /	в	∕c	D		
1	10125	1.0125	-1.8 <del>6612</del>	> 0		
2	10175	1.0175	-2.08419	$\sim$		
3	10225	1.0225	0.200342	213		
4	10275	1.0275	10.15712	222		L
5	10325	1.0325	15.2885	222		
6	10375	1.0375	16.67144	217		
7	10425	1.0425	17.55127	213		
8	10475	1.0475	17.4819	213		
9	10525	1.0525	17.4819	208	_	
10	10575	1.0575	17.69213	208		
11	10625	1.0625	17.81215	208	_	
12	10675	1.0675	17.83447	208		
13	10725	1.0725	18.0907	208	. <b>-</b>	
	▶ ▶ ∖ <u>test</u>	/			Ш	
図形	の調整( <u>R</u> )・	▶ ७ オー	ヽシェイプ( <u>U</u> )・	• 🔨 🔪 🗐	*	
				NUM	_ //	

音声データファイル内のサンプル番号

- 音声データファイル内のファイル先頭からの時間 [sec]

フレーム内平均音声パワー [dB]

― ピッチ周波数 [Hz] (0は無音、7は無声を示します。)

#### 3-4. ホルマント分析

音声データファイルを読み込み、ホルマント分析とピッチ抽出を行います。音声の再生、解析結果のファイルへの 保存、印刷を行うこともできます。

ホルマント分析方法	:	LPCスペクトル包絡のピークピッキング。	
		バンド幅は 2次曲線あてはめにより求めます	0
ピッチ抽出方法	:	「柔軟性のあるピッチ抽出法」川浦、粕谷	宇都宮大学工学部
		電子情報通信学会技術研究報告 SP87-3	1987.4.23

#### 3-4-1. 画面



#### 解析波形の説明

青の点	:	ホルマント周波数 [Hz]
		無音区間または、ホルマントバンド幅が 900Hz以上の場合には点は非表示となります。
赤の点	:	ピッチ周波数 [Hz]
水色の線	:	音源パワー(残差エネルギー) [dB] (0 [dB]は Silent Levelを示します。)

### 3-4-2. メニュー

#### File

と読み込みます。
ルがステレオ音声の場合、左右どちらのチャネルの音声を読
が表示されますので、読み込むチャネルを指定して下さい。
チ周波数データを CSVファイル形式で保存します。

音声波形上の Head-Tailカーソル区間の音声を出力します。 読み込んだ音声データファイルがステレオ音声だった場合、実際に読み込んだ チャネルに係わらず、ステレオ音声が再生されます。
音声波形上の Head-Tailカーソル区間を拡大表示します。
全体の波形を再表示します。
指定の条件でホルマント分析およびピッチ抽出を実行し、結果を表示します。
分析条件を全て初期状態に戻します。
本プログラムのバージョン情報を表示します。

#### 3-4-3. ボタン等

画面上のボタン	メニューと同じ名前を持つボタンは対応するメニューを同じ機能を持ちます。
Waveform	
Zoom Length Fix	チェック時は、拡大表示時に画面表示を指定時間に固定します。
Formant and Pitch Extraction	
ピッチ抽出関連	
Frame Shift [msec]	分析フレームシフト幅。通常は 5msec程度。
Silent Level	本プログラムが無音と判断する閾値。この値はフレーム内平均振幅です。通常
	300程度を設定します。
F0 Maximum [Hz]	ピッチ探索最大値。
F0 Minimum [Hz]	ピッチ探索最小値。
Voiced / UnVoiced	このスクロールバーは、自己相関関数のピーク検出時、有声/無声の判定に用
	いる閾値を調整します。既定値は 1.0です。
	Voiced側にすると判定が有声傾向になり、UnVoiced側にすると判定が無声傾向
	になります。上記 Silent Levelとこのスクロールバーを調整しながらより正確な抽
	出を行います。
ホルマント抽出関連	
F. Window [msec]	分析窓長。通常は 20~30msec。
Peak Enhance	チェック時は、ピーク強調処理を行います。
	隣り合うホルマントが近い時などにホルマント抽出が困難になりますが、このよう
	な時はピークを強調すると良い結果が得られることがあります。
	(Off-axis spectral enhancement procedure : Markel 1976)



3-4-4. 音声波形上での操作

マウス左ボタンクリックで Headカーソルを、マウス 右ボタンクリックで Tailカーソルを指定します。

#### 3-4-5. 解析波形上での操作

マウスの左ボタンをクリックすると、その位置の各データ値を表示します。



#### 3-4-6. 解説

- LPC分析次数・FFT窓長について サンプリング周波数が 12kHzより高いとき、LPC分析次数 22, FFT窓長 1024。 サンプリング周波数が 12kHz以下のとき、LPC分析次数 12, FFT窓長 512。
- ② 波形拡大手順
  - (1) 音声波形上にマウスを置き、左ボタンで開始時点(Headカーソル)を、右ボタンで終了時点(Tailカーソル) を指定します。
  - (2) Zoomボタンをクリックすると、音声波形とホルマント・ピッチ曲線が拡大されます。



③ 本プログラムで読める音声データ (.wav)ファイル

下表のサンプリング周波数のモノラルまたは、ステレオ音声データファイルを読み込むことができます。 ステレオの場合は、読み込み時に読み込むチャネル(右または左)を指定します。

サンプリング周波数 (kHz)	48	24	12	10	8	44.1	22.05	11.025
Main分析窓長 (msec)	30	30	30	36	35	32.6	32.6	32.6
全分析窓長 (msec)	45	45	45	54	52.5	48.9	48.9	48.9

上記分析窓長はピッチ抽出窓長です。

④ 解析結果の保存ファイル例(CSVファイル形式)保存した解析結果を Microsoft Excelに読み込んだときの例を示します。

M	icrosoft E	xcel – test	.csv																				_ [/	□×
	ファイル(E)	編集(E) :	表示(⊻	)挿	入① 書	:式( <u>0</u> )	ツール(エ	) デー	タ( <u>D</u> ) ウ	ィンドウ	κ <u>ω</u> ) Λ.	ルプ( <u>H</u> )										Ī	_ /	al XI
	≆ 🖬 🔒	l 🗐 🗟 🖤		ю. •	14	<b>Σ</b> j	f <sub>≈</sub> A∣Z	l 🛍	<i>🚯</i> 100	% 🗸	2	: 🔟	a 🖓	I abl		۰	=ŧ		≓		* *	A		8
MS	Pゴシック	•	10 👻	B	ΙU	r 🗐	= =			<b>*</b> - <b>}</b> •-	¥₽	⊒• ¥	' 😨 🤊	6,	.0 .0	.0	€≣ ·	Ē			)   <del>[</del>	1	עםי	プト
	A1	•		= 91	775																14 -		_	
	A	В	С	D	Е	F	G	Н	Ι	J	K	L	M	N	0	Ρ	Q	R	S	Т	U	V	W	X
34	11425	1.1425	250	38	746	133	1357	451	2358	190	3046	168	3893	560	0	0	0	0	0	0	0	0	0	0
35	11475	1.1475	263	38	757	108	1348	430	2384	238	3016	152	3915	333	0	0	0	0	0	0	0	0	0	0
36	11525	1.1525	263	38	766	88	1440	453	2299	403	3028	140	3955	273	0	0	0	0	0	0	0	0	0	0
37	11575	1.1575	270	40	790	103	1548	730	2261	340	3044	172	4000	249	0	0	0	0	0	0	0	0	0	0
38	11625	1.1625	278	39	791	93	2253	365	3046	209	3950	207	0	0	0	0	0	0	0	0	0	0	0	0
39	11675	1.1675	286	39	804	102	2158	278	3027	261	3939	167	0	0	0	0	0	0	0	0	0	0	0	0
40	11725	1.1725	286	39	825	103	2147	324	3019	319	3925	200	0	0	0	0	0	0	0	0	0	0	0	0
41	11775	1.1775	286	39	848	113	2141	388	3030	339	3912	189	0	0	0	0	0	0	0	0	0	0	0	0
42	11825	1.1825	286	39	860	160	2219	626	3022	367	3924	238	0	0	0	0	0	0	0	0	0	0	0	0
43	11875	1.1875	294	37	885	192	1873	764	2972	366	3867	212	0	0	0	0	0	0	0	0	0	0	0	0
44	11925	1.1925	303	36	933	335	1760	296	2441	412	3021	274	3837	177	0	0	0	0	0	0	0	0	0	0
45	11975	1.1975	303	34	971	307	1705	242	2433	353	3023	288	3836	217	0	0	0	0	0	0	0	0	0	0
46	12025	1.2025	303	32	959	447	1725	213	2501	355	2959	484	3789	220	0	0	0	0	0	0	0	0	0	0
47	12075	1.2075	303	29	1030	403	1682	176	2535	276	3055	324	3882	261	0	0	0	0	0	0	0	0	0	0
		est	202	07	205	E40	054	001	1050	100	0500	<b>√</b> ]^^	0	0		0	0	0	0	0	0	0	Ê	۰Ť
   図形	の調整( <u>R</u> )	)• 🗟 🌀	オート	シェイ	プ(U) +	1 1		e (6	4 🙎	<u>ð</u> -	. <u>/</u> - A	- =	≡ ≓ I		•									
- רב	ンド																	NU	M					

- A列 音声データファイルサンプル番号
- B列 音声データファイルの先頭からの時間 [sec]
- C列 ピッチ周波数 [Hz] (0は無音、7は無声を示します)
- D列 音源パワー [dB]
- E列 ホルマント周波数 (1) [Hz]
- F列 ホルマント・バンド幅 (1) [Hz]
- G列 ホルマント周波数 (2) [Hz]
- H列 ホルマント・バンド幅 (2)[Hz]
- : U列 ホルマント周波数 (9) [Hz]
- V列 ホルマント・バンド幅 (9) [Hz]
- W列 ホルマント周波数 (10) [Hz]
- X列 ホルマント・バンド幅 (10) [Hz]

#### 第4章 技術解説

4-1. 用語説明

- CSVファイル形式 レコード中の各項目を、コンマ(,)を区切りとしたテキストデータフォーマット形式。 Microsoft Excelやロータス 1-2-3などのスプレッドシートアプリケーションでは、汎用フォ ーマットの1つとして、CSVファイル形式でのデータ入出力機能が備えられています。
- Headカーソル 編集のときなど、その対象となる範囲の始まりを示す縦線。
- Tailカーソル 編集のときなど、その対象となる範囲の終わりを示す縦線。
- Paste Buffer 削除した内容等を保存しておく一時的な退避保存用の領域。

 wavファイル
 Windowsで標準的に使われている音声データのファイル形式。例えば、Windowsの起動 音やエラー警告音なども、このファイル形式で保存されています。
 また、Windowsに標準装備されているサウンドレコーダで録音した音や声、音楽も wavフ ァイルとして保存されます。このサウンドレコーダーで、wavファイルを再生することもでき ます。
 wavファイルは、名前の後に「.wav」という拡張子が付いています。しかし、Windowsでは そのままでは拡張子が見えない設定になっています。

ポップアップメニュー Windowsでは開いているウィンドウ毎に、「ファイル」・「編集」といったメニューが並んでいます。このメニューにマウスカーソルを合わせてクリックすると、さらに詳しいメニューが下方向に現れます。このメニューをプルダウンメニューとかドロップダウンメニューと呼びます。プルダウンは「引き下ろす」、ドロップダウンは「垂れ下がる」といった意味で、下方向に出るメニューのことです。 これに対し、画面上のアイコンなどにマウスの矢印を合わせて右ボタンをクリックすると、その場にメニューが開きます。これを、ポップアップメニューと呼びます。ポップアップは「ポンとはね上がる」といった意味を持っています。

#### 4-2. サウンドスペクトログラム

#### 4-2-1. Dynamic Measureについて

メニューの 'Option'- 'Dynamic Measure'チェック時は、WLR距離尺度によるスペクトル変化量を表示します。 Dynamic Measureは、音素セグメンテーションの手掛かりとして有用な場合があると考えています。

計算方法  
DM = 
$$\sqrt{\left|\sum_{n=1}^{P} (r_n^{(R)} - r_n^{(L)})(c_n^{(R)} - c_n^{(L)})\right|}$$
  
DM : Dynamic Measure  
P : LPC分析次数 (24)  
r : LPC自己相関関数  
c : LPCケプストラム  
(R) : サウンドスペクトログラム表示フレームの後 30msec区間計算値

(L) : サウンドスペクトログラム表示フレームの前 30msec区間計算値

#### <u>表示例</u>

サウンドスペクトログラム表示区間内の DMの最大値を 1、DMの平均値の 1/2を 0とノーマライズして表示します。



要注意点: アルチファクト例

上図丸印内のように、30msec計算区間内で生起・終了する短時間のスペクトル変化に対しては DMの値は谷(前後にピーク)になることがあります。

① 分析手順

分析は以下の手順で行われます。



② プリエンファシス (Pre-Emphasis)

有声音・音源のスペクトル全体の概略的な傾きは約 -12dB/octです。一方口唇からの放射特性のスペクトルの概略的な傾きは約 +6dB/octです。このため、有声音のパワースペクトルは -6dB/octの傾きを持つことになります。音声信号処理では、通常 +6dB/octの高域強調(Pre-Emphasis)を施して、スペクトルの傾斜の平 坦化を行うことにより信号対雑音比(S/N比)を改善します。

本プログラムでは、Pre-Emphasisチェック時に Pre-Emphasisをかけます。 例として、白色雑音(一様乱数系列により作成)のパワースペクトルを示します。



#### ③ 窓掛け

#### 〇 時間窓の種類

サウンドスペクトログラムは FFTの計算結果を濃淡図形にして、周波数スペクトルの時間変化を表示する ものです。FFTは有限長の離散的数値データのフーリエ変換であり、しかもその結果は、有限長のデータ がそのまま繰り返している信号のフーリエ変換となっています。このため、音声信号をそのまま FFTした のでは、原信号に本来ならば含まれていないはずの周波数成分が現れることが多いので、それを少なく するために時間窓という重み関数が使われます。

本プログラムでは、時間窓として、Rectangular、Hamming、Hanning、Blackmanの 4種類が選択できます。

以下に、2000Hz, 2200Hz, 5000Hzの3つの正弦波を加え合わせた信号を入力とし、各時間窓を使用した時のパワースペクトルを例示します。

入力信号のサンプリング周波数は 22.05kHzです。

分析条件は、分析窓長 (Window Length) 20msec、Pre-Emphasisあり、2048点 FFTとなっています。



2000Hz, 2200Hz, 5000Hzの 3つの正弦波を加え合わせた信号のパワースペクトル

この入力信号の理想的なパワースペクトルは 2kHz, 2.2kHz, 5kHzの 3本の線となりますが、FFTを用い ている以上はそうはいきません。Rectangular窓を使用したスペクトルでは、2kHz, 2.2kHz, 5kHzの各信号 成分(メインローブ)の幅が狭いので周波数分解能は良さそうですが、他の周波数に大きなサイドローブ が広がるので周波数分析には使いにくくなっています。サイドローブは Blackman窓を使用した場合に最 も小さくなりますが、周波数分解能は低下します。 まとめると、

周波数分解能 : Rectangular > Hamming > Hanning > Blackman サイドローブ : Rectangular > Hamming > Hanning > Blackman となります。サウンドスペクトログラムでは、通常は Hanning窓を使用すれば良いと考えます。

◎ 分析窓長(時間窓長)と周波数分解能

KAY社の旧製品 Sona-Graph Type 6061Aでは Narrow Band分析 45Hz (3dB down points)と Wide Band分析 300Hz (3dB down points)の 2種類の分析が可能でした。

すなわち、周波数分解能を目的に応じて切り替えて分析を行います。周波数分解能と時間分解能にはトレードオフがあるので、Narrow Band分析ではピッチ構造は周波数軸方向の細かい繰り返しとして、Wide Band分析では時間軸方向の細かい繰り返しとして表示されます。

サウンドスペクトログラムでは、Window Length [msec]に分析窓長を入力することで目的の周波数分解能 を実現します。(Narrow)ボタンをクリックすると 60が入り、Wide」ボタンをクリックすると 6.6が入ります。 ただし、指定した値が FFT窓長より長い場合は FFT窓長になるので注意して下さい。) 前述の 2000Hz, 2200Hz, 5000Hzの 3つの正弦波を加え合わせた信号を再び用いて、分析窓長を変化 させた時のスペクトルを例示します。分析条件は Hanning窓、Pre-Emphasisあり、2048点 FFTです。



サウンドスペクトログラム パワースペクトル

前ページのパワースペクトルの 5kHzメインローブを用いて、分析窓長と 3dB down pointsの帯域幅を大 ざっぱではありますが測定してみました。



◎ 分析窓長(時間窓長)と FFT窓の関係

分析窓長が FFT窓長に満たない時は中央部に音声データを置き、残り部分に 0を詰めて FFTを実行 します。



最長分析窓長は FFT窓長と音声信号のサンプリング周波数で決まります。

サンプルノガ国波粉		FFT	窓長	
リンノリンク回収数	256	512	1024	2048
11.025kHz	23.2msec	46.4msec	92.8msec	185.7msec
22.050kHz	11.6msec	23.2msec	46.4msec	92.8msec
44.1kHz	5.8msec	11.6msec	23.2msec	46.4msec

#### 出力時の高域強調

+6dB/octの高域強調(Pre-Emphasis)を施すと、スペクトルの傾斜の平坦化を行うことにより信号対雑音比 (S/N比)を改善することは前述しました。それとは別に出力時(表示時)において単純に高域のスペクトルを 強調することができます。

サウンドスペクトログラムでは、High-Boostチェック時に、24dB/10kHzの傾斜で高域を強調します。

例として、白色雑音(一様乱数系列により作成)のパワースペクトルを示します。 プリエンファシスとは独立しているので、その組み合わせを示します。



#### ⑤ 音声信号のサンプリング周波数とスペクトル上限周波数

サンプリングとは、時間的に連続な波形(アナログ信号)を、時間的に離散的な時点の値の系列(ディジタル 信号)で表現することです。

サンプリング周波数とは、サンプリングを行う時の離散的な時点の時間幅の逆数(簡単に言えば、1秒間に行うサンプリングの回数)です。wavファイルの標準的なサンプリング周波数は 11.025kHz, 22.05kHz, 44.1kHz です。音声分析のためには通常 22.05kHzまでで充分です。

FFTで計算して求めた周波数スペクトルの上限の周波数はサンプリング周波数の 1/2です。例えばサンプリング周波数が 22.05kHz の場合のスペクトル上限周波数は、11.025kHzとなります。

#### 5-1. 共通仕様

・音声データファイル読み込み								
ファイル形式	WAVファイル形式							
フォーマットタイプ	PCM (ADPCM等の圧縮音声データファイルは読み込めません。)							
量子化ビット数	8ビット、または、16ビット							
エレラル粉	1(モノラル)、または、2(ステレオ)							
	ステレオの場合、左右どちらのチャネルを読み込むか指定します。							
星十立吉データ트	音声波形編集は、24秒							
	音声波形編集以外は、無制限(ただし、パソコン搭載のメモリ容量に依存します。)							
サンプリング周波数	8/10/11.025/12/22.05/24/44.1/48kHz							
•音声再生								
音声波形編集は、常にモ	ミノラルで再生されます。							
音声波形編集以外は、	読み込んだ音声データファイルがステレオ音声の場合、読み込んだチャネルに係わらず							

ステレオ音声が再生されます。

### 5-2. 音声波形編集

·音声録音·再生			
量子化ビット数	16ビット		
チャネル数	1(モノラル)		
サンプリング周波数	8/10/11.025/12/22.05/24/44.1/48kHz		
最大録音時間	24秒		
・音声データファイル保存			
ファイル形式	WAVファイル形式		
フォーマットタイプ	PCM		
量子化ビット数	16ビット		
チャネル数	1(モノラル)		
サンプリング周波数	8/10/11.025/12/22.05/24/44.1/48kHz		
最大音声データ長	24秒		
・編集機能			
削除、コピー、貼り付け、	無音化、加算、減算、フェードイン、フェードアウト、正弦波発生、ノイズ発生、インパルス		
列発生、振幅増減、振幅正規化			

### 5-3. サウンドスペクトログラム

·分析条件	
分析窓長	Narrow (60msec), Wide (6.6msec), 任意設定
時間窓	Hamming, Hanning, Blackman, Rectangular
FFT窓長	256, 512, 1024, 2048
フレームシフト	自動、任意設定
高域強調	Pre-Emphasis, High-Boost

### 5-4. ピッチ抽出

・ピッチ抽出	
ピッチ抽出方法	「柔軟性のあるピッチ抽出法」 川浦、粕谷 宇都宮大学工学部
分析条件	フレームシフト、無音閾値、ピッチ探索最小・最大値、有声・無声閾値

編集機能	無声化、2倍、1/2、指定区間の内挿、指定値に変更
・ファイル保存	
ファイル形式	CSVファイル形式
保存データ	音声パワー、ピッチ周波数

## 5-5. ホルマント分析

・ホルマント分析		
ホルマント分析方法	LPCスペクトル包絡のピークピッキング	
	バンド幅は2次曲線あてはめによる	
分析条件	分析窓長、ピーク強調	
<ul> <li>・ピッチ抽出</li> </ul>		
ピッチ抽出方法 「柔軟性のあるピッチ抽出法」川浦、粕谷 宇都宮大学工学部		
分析条件 フレームシフト、無音閾値、ピッチ探索最小・最大値、有声・無声閾値		
・ファイル保存		
ファイル形式 CSVファイル形式		
保存データ	ホルマント周波数、ピッチ周波数	

### 第6章 初期値

音声録聞見 for Windowsの各プログラムは、終了時の設定を自動的に保存し、次回の起動時にはその設定を引き継ぎます。また、Resetメニュー/ボタンを実行すると、全ての設定を初期値に戻すことができます。 本章では、各プログラムの初期値を記しています。

### 6-1. 音声波形編集

(1)	録音周波数	11025Hz
2	録音時間	12秒
3	誤録音防止機能	オフ
4	再生繰り返し	無し
5	Hear/Tail Cursor Part表示部時間幅	70msec
6	Sine Wave振幅、周波数	90%, 1000Hz
$\bigcirc$	Noise振幅	90%
8	Impulse振幅、周波数	90%, 150Hz
9	Amp. Increase	25%
10	Amp. Decrease	25%
	Amp. Normalize	100%

### 6-2. サウンドスペクトログラム

① Pre-Emphasis	オン
② High-Boost	オフ
③ Auto Frame Shift	オフ
④ Frame Shift	5.0msec
⑤ Window Length	Wide (6.60msec)
⑥ Frequency Auto Detect	オン
⑦ Frequency Upper	5000Hz
⑧ Frequency Lower	0Hz
(9) FFT	1024ポイント
10 WIndow Type	Hanning
1 Black Level	240
12 White Level	50
$(3) \gamma +$	オフ
(14) γ -	オフ

### 6-3. ピッチ抽出

(1)	Draw Power	オン
2	Log Scale	オン
3	Zoom Length Fix	オフ
4	Zoom Length	3.0sec
(5)	Frame Shift	5msec
6	Silent Level	300
$\bigcirc$	F0 Maximum	310Hz
8	F0 Minimum	60Hz
9	VUV Level	1.0

## 6-4. ホルマント分析

$\bigcirc$	Zoom Length Fix	オフ
2	Zoom Length	3.0sec
3	Frame Shift	5msec
4	Silent Level	300
5	F0 Maximum	310Hz
6	F0 Minimum	60Hz
$\bigcirc$	VUV Level	1.0
8	F.Window	25msec
9	Peak Enhance	オフ

### 第7章 参考書·参考文献

今川 博・桐谷 滋	「DSPを用いたピッチ、フォルマント実時間抽出とその発音訓練への応用」 電子情報通信学会技術研究報告, SP89-36, 17-24, 1989.
Hiroshi Imagawa and Shigeru Kirita	ni 「High-speed Speech Analysis System Using a Personal Computer with DSP and its Applications to Pronunciation Training.」 Ann. Bull. RILP, University of Tokyo No.23, 173–185, 1989.
今石 元久	「日本語音声の実験的研究」,和泉書院,1997.5.29
川浦 ・ 粕谷	「柔軟性のあるピッチ抽出法」,宇都宮大学工学部 電子情報通信学会技術研究報告,SP87-3,1987.4.23
粕谷 英樹・他	「音声合成のための音源モデルに関する検討」 日本音響学会講演論文集, 平成2年9月, 1-6-19, 213
三輪 譲二	「パソコン音声処理」, [パソコン活用シリーズ8], (株)昭晃堂, 1991.7.5
斎藤 収三・中田 和男	「音声情報処理の基礎」, (株)オーム社, 昭和57年
古井 貞熙	「ディジタル音声処理」,[ディジタル テクノロジー シリーズ ⑥] 東海大学出版会, 1988.9.25
杉山 雅英・鹿野 清宏	「ピークに重みをおいたLPCスペクトルマッチング尺度」 信学論, J64-A, 5, 409-416, 1981 (昭和56)
嵯峨山 茂樹 ・ 板倉 文忠	「音声の動的尺度に含まれる個人性情報」 音学春季講論, 3-2-7 (昭和54)
城戸 健一	「周波数分析用と畳み込み演算用の時間窓選択」 日本音響学会誌, Vol.55 No.5, 360-363, 1999.

索引

### А

Add	
All	14, 20, 25
Amp. Decrease	
Amp. Increase	12
Amp. Normalize	
Auto frame Shift	15

#### В

С

Copy					12
CSVファイル	19,	23,	24,	27,	28
Cut					12

#### D

Delayed Auditory Feedback	13
Draw Power	20
Dynamic Measure1	5,29

#### Е

Execute.....14, 20, 25

F

F Window	25
Fado In	
Fade Out	
Faue Out	15 17 20 21 24
	10, 17, 50, 51, 54
FF1念友	15, 17, 26, 31, 33
Frame Shift	15, 20, 25

#### Η

Hamming	15,	30,	33
Hanning15, 30,	31,	32,	33
Headカーソル11, 12, 13, 15, 17, 20,	23,	26,	28
High-Boost		15,	34

#### Ι

Impulse	12
Interpolate	22

L
Log Scale
М
Main分析窓長23, 26
N
Narrow15, 31 Noise12
0
OS2
Р
Paste Buffer       12, 28         Paste(H)       12         Paste(T)       12         Pause       12         Peak Enhance       25         Play       12, 14, 19, 25         Pre-Emphasis       15, 30, 32, 34
R
Record
S
Settings       12         Silence       12         Silent Level       19, 20, 24, 25         Sonagraph.bmp       18         Sound       12, 14, 19, 25         Spectrum       17         Stop       12         Sub       12
Т
Tailカーソル11, 12, 13, 15, 20, 26, 28

U	
Unvoiced	20
UnVoiced	

### V

Version	13,	15,	20,	25
Voiced	•••••		20,	25

#### W

wav11	, 14, 19	9, 23, 2	24, 26,	28,	34
Wide				15,	31
Window Length			15,	30,	31
Window Type					15
WLR距離尺度		•••••		15,	29

#### Ζ

Zoom	20,	23,	25,	26
Zoom Length Fix			20,	25

#### あ

アプリケーションの追加と削除		8
アンインストール	8,	9

### い

印刷	12, 14, 19, 24
インストール	
インパルス	

### お

音源パワー		24,	27
音声波形編集	1,	10,	11
音声パワー 1	9	20.	21
н, ,	~,	_0,	

### か

加筲 19		
$\pi + 12$	加算	. 12
ガンマ15	ガンマ	.15

### き

共有ファイル ......9

### Ĺ

コントロールパネル		••••	•••	2	, 8
コンピュータ	2,	3,	6,	8,	13

#### さ

4
5
31
14, 20, 25
14, 29, 30
2
8, 12
24
30, 33, 34
16, 21, 27

### L

時間分解能	31	
自己相関関数 20, 25,	29	
周波数分解能	31	

### す

スタートメニュー	
ステレオ	11, 13, 18, 23, 26
スピーカ	2
スペクトル	17, 29, 30, 31, 32, 34

### せ

正規化	
正弦波	
セットアップ	6
全分析窓長	23, 26

5		
遅延エコー	1,	13

## て ディスプレイ.....2

と	

動作環境......2

# $\mathcal{O}$

ノイズ	12
ノーマライズ	29

#### は

バージョン	13,	15,	20,	25
ハードディスク				2
パワースペクトル		.30,	31,	34

ひ

ピッチカーブ	20,	21
ピッチ周波数 19, 21,	24,	27
ビッチ探索	20,	25
ピッチ抽出1, 10, 11, 13, 19, 20, 23,	24,	25
ビットマップ	•••••	18

### Ş

フレーム.....17, 20, 25

 $\sim$ 

#### ほ

ポップアップメニュー	
ボリュームコントロール.	
ホルマント・バンド幅	
ホルマント周波数	

ホルマントバンド幅	
ホルマント分析	1, 10, 24, 25
	, , ,
±	
6	
マイク	2
<i>t</i> o	
2	
無音	12, 27
無声	20, 21, 25, 27
	, , ,
×	
<i>wy</i>	
メインローブ	31, 33
メモリ	2
¢.	
0	
モノラル	
Ŵ	
有声	20, 25
3	
録音音量	3

### ご注意

- (1) 本書は、Windows対応音声分析合成ソフトウェア「音声録聞見 for Windows」のユーザーマニュアルです。
- (2) 本書に記載されている事柄は、将来予告なしに変更することがあります。
- (3) 本書の一部あるいは全部について、無断で複製することはできません。

音声録聞見 for Windows (CD-ROMフリーウェア版) ユーザーマニュアル

2005年3月 第8版