音声認識・音声対話技術講習会

実習1：
音声の音響分析と特徴量抽出

立命館大学　情報理工学部
西浦　敬信
nishiura@is.ritsumei.ac.jp

College of Information Science and Engineering,
Ritsumeikan University

Aug. 28. 2006

実習の準備

電源投入

- 正面にあるDELLの黒い計算機 (電源スイッチは本体の正面中央付近)
- ディスプレー (電源スイッチはディスプレーの裏側)

立ち上がるまで待つ. . .

ログイン

Login画面 ⇒ ws***へようこそ！
「ユーザ名:」 ⇒ sp** (各自のアカウントを入力)
「パスワード」 ⇒ **** (各自のパスワードを入力)

実習の準備

ターミナルの起動

“右”クリック

⇒ 「新しいターミナル」を“左”クリック

パスワードの変更

% yppasswd

old passwd:***** (最初のパスワードを入力)

new passwd:***** (各自が考えた新しいパスワードを入力)

作業ディレクトリへ移動 (今後の作業はこのディレクトリにて行う)

% cd ~/exercise/work

実習1の資料 (スクリーンが見づらい場合はこちらを参照のこと)

$ acroread /n/media/sp/doc/exercise1.pdf &

はじめに

・実習内容：音声の取り込み，分析，特徴量の抽出までの実践

(1) 音声の研究でよく使われる音声ファイル形式
(2) 音声の取り込みと学習用データの作成
(3) スペクトログラム
(4) ケプストラム分析
(5) 特徴量の抽出

音声の研究でよく使われる音声ファイル形式

(1) ヘッダーがあるかどうか

(2) サンプリング周波数・・・公倍数，公約数になっていないものの変換は困難

- 音声の研究：8000 Hz，16000 Hz，
  DATが48000 Hz
  G.772（IEEE国際標準規格）が16000 Hzである事などに起因
- マルチメディア系：11025 Hz，22050 Hz，44100 Hz
  CDが44100 Hzであるため

(3) チャネル数

・・・1ならモノラル，2ならステレオ

音声の研究でよく使われる音声ファイル形式 (2)

(4) ビット数（ビット / サンプル）

・・・音声研究では16ビットがほとんど

(5) バイトオーダー（エンディアン）

通常，計算機は，1バイト（8ビット）より大きなデータは1バイトごとに記憶するが，その時の記録する順番．この順番は一般にCPUにより異なる

(6) 圧縮方式

・・・音声の研究では通常圧縮を行わない

音声の研究でよく使われる音声ファイル形式 (ヘッダーなし)

・Rawファイル（.raw，.ad，.datなど）

ーサンプリング周波数：不定，チャネル数：不定，ビット数：不定
ーパラメータは，自分で覚えておくか，ファイル名に付けておく必要性
ーバイトオーダーが環境によって違うことに注意
・Sun，Sparc，Macintosh，SGI・・・Big Endian
・Windows，Linux (i386)，Compaq Alpha・・・Little Endian

・Text（.txt，.text）

ーサンプリング周波数：不定，チャネル数：不定
ー読み込み方，書き込み方によっては値が丸められる
ー環境によってそれほど変化はない
ー改行コードに注意
・Unix系・・・LF，Windows・・・CR+LF，Mac・・・CR

音声の研究でよく使われる音声ファイル形式 (ヘッダーあり)

- Aiff format（.aiff，.aif）

ーAudio Interchange File Formatの略
ーAppleによって提案されたフォーマット
ー音声データだけでなく，楽器情報なども保存可能
ーAIFC(.afc)は，AIFFに圧縮のサポートが加わったもの
[バイトオーダー] Big Endian
[環境] Mac，SGI
[変数] rate，#channels，#bits / sample (1-32)，etc.

音声の研究でよく使われる音声ファイル形式 (ヘッダーあり)(2)

・Riff WAVE file format（.wav）

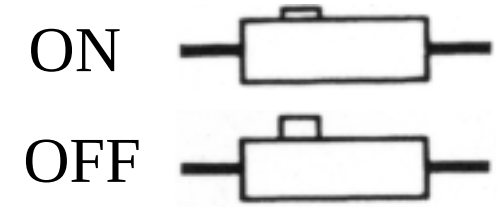
ーMicrosoftとIBMが，AIFFを参考にして作ったといわれる
　フォーマット
ーWindowsで標準の音声ファイル形式であるため，広く利用
[バイトオーダー]　Little Endian
[環境]　Windows
[変数]　rate，#channels (mono or stereo)，bits / sample(8-32)，
　　etc.

今回は，Big EndianのRawファイル（ヘッダーなし）を使用

音声の取り込み準備

・ヘッドセットマイクの接続

－計算機の裏側にある赤，青，ピンク，緑，黒，黄端子を持つボードにヘッドセットマイクを接続
－ヘッドセットマイクの「ピンク」端子　⇒　ボードの「ピンク」端子へ
－ヘッドセットマイクの「ベージュ」端子　⇒　ボードの「緑」端子へ
－ヘッドセットのスイッチは「ボタンがより押し込んだ状態」がON

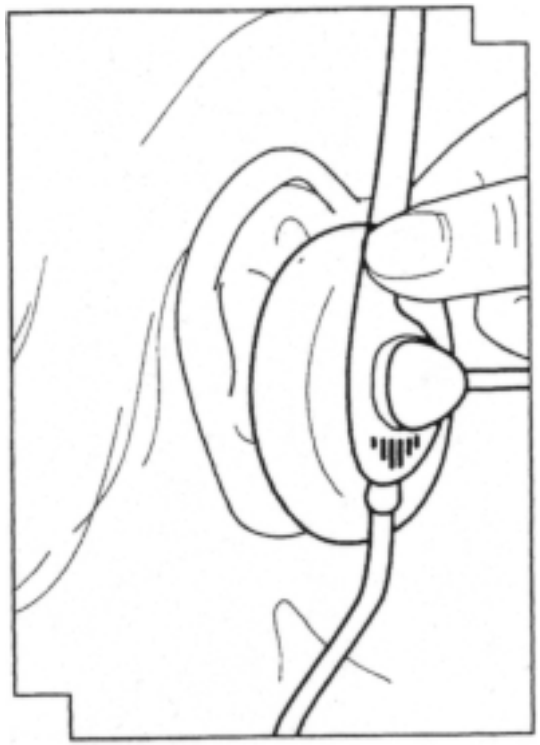


・音量の設定

－左下「赤帽子」ボタンを左クリック
－　⇒　「サウンドとビデオ」を選択後，「音量コントロール」を左クリック
－「vol」，「pcm」，「speaker」のスライダを真ん中付近まで上げる。
－「mic」と「igain」のスライダを真ん中より少し上くらいまであげる
－「mic」は無印，「igain」は「連動」と「録音をクリック」（印をつける）

ヘッドセットの装着

・先にイヤーピースを耳に当ててから装着するとより良いフィット感が得られます。



・マイクロフォンの装着位置は、直接息がかからないように、図のように口元の脇からブームの先端まで約2cm程度が最適です。

音声の取り込みテスト

・Wavesurferによる録音確認

```
$ wavesurfer
```

(1)画面の「sound」と書かれた青色のバーを右クリック
　　⇒　「Create Pane」を選択後、「Waveform」を選択
(2)録音ボタン ● を押すと取り込み開始，停止ボタン ■ を押すと終了
　　⇒　録音波形が表示される。
(3)表示波形を確認、必要に応じて再生ボタン ▶ により録音音声を再生
(4)録音レベルを確認して、「音響コントロール」の「mic」と「igain」のスライダや「vol」，「pcm」，「speaker」のスライダを調節する

音声の取り込みと学習用音声データの作成

■ ■ ■演習課題1 ■ ■ ■

リスト1に示す50単語(数字：学習用)と5文(テスト用)を，音声録音用ツール(record.pl)を用いてワークステーション上に取り込んでください。ファイル名はrecord.plを実行した際に一番上に表示され，自動で発話を検出しファイルに保存します。なお本演習では50単語（数字：学習用），5文(テスト用)の順にファイル名が表示されますので注意願います。

(注)

- 発話は読み誤りがないよう、必ず発話後に内容を確認してください。
- 今後の演習では、これらの音声ファイルを使用しますので、間違いがないよう収録して下さい。
- ファイルは、~/exercise/workディレクトリで作業すれば自動的に~/exercise/speechの下に保存されます。

音声の取り込みと学習用音声データの作成 (2)

・録音用ツールの起動

－ record.pl

－使い方

・録音後，%のあとにコマンド入力

l	：	今録音した音声を再生
r	：	今の録音をやり直し
b	：	前の文に戻る
n	：	次の文に進む
m X	：	文章 X へ移動 (例: m d01)
q	：	終了

音声の取り込み例

record.plの実行

```
$ record.pl
```

```
[d01]  *未録音
       0  (ぜろ：1回目)
```

録音ファイル名　と　現在の状態

発話内容の提示(回数：数字のみ)

```
fragment size = 1024 bytes (32msec)
AD-in thread created
<<<  please speak  >>>
```

この状態で音声入力OK
「……　」が表示されれば録音中

```
%
```

%　が表示されたら録音完了。
次に進む場合は％ｎを実行すること
（　録音しなおす場合は％ｒ　を実行）

音素バランス文の取り込み (1)

■ ■ ■演習課題2 ■ ■ ■

演習課題1と同じ要領で、音素バランス文(20文)を，音声録音用ツール(record_b20.pl)を用いてワークステーション上に取り込んでください。
なお、音素バランス文の録音では「お手本音声」が再生されますので、できる限りお手本のイントネーションを参考に発話してください。
さらに、時間に余裕のある受講者は、追加で音素バランス文(30文)を，音声録音用ツール(record_b30.pl)を用いてワークステーション上に取り込んでください。

(注)

- 発話は読み誤りがないよう、必ず発話後に内容を確認してください。
- なお、この収録音声は音声認識・音声合成の両方の講習で使用しますので、間違いがないよう収録して下さい。
- たくさん時間をとりますので、できるだけ追加の音素バランス文（30文）も録音してください。

音素バランス文の取り込み(2)

・音素バランス文録音用ツールの起動

ー record_b20.pl （30文の場合は record_b30.pl ）

ー使い方

・録音後，%のあとにコマンド入力

l	：録音した音声を再生
r	：録音やり直し
R	：録音やり直し（お手本無し）
t	：お手本を聞き直す
b	：前の文に戻る
n	：次の文に進む
m X	：文章 X へ移動（例： m d01）
q	：終了

音素バランス文の取り込み例

record_b20.plの実行 (record_b30.pl も同じ)

最初にターミナルを最大化する
(ターミナルウィンドウの右上の3つのボタンの真ん中を左クリックする)

```
$ record_b20.pl
```

[d01]　*未録音

録音ファイル名　と　現在の状態

あらゆるげんじつを、すべてじぶんのほうへねじまげたのだ。
あらゆる現実を、すべて自分のほうへねじ曲げたのだ。

お手本音声のイントネーションを参考に発話すること

fragment size = 1024 bytes (32msec)
AD-in thread created
<<< please speak >>>

%

この状態で音声入力OK
「……　」が表示されれば録音中

音素バランス文の取り込みに関する注意点

- 1. **文章を読み間違えない**
  - * 文章を読み間違えると、後続の音声認識や音声合成講習において、大きな問題が生じます。録音後は必ず発声内容を確認してください。例えば、「日本」：にほん or にっぽん，など。
- 2. **テキストにある通りに「、」でポーズを入れる**
  - * もし提示音と食い違う場合は，テキストに従ってください
  - * 昨年あった例：　単語ごとに切って読む（×）
- 3. **アクセントは，ガイド音声とできるだけ同じにする**
  - * アクセントは音の強弱よりも，高低でつけるように。
  - * 提示音声のアクセントを参考にしてください。
- 4. **発音は恥ずかしがらずにハキハキと**
- 5. **その他**
  - * ヘッドセットをしっかりと頭に装着
  - * 入力レベル調整を念入りに
  - * マイクに近づきすぎない

上記の注意事項に注意して録音すると，認識精度や合成音声の品質が良くなります．理由は各実習にて詳細に説明予定

スペクトログラム

・スペクトログラム

・・・短時間スペクトルの変化を濃淡表示したもの

・スペクトログラムの表示

(1) WaveSurferの実行
(1.0.1より新しいものであればrawファイルにも対応)

```
$ wavesurfer &
```

(2)「File」メニューから「Open...」を選択し，ファイルダイアログを開く
(3) ダイアログの「Files of type:」で「All Files (*)」を選択し，ファイルを開く
(4)「Interpret Raw File As」というダイアログが開くので，
16000 Hz，Lin16，Mono，Big Endianをそれぞれ選択し，OKを押す
(5)「Choose configuration」というダイアログが開くので，リストから
「Speech analysis」を選択し，OKを押す

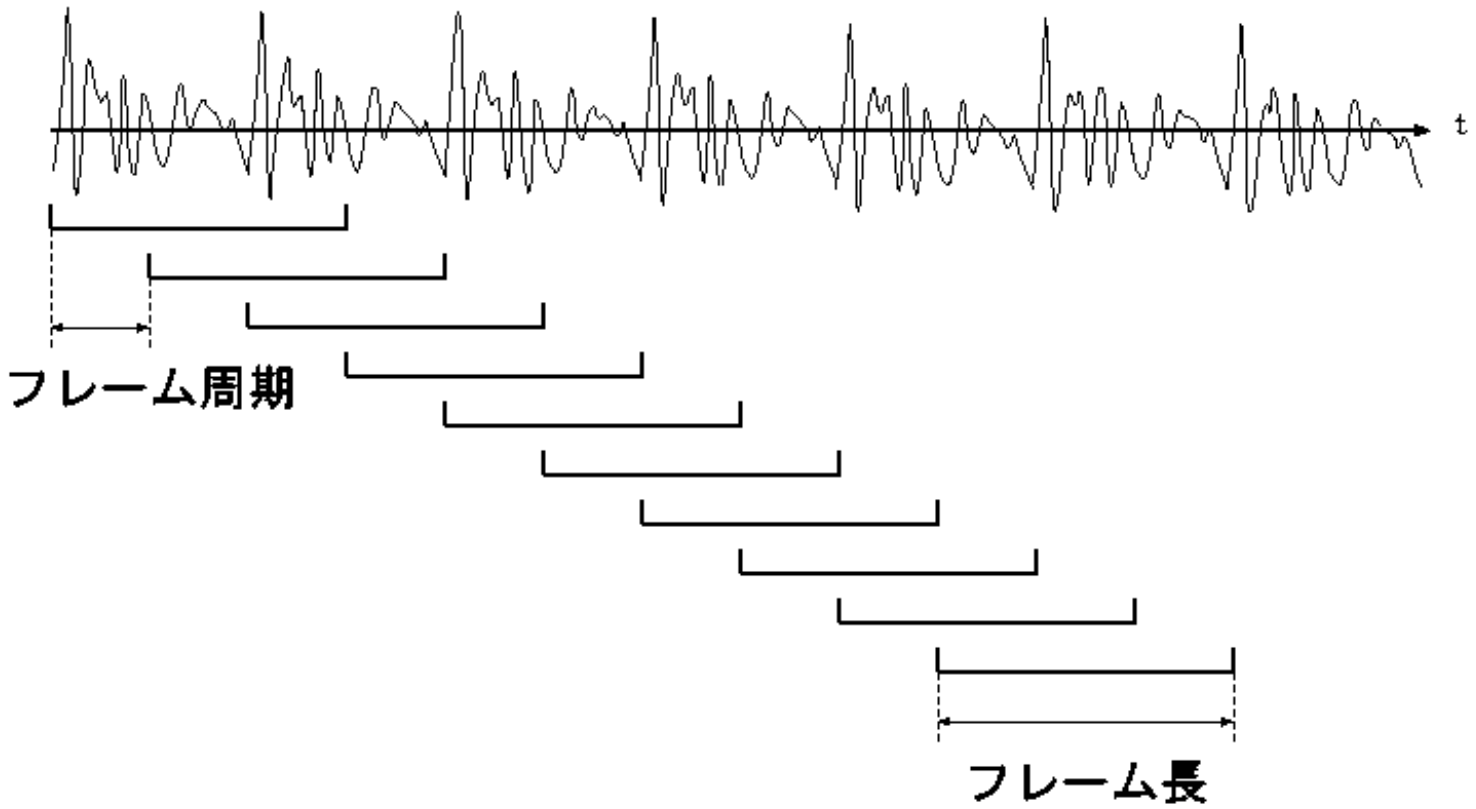
スペクトログラム (2)

■ ■ ■ 演習課題3 ■ ■ ■

課題1で収録した，数字音声のデータのうち1つをwavesurferを用いて表示し，数字を構成する音素が，スペクトログラム上のどの区間に対応するかを調べて下さい。次に/SAN/，/NANA/に含まれる音素/a/の中心付近における短時間スペクトルを表示し，それらがほぼ同一の形状をもつことを確認して下さい。最後に，/KYU:/の第1回目と5回目の発声を比較し，両者の相違が主として時間軸上の(非線形)伸縮で正規化可能であることを確認して下さい。

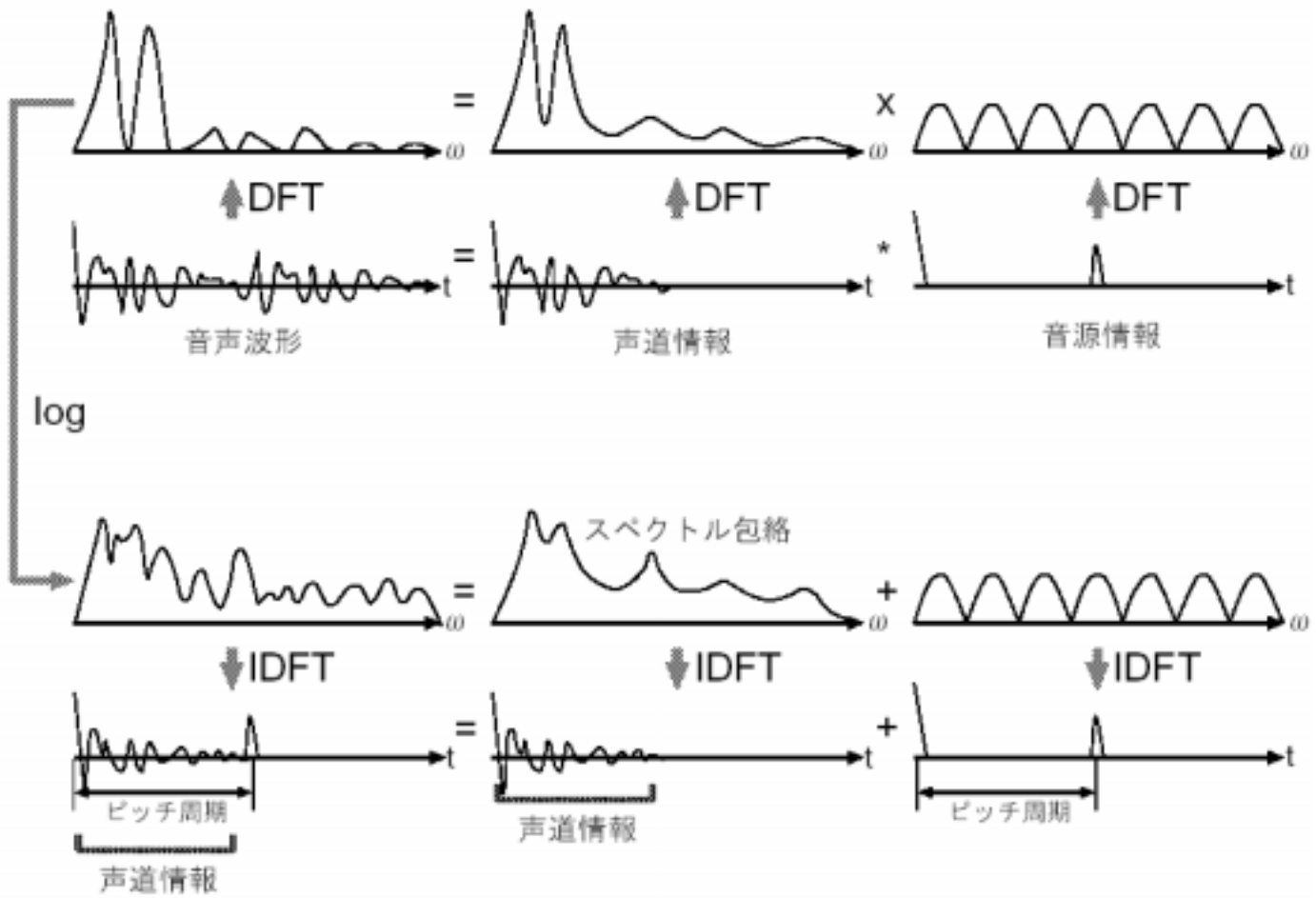
ケプストラム分析

(1) 短時間分析



ケプストラム分析 (2)

(2) ケプストラムの低次成分からスペクトル包絡を抽出



特徴量の抽出

■ ■ ■ 演習課題4 ■ ■ ■

課題1、課題2で収録したデータをHCopyを用いて分析し，対応するMFCCパラメータファイルを作成して下さい。分析により得られたパラメータファイルには，リストに示す通りの名前を付けて，リスト2に示すディレクトリ構成に従ってディスク上に格納して下さい。

- HCopyを用いてスペクトル包絡列を抽出
- MFCC (Mel Frequency Cepstrum Coefficient)を使用

全受講者

```
$ HCopy -T 1 -C ../config/config.HCopy -S ../script/HCopy.scp
```

上記実行後、さらに音素バランス30文も録音した受講者

```
$ HCopy -T 1 -C ../config/config.HCopy -S ../script/HCopy_b30.scp
```

最後に

実習を終えるときは必ずログアウトすること

左下の「赤帽子」ボタンを左クリック
⇒ 「ログアウト」を左クリック
⇒ 「ログアウト」を選択し「OK」を左クリック

☆ 注意 ☆

実習終了後、ヘッドセットマイクロホンは出口にて回収しますので、よろしくお願いします。おつかれさまでした。