TR-IT-0340

CHATRで用いる設定パラメータの最適化

北川 敏
Satoshi Kitagawa

2000.2

概要

本報告は，CHATRで得られる合成音声の韻律の向上のために，CHATRで用いている設定パラメータの最適化の手法について記すものである．

目次

第 1 章

はじめに

第二研究室で開発された自然音声波形接続型音声合成システム CHATR [1] は，予め録音した音声データベース中の音素単位の音声波形を，信号処理を行わずに接続して連続音声として出力するため，自然性の高い合成音声が得られることが最大の特長である．しかし，単位選択時の不完全な音素単位アルゴリズムなどのために，常に聴覚的に最適な音素が選択されるとは限らず，得られる合成音声のイントネーションが不自然であるという問題がある．CHATR では，音素単位選択のためにさまざまな設定パラメータを使用するが，その設定の仕方によって得られる合成音声の韻律は変化する．これまでは経験則にもとづいた値が使用されてきており，必ずしも最適値であるとは言い難い．

そこで，本レポートでは CHATR の日本語の韻律に関する研究として，音素単位選択の際に使用するパラメータの最適値の求め方について記述する．また，発話様式が異なる場合のパラメータへの影響についても調べた．

第 2 章

パラメータの最適化

2.1 CHATR における単位選択

CHATR は音素単位による波形接続型音声合成システムであるため，音素単位選択の結果が合成音声の品質に大きく影響を与える．ここで使用する単位選択には複数の方法があるが，いずれの方法においても設定パラメータが必要である．このパラメータの値はおおまかに設定されたものであり，最適な値であるかどうかの検証は行っていなかった．それは，合成音声の品質の評価は聴取実験に頼るしかなかったためである．この場合，評価は聴取する人の主観により左右され客観的ではなく，また評価できるデータ量も限定されてしまう．ここで，計算機によって自動評価が行えるようになると，これらの問題は解消される．このことに対して，自然音声と評価対象の合成音声の比較による評価方法を開発したことにより，計算機上での評価ができるようになった．[2]

そこで，この比較による違いを元にしてパラメータの最適値を求める方法について調べてみた．

2.2 評価対象音声の作成

CHATR では，さまざまな設定パラメータを使用している．そのため，全てのパラメータの組み合わせについて計算を行うと，膨大な処理時間がかかるために，2 段階に分けて計算を行うことにした．

はじめに，設定パラメータのうち，以下の組み合わせについて合成音声の作成を行った．

```
(beam_width 25 50 100 200 500)
(cand_width 25 50 100 200 500)
(cand_max 25 50 100 200 500)
```

この中で，最も良い結果が得られた組み合わせに対して，以下のパラメータの最適値を求めることにした．

今回，合成音声の作成に際して単位選択手法としては UDB を用いた．この UDB で使用する設定パラメータにも様々なものがあるが，ここでは以下のパラメータの値を変更して使用した．

```
(join_wt 0.2 0.5 0.8)
(unit_wt 0.2 0.5 0.8)
(unit_phone_wt 0.2 0.5 0.8)
(lcontext_wt 0.2 0.5 0.8)
(power_wt 0.2 0.5 0.8)
(dur_wt 0.2 0.5 0.8)
(pitch_wt 0.2 0.5 0.8)
(pros_context_wt 0.2 0.5 0.8)
(vq_wt 0.2 0.5 0.8)
(vq_f0_wt 0.2 0.5 0.8)
(vq_pow_wt 0.2 0.5 0.8)
(phone_context_wt 0.2 0.5 0.8)
```

2.3 音声の比較

2つの音声を比較する場合，次の2つの特徴の比較があげられる．

1. 音声波形の物理的形状
2. 韻律情報

そこで，はじめに物理的形状を，次に韻律情報を利用して調べた結果について述べる．

第 3 章

音声波形の物理的形状の比較

3.1 比較方法

ここでは、物理的形状を比較することによりその類似度を調べた.

2つのケプストラム距離を求め，その差が小さいものほど類似性が高いとみなし，高品質であると判断する．この時に用いた関数は、 CHATR の内部コマンドである Compare_Cepstrums である.

評価対象の話者データベースとして MHT，評価用の合成音声は情報量の多い MHT_SD_E43 を用いた．検証を目的としているため，評価は 1 文だけに対してのみ行っている.

3.2 beam_width, cand_width, cand_max の最適値の算出

beam_width, cand_width, cand_max を次の組み合わせで用いた場合の結果を図 3.1に示す.

```
(beam_width 25 50 100 200 500)
(cand_width 25 50 100 200 500)
(cand_max 25 50 100 200 500)
```

図 3.1の左から順に，beam_width，cand_width，cand_max のパラメータ別にまとめたものである．縦軸はケプストラム距離を示し，数値が小さいほど類似度が高いことを示す.

この結果をもとに,

```
(beam_width 25)
(cand_width 25)
(cand_max 100)
```

の組み合わせを決定し、次節のような評価を行った.

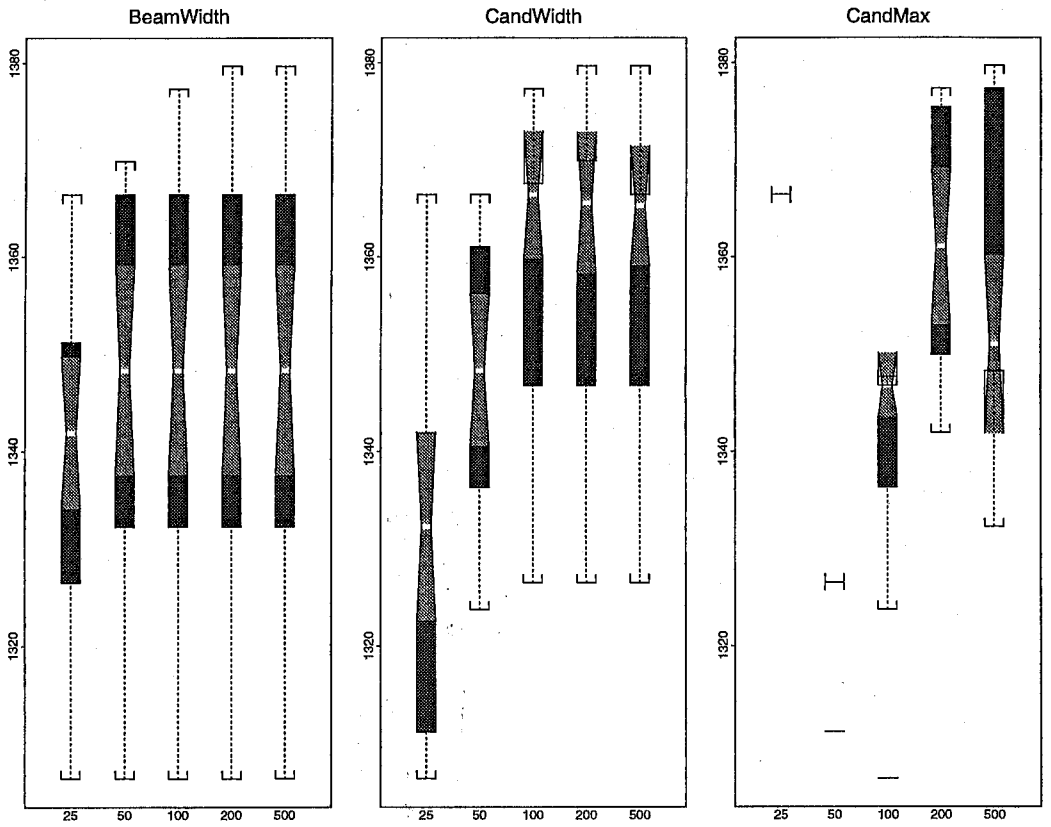

図 3.1: beam_width, cand_width, cand_max によるケプストラム距離の変化

3.3 nus_params で用いるパラメータの最適値の算出

ここで使用した nus_params の組み合わせは，次の通りである．

```
(join_wt 0.2 0.5 0.8)
(unit_wt 0.2 0.5 0.8)
(unit_phone_wt 0.2 0.5 0.8)
(lcontext_wt 0.2 0.5 0.8)
(power_wt 0.2 0.5 0.8)
(dur_wt 0.2 0.5 0.8)
(pitch_wt 0.2 0.5 0.8)
(pros_context_wt 0.2 0.5 0.8)
(vq_wt 0.2 0.5 0.8)
(vq_f0_wt 0.2 0.5 0.8)
(vq_pow_wt 0.2 0.5 0.8)
(phone_context_wt 0.2 0.5 0.8)
```

前節と同様の方法で求めた結果が

/home/as68/satoshi/data1/work/opt-param/MHT/Cepstrum/MHT_SD_E43.25_25_100.out

である．

ファイルサイズが膨大になったために，ケプストラム距離の値が 2 以下のものについてまとめた。該当する組み合わせは２７通りで、その内訳は次の通りであった．各パラメータの上段がパラメータの値，下段がその度数である．

```
(join_wt)           0.5 0.8
                      9  18
(unit_wt)           0.2 0.5
                     18   9
(unit_phone_wt)     0.2 0.5 0.8
                      9   9   9
(lcontext_wt)       0.2 0.5
                     18   9
(power_wt)          0.2 0.5 0.8
                      9   9   9
(dur_wt)            0.2 0.8
                      9  18
(pitch_wt)          0.5 0.8
                      9  18
(pros_context_wt)   0.2 0.5
                      9  18
(vq_wt)             0.8
```

```
                    27
(vq_f0_wt)          0.2
                    27
(vq_pow_wt)         0.8
                    27
(phone_context_wt)  0.2
                    27
```

この結果をもとに，次の組み合わせによる評価を行った．

3.4 nus_params で用いるパラメータの最適値の算出（その２）

前節ではパラメータの変化の範囲がおおまかであったため，本節では変化幅を狭めて次の組み合わせによる評価を行った．

```
(join_wt 0.7 0.8 0.9)
(unit_wt 0.1 0.2 0.3 )
(lcontext_wt 0.1 0.2 0.3)
(dur_wt 0.7 0.8 0.9)
(pitch_wt 0.7 0.8 0.9)
(pros_context_wt 0.4 0.5 0.6)
(vq_wt 0.8)
(vq_f0_wt 0.2)
(vq_pow_wt 0.8)
(phone_context_wt 0.2)
```

結果を図3.2～図3.7に示す．

以上のことを繰り返すことで，最適値を求めることが可能になった．

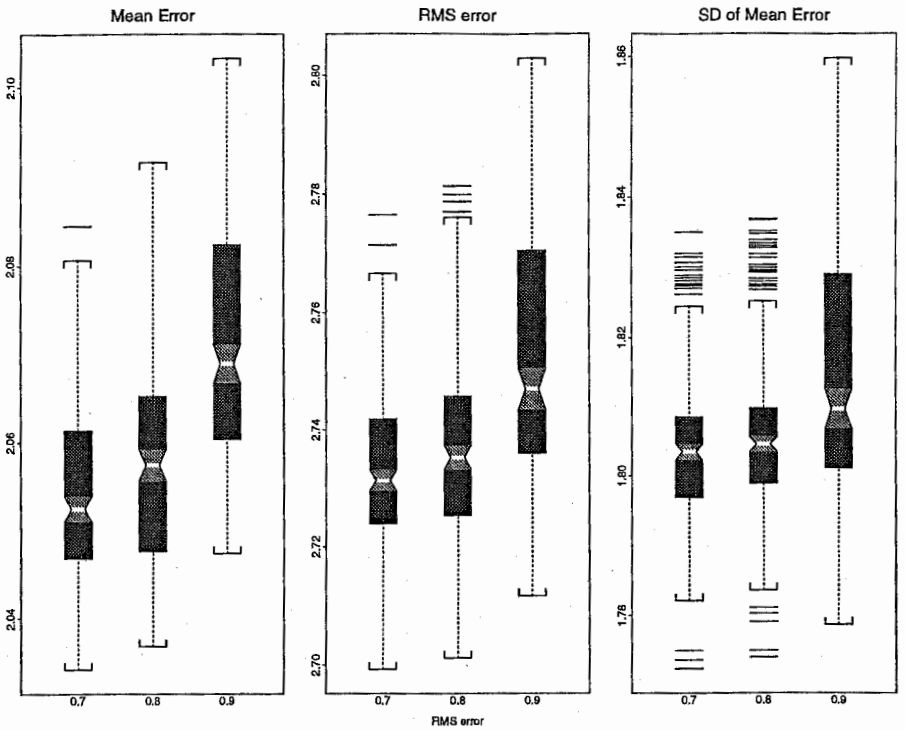

図 3.2: dur_wt によるケプストラム距離の変化

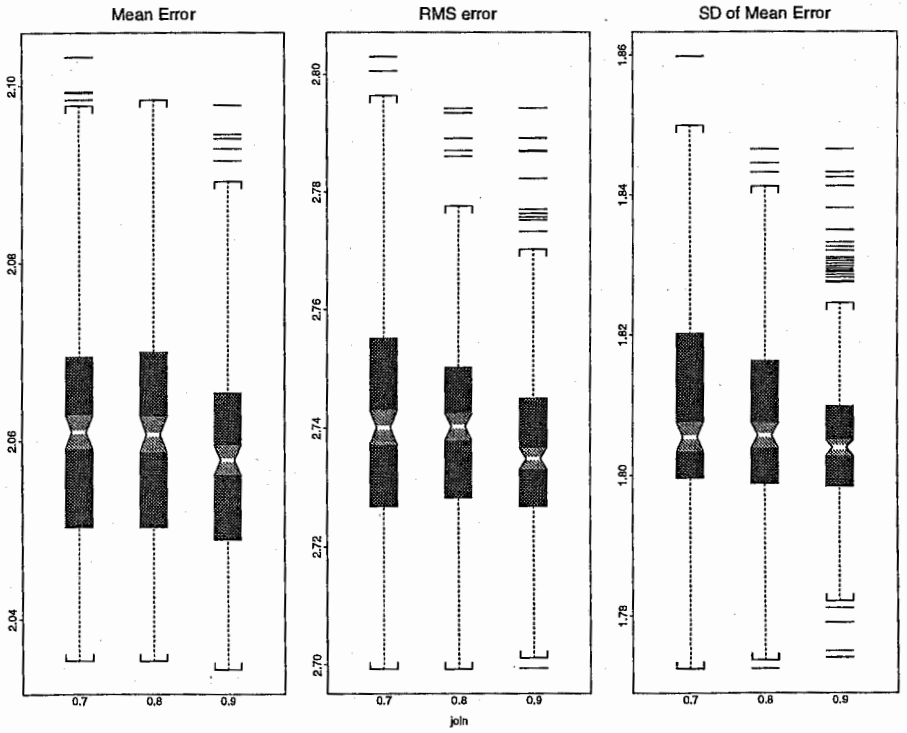

図 3.3: join_wt によるケプストラム距離の変化

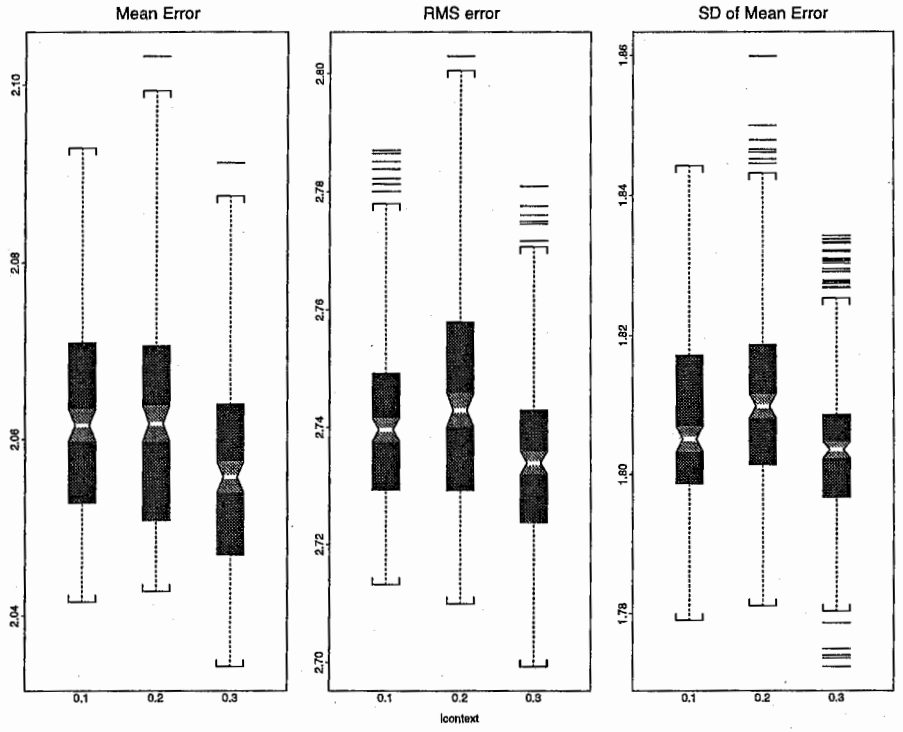

図 3.4: lcontext_wt によるケプストラム距離の変化

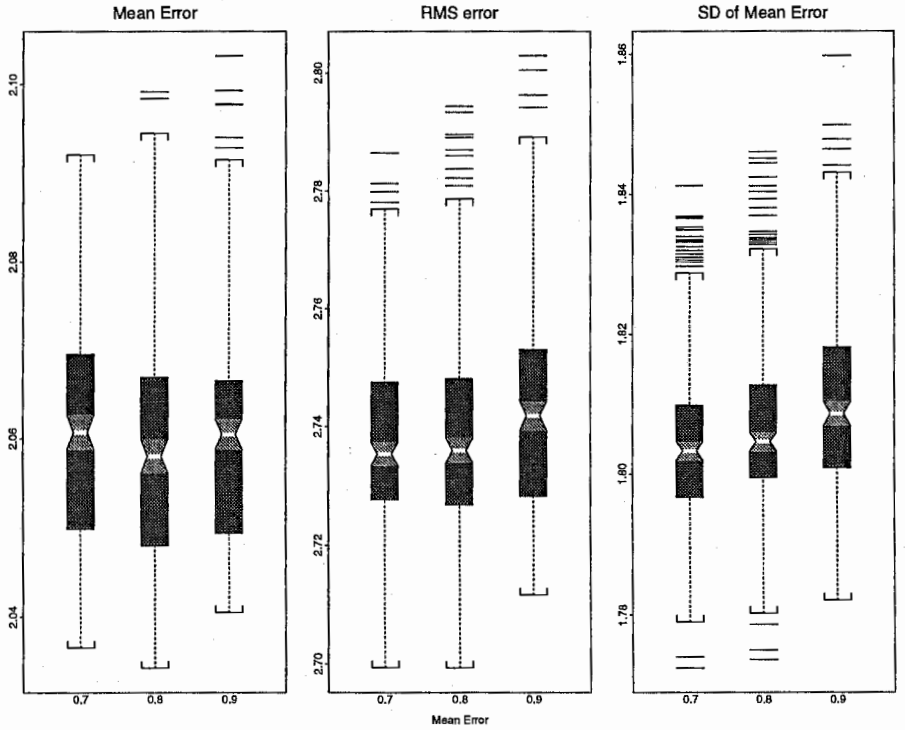

図 3.5: pitch_wt によるケプストラム距離の変化

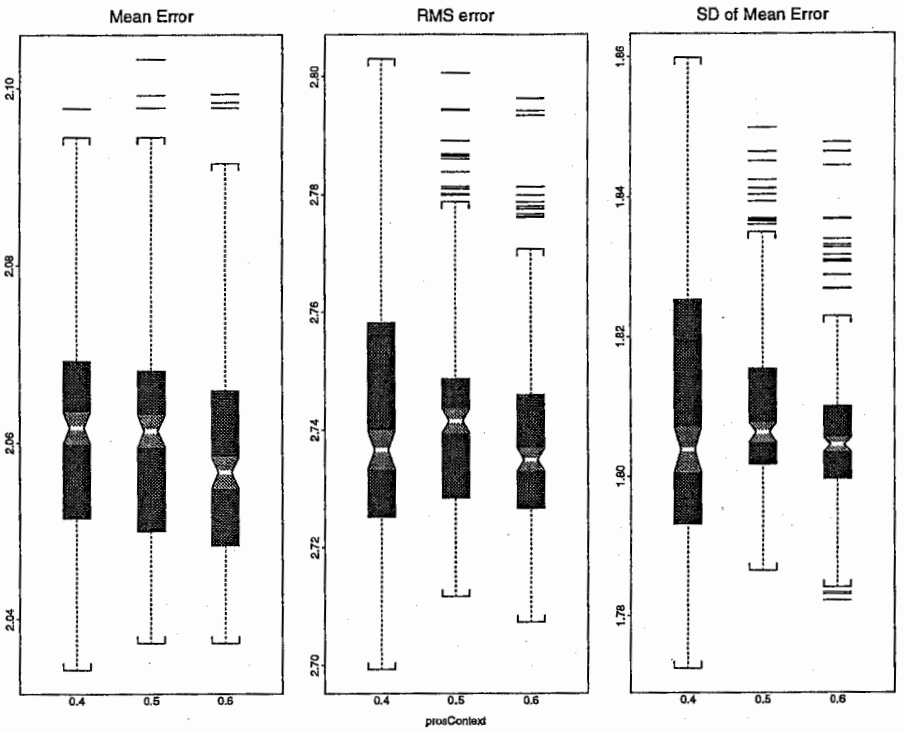

図 3.6: pros_context_wt によるケプストラム距離の変化

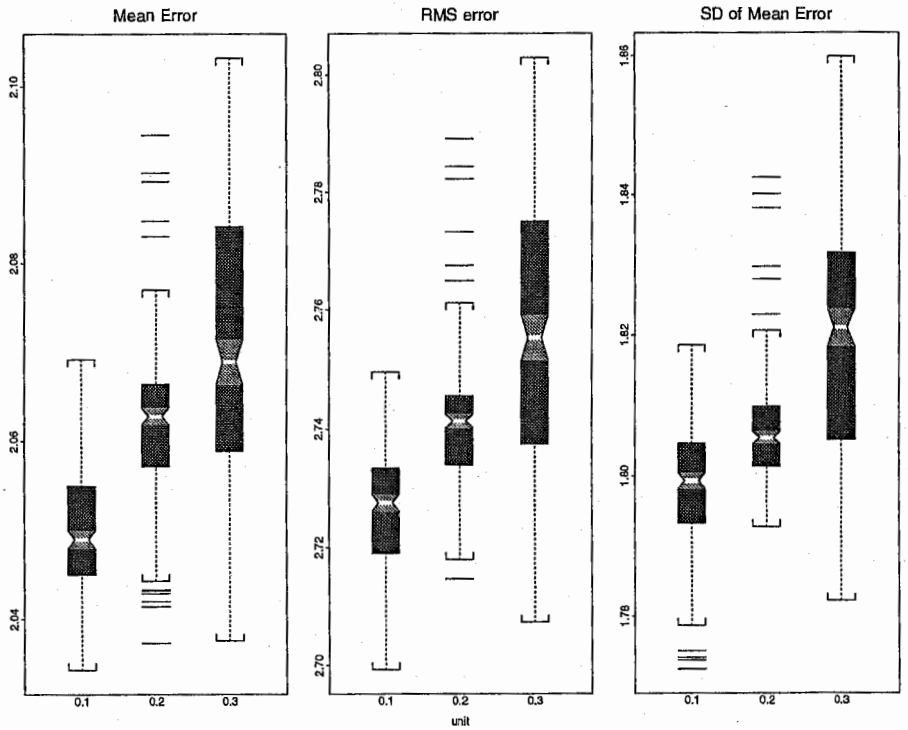

図 3.7: unit_wt によるケプストラム距離の変化

第 4 章

韻律情報の比較

4.1 比較方法

本章では、韻律情報を比較することにより，その類似度を調べた.

CHATR の関数である (Save UdbSelectInfo "-") によって求めた単位選択における予測値と選ばれた音素の平均値の差分から，1 音素あたりの平均値を求める.

前章の場合と同様に，差が小さいほど類似性が高いとみなし，高品質であると判断する.

評価対象の話者データベースとして MHT，評価用の合成音声は前章と同様に MHT_SD_E43 を用いた．検証を目的としているため，評価は 1 文だけに対してのみ行っている.

4.2 nus_params で用いるパラメータの最適値の算出

用いた nus_params の組み合わせは，3.4 節と同様で次の通りである.

```
(join_wt 0.7 0.8 0.9)
(unit_wt 0.1 0.2 0.3 )
(lcontext_wt 0.1 0.2 0.3)
(dur_wt 0.7 0.8 0.9)
(pitch_wt 0.7 0.8 0.9)
(pros_context_wt 0.4 0.5 0.6)
(vq_wt 0.8)
(vq_f0_wt 0.2)
(vq_pow_wt 0.8)
(phone_context_wt 0.2)
```

結果をパラメータ別に図 4.1～図 4.6に示す.

左から順に pitch, duration, power の結果を示す．各図とも，縦軸方向において数値が小さいほど類似度が高いことを示す.

前章と同様に繰り返すことで，最適値を求めることが可能になった.

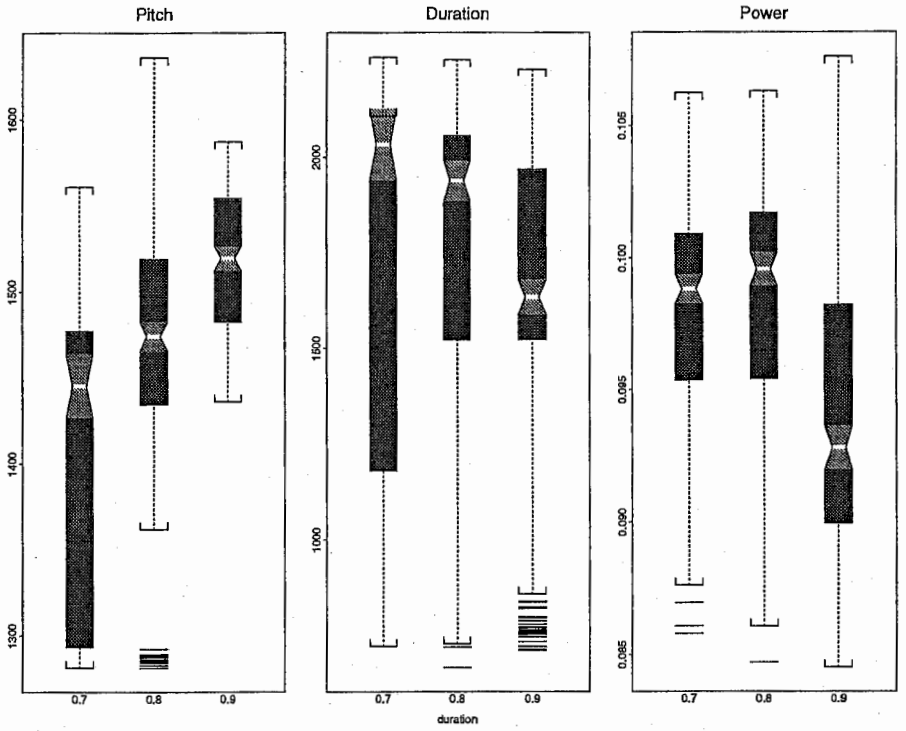

図 4.1: dur_wt による UdbSelectInfo の変化

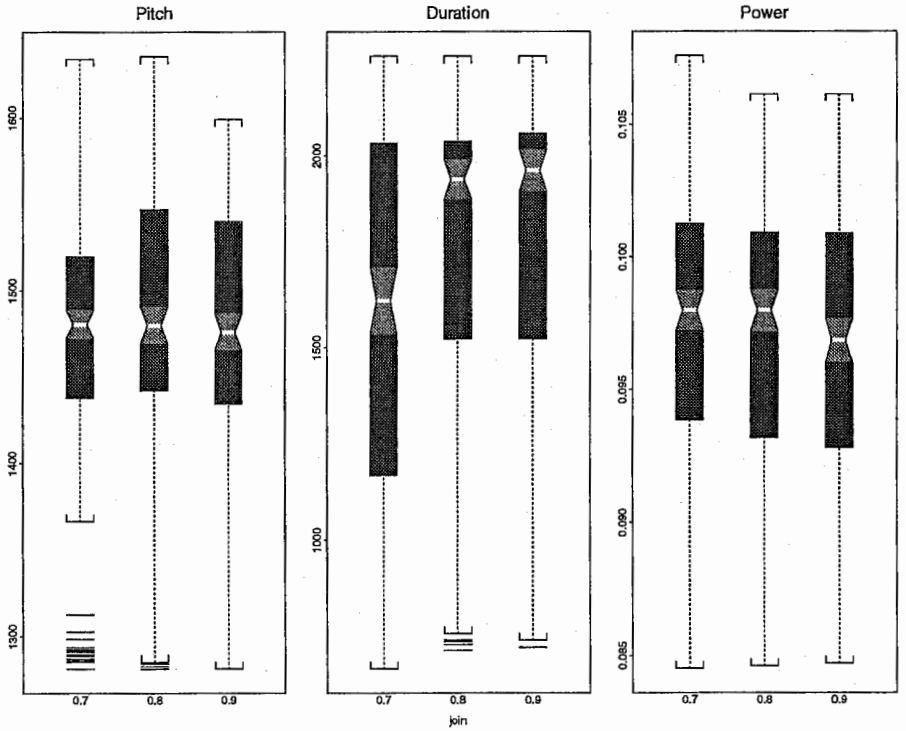

図 4.2: join_wt による UdbSelectInfo の変化

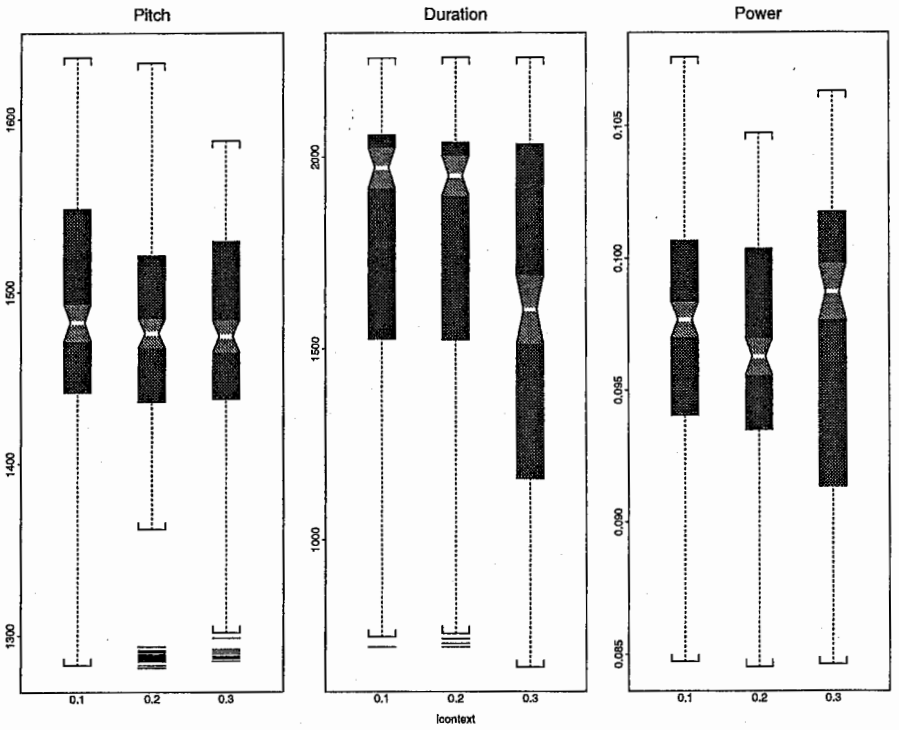

図 4.3: lcontext_wt による UdbSelectInfo の変化

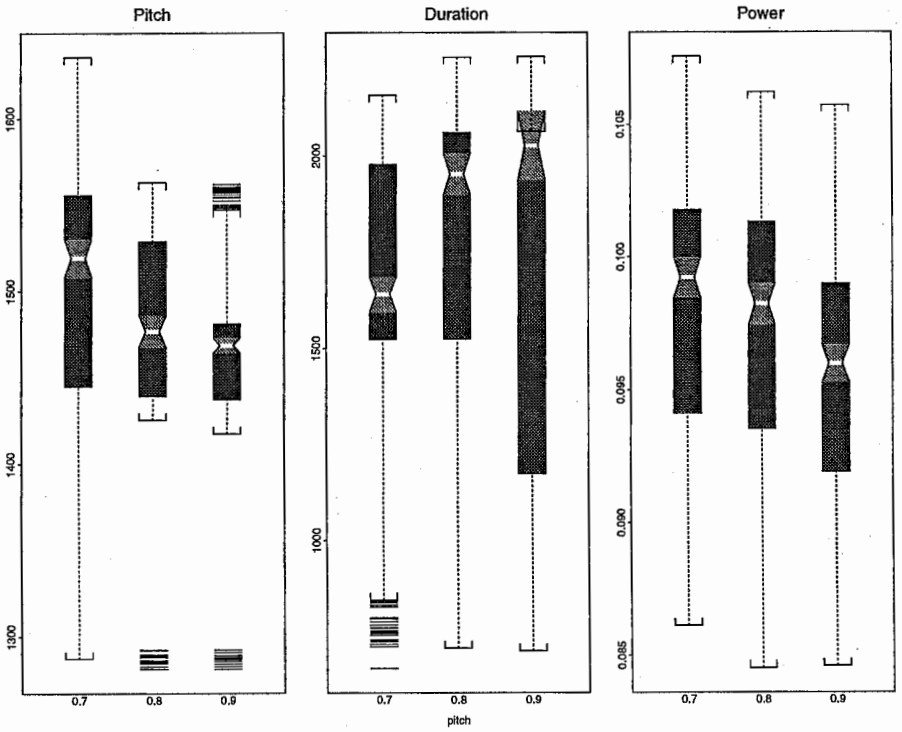

図 4.4: pitch_wt による UdbSelectInfo の変化

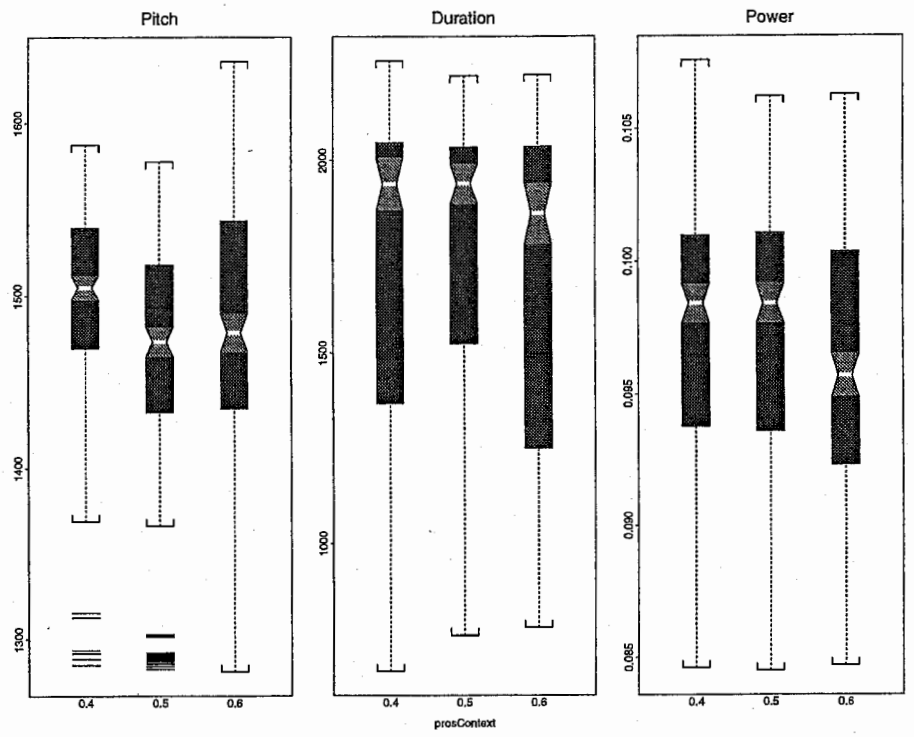

図 4.5: pros_context_wt による UdbSelectInfo の変化

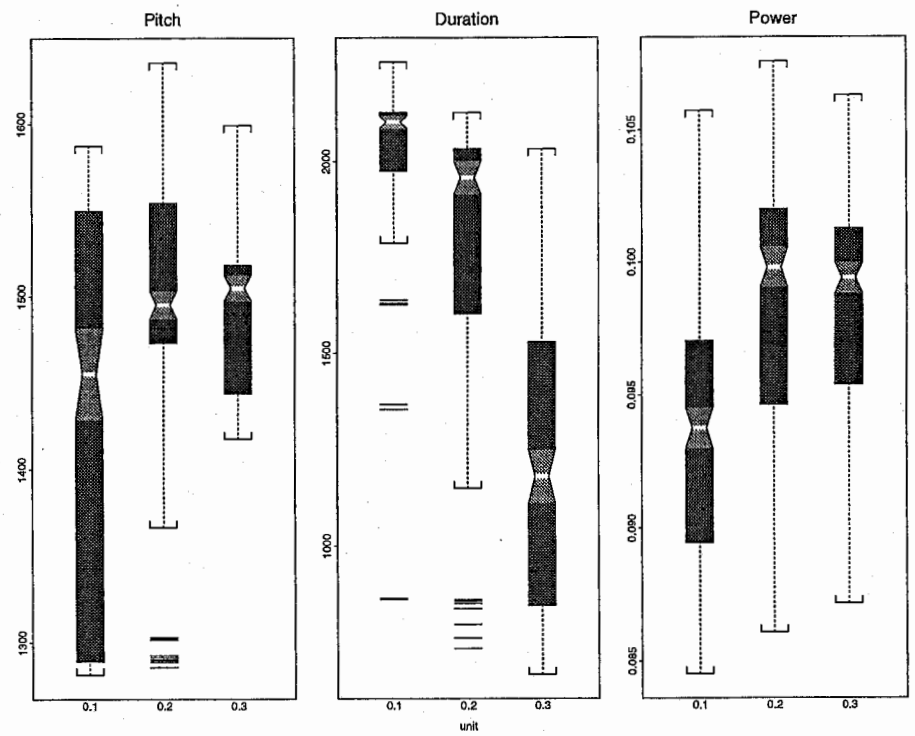

図 4.6: unit_wt による UdbSelectInfo の変化

第 5 章

発話様式の違いによる影響

5.1 発話様式

これまでに求めてきたnus_paramsの値は，ある話者のただ一つの発話様式についてのみ調べてきた．そこで，本章では同一話者の異なる発話様式のデータベースに対して，パラメータの分布を調べ，発話様式による違いをみることにした．

なお、評価対象の話者データベースとしてFIAa, FIAj, FIAsを用いた．これらは、同一の読み上げテキストを使用していないため，それぞれ任意の１０文で音声合成を行っている．

5.2 nus_paramsで用いるパラメータへの影響

用いたnus_paramsの組み合わせは，次の通りである．

```
(join_wt 0.2 0.5 0.8)
(unit_wt 0.2 0.5 0.8 )
(lcontext_wt 0.2 0.5 0.8)
(dur_wt 0.2 0.5 0.8)
(pitch_wt 0.2 0.5 0.8)
(pros_context_wt 0.2 0.5 0.8)
(vq_wt 0.8)
(vq_f0_wt 0.2)
(vq_pow_wt 0.8)
(phone_context_wt 0.2)
```

パラメータ別にまとめた結果を図5.1～図5.6に示す．

図の上段はFIAa，中段はFIAj，下段はFIAsを示す．また，図の左列はPitch，中列はDuration，右列はPowerを示す．各図とも，縦軸方向において数値が小さいほど類似度が高いことを示す．

これらの結果から，発話様式によってパラメータの重み，つまり重要視すべき要素が変わる傾向かみられた．この点については，さらに詳しく調べることによって，発話様式に応じたパラメータの最適値を求めることができると思われる．

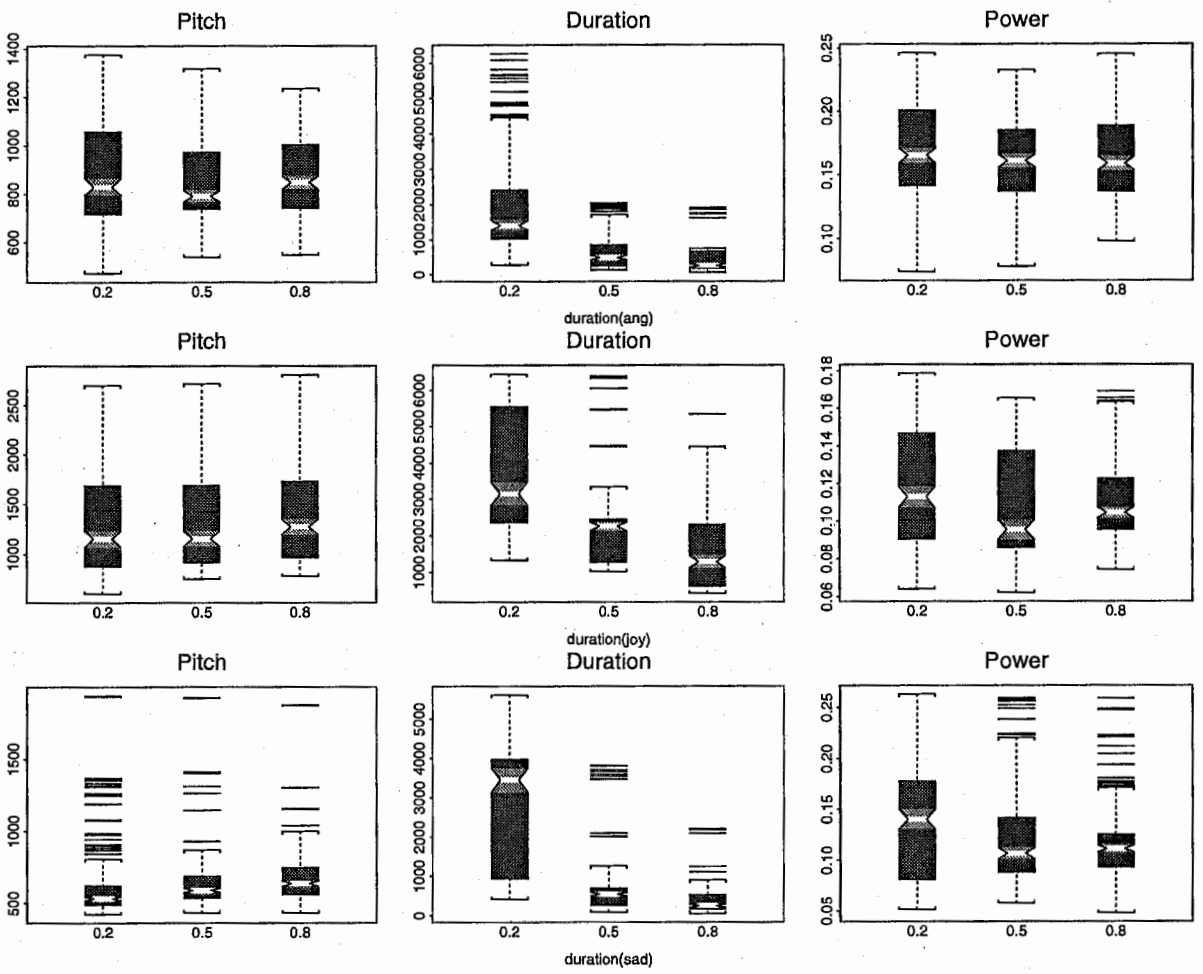

図 5.1: dur_wt による UdbSelectInfo の変化
(上段‥ FIAa, 中段‥ FIAj, 下段‥ FIAs)

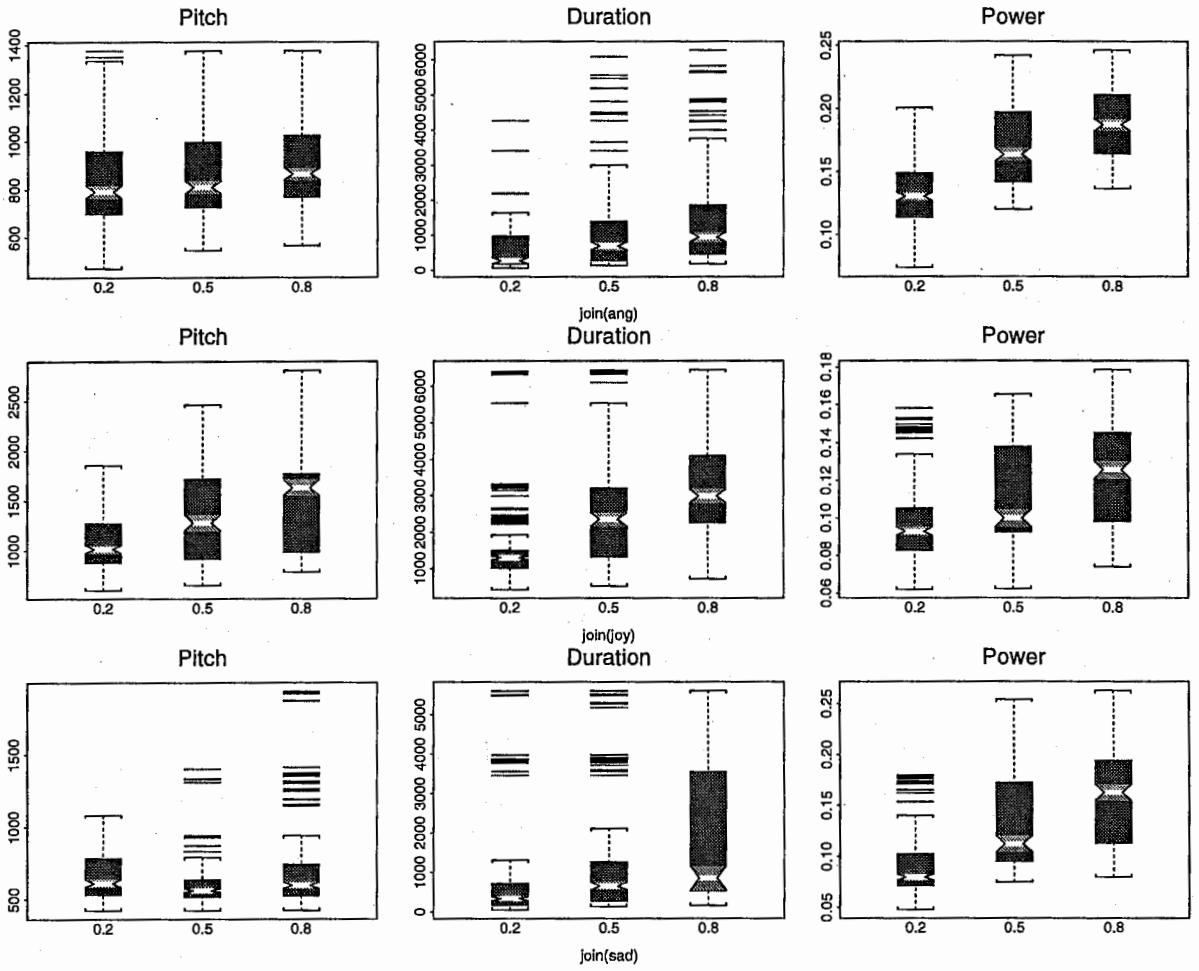

図 5.2: join_wt による UdbSelectInfo の変化
(上段‥ FIAa，中段‥ FIAj，下段‥ FIAs)

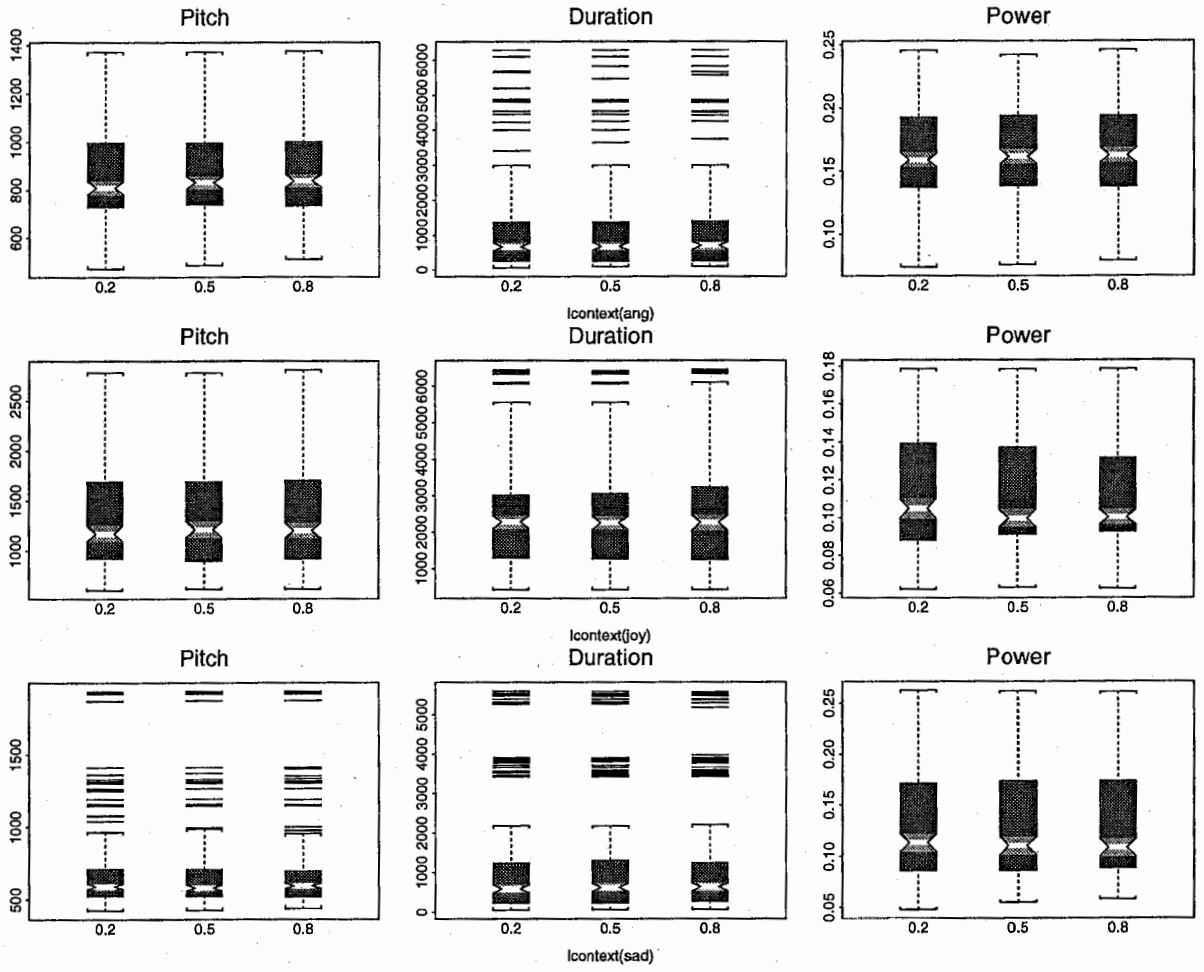

図 5.3: lcontext_wt による UdbSelectInfo の変化
(上段‥ FIAa, 中段‥ FIAj, 下段‥ FIAs)

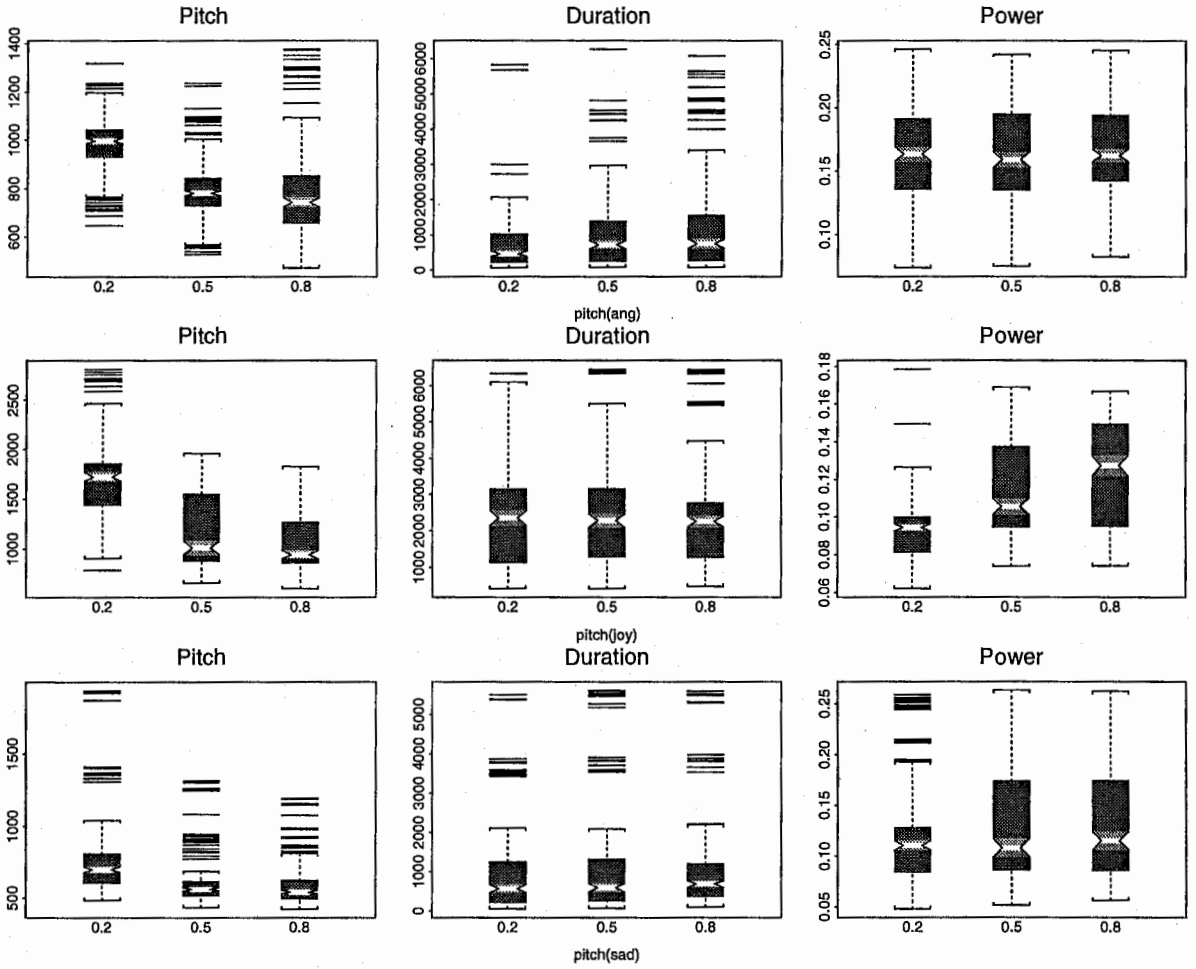

図 5.4: pitch_wt による UdbSelectInfo の変化
(上段‥ FIAa, 中段‥ FIAj, 下段‥ FIAs)

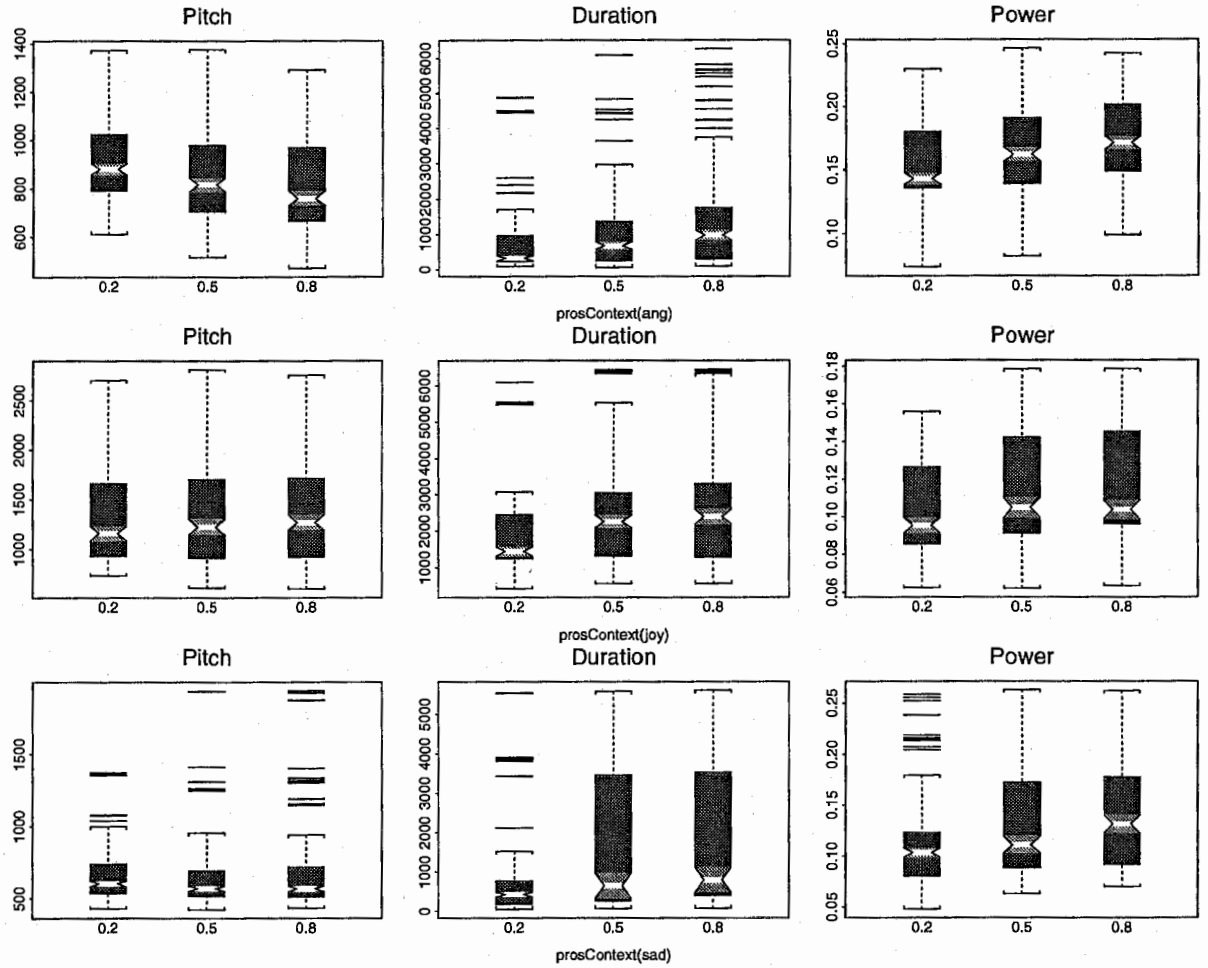

図 5.5: pros_context_wt による UdbSelectInfo の変化
(上段‥ FIAa, 中段‥ FIAj, 下段‥ FIAs)

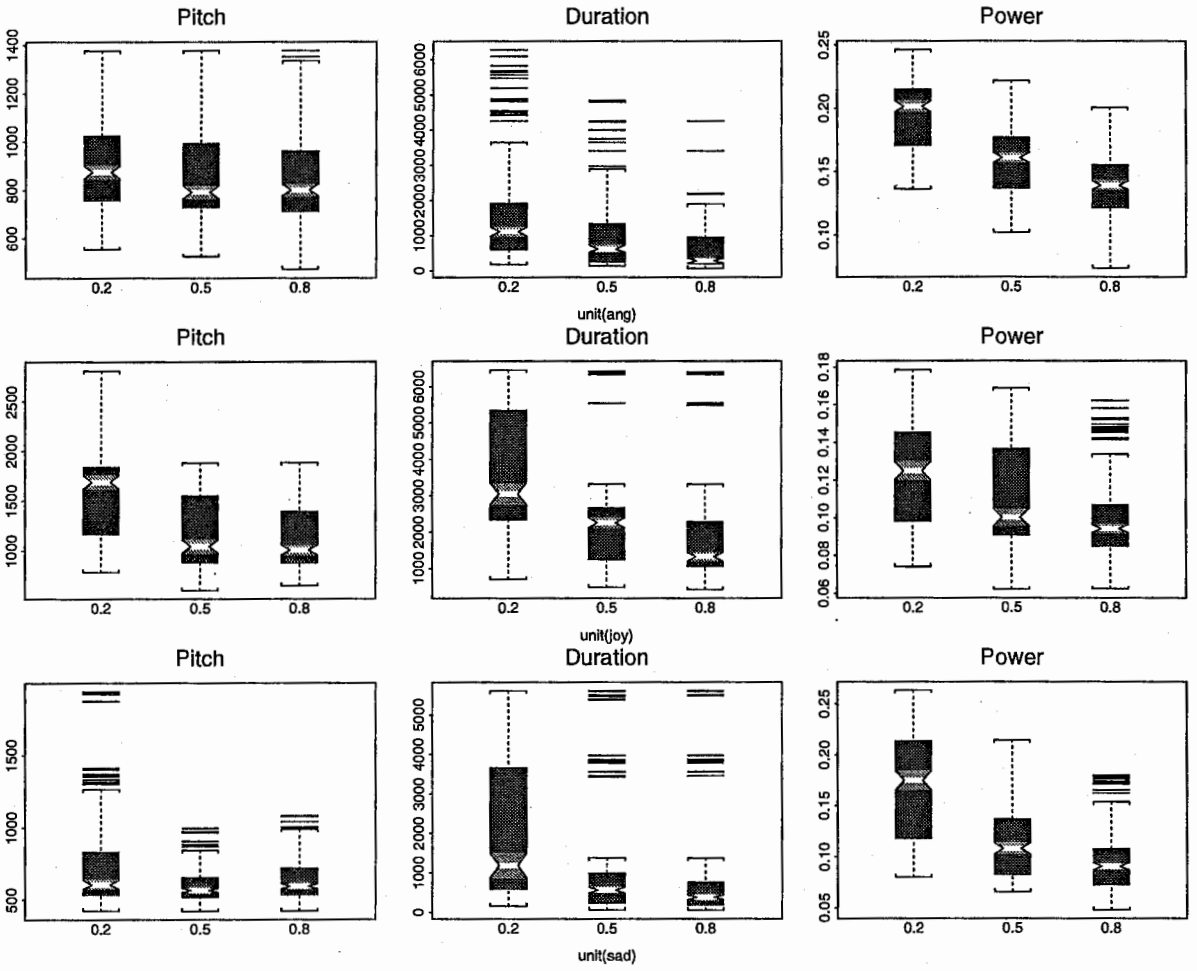

図 5.6: pros_context_wt による UdbSelectInfo の変化
(上段‥ FIAa, 中段‥ FIAj, 下段‥ FIAs)

第 6 章

まとめ

本稿では波形接続型音声合成システム CHATR の最も大きな問題点であるイントネーションの自然性の改良のために，音素単位選択の際に使用するパラメータの最適化について、将来の自動化を視野にいれて，その求め方について示した．

また，データベースだけではなく，発話様式に合わせてパラメータを設定する必要性も明らかにすることができた．

今回，パラメータ設定の自動化を重要視したため，得られたパラメータの組み合わせについての検証は不十分である．また，評価対象を検証を目的としたため 1 文に限定しているので，複数の文による評価が不可欠である．その際，評価に適した文の選定も必要と思われる．

今後，評価方法の決定にあたっては，聴覚実験との相関を調べる必要があると考えられる．

第 7 章

謝辞

ATR 音声翻訳通信研究所での研究の機会を与えてくださった山崎元社長，樋口元室長，出向元の関連部署の皆様に深謝します．研究に関して有益な助言を頂いた，山本社長，キャンベル室長，第二研究室の皆様に感謝します．

参考文献

[1] W.N.Campbell, A.W.Black :, "CHATR: 自然音声波形接続型任意音声合成システム", 信学技報, SP96-7, (1996,5).

[2] J.D.Chen, W.N.Campbell : "Speech Synthesis Evaluation by Objective Distance Measures", 信学技報, SP99-3, (1999,5).

付録 A

使用したプログラム

A.1 CHATR バッチ処理スクリプト，使用法

3.2 節

○スクリプトファイル

/home/as68/satoshi/work/opt-param/MHT/VarCandBeam/
MHT_SD_E43.varCandBeam.ch（内容は 27 ～ 29 ページ）
mkscore.awk

○使用例

```
%/DB/CHATR/$OS/bin/chatr -b MHT_SD_E43.varCandBeam.ch | \
    awk -f ./mkscore.awk |sed 's/[();]//g' > MHT_SD_E43.varCandBeam.sData
```

3.3 節

○スクリプトファイル

/home/as68/satoshi/work/opt-param/MHT/Cepstrum/
MHT_SD_E43.25_25_100.ch
fil.awk

○使用例

```
%/DB/CHATR/$OS/bin/chatr -b MHT_SD_E43.25_25_100.ch | \
    sed 's/[()]//g' |awk -f ./fil.awk > MHT_SD_E43.25_25_100.sData
```

3.4 節

○スクリプトファイル

/home/as68/satoshi/work/opt-param/MHT/Cepstrum/
MHT_SD_E43.25_25_100.ver2.ch

○使用例

```
%/DB/CHATR/$OS/bin/chatr -b MHT_SD_E43.25_25_100.ver2.ch | \
    sed 's/[()]//g' |awk -f ./fil.awk > MHT_SD_E43.25_25_100.sData
```

4.2 節

○スクリプトファイル

```
/home/as68/satoshi/work/opt-param/MHT/UdbSelectInfo/
    MHT_SD_E43.25_25_100.ver2.selData.ch
```

○使用例

```
%/DB/CHATR/$OS/bin/chatr -b MHT_SD_E43.25_25_100.ver2.selData.ch | \
    sed 's/[()]//g' |awk -f ./fil.awk > MHT_SD_E43.25_25_100.ver2.sData
```

5.2 節

○スクリプトファイル名

```
/home/as68/satoshi/work/opt-param/FIA/mksData.sh
```

○使用方法

```
%mksData.sh
```

MHT_SD_E43.varCandBeam.ch の内容 (29 ページまで)

```
    (speaker_MHT)

    (Parameter Concat_Method NONE)
    (Database Set Strategy Generic)

;;;  For long searches need to be aggressive with garbage collection
    (gc_info 'EnoughRope)

   (set beam_widths
     '(
       (beam_width 25 50 100 200 500)
       (cand_width 25 50 100 200 500)
       (cand_max 25 50 100 200 500)
       )
    )
;;;  So that cepstrum resynthesis doesn't happen (not the cleanest way)
    (set cep_no_cep_synth 't)
    (set cep_use_cache 't)  ;; load cep files and keep them loaded

;;;  Ensure no signal processing occurs
    (Parameter Concat_Method DUMB+)
    (set power_modify nil)
    (set synth_hook nil)

;;;  The desired objective distance measure should be defined
;;;  for the speaker in its *_synth.ch file  but just in case it
;;;  isn't we ensure it has a value
    (defvar cep_dist_params nil)
    (set comp_parms (cons '(align_type tw) cep_dist_params))

;;;  This defines the weights you wish to set and their various values
;;;  note that the bigger this is the longer the search will take
   (set weight_table
     '(
       (join_wt 1.0)
       (unit_wt 1.0)
       (lcontext_wt 1.0)
       (cand_thresh 1.0)
       (vq_wt 0.9)
       (vq_f0_wt 1.5)
       (vq_pow_wt 1.0)
       (phone_context_wt 0.0)
       (pros_context_wt 0.0)
       (dur_penalty 1.0)
       (endpoint_weight 0.0)
       )
   )
```

```
   (define the_weights (g)
       (copy (car (cdr g))))

   (define test_score (name weights)
   "Tests a particular set of weights with an utterance."
       (set nus_params
           (append (copy weights)
               (cons
                (list (copy 'exclude_list)
                      (copy name))
                (copy weight_table))))
      (Synth utt_seg)
      (set wv (reverse (mapc the_weights weights)))
;
      (print wv "-")
      (Save UdbSelectInfo "-")
;
      (set wv nil)
      (free_val 'nus_params)
  't)

   (define mk_pairs (n v)
      (list (copy n) (copy v)))

  (define do_all (name)
      (Parameter Concat_Method NUUCEP)
      (Database Set Strategy Generic)
      (set utt_seg (load_segs name))
      (set cep_no_cep_synth 't)
      (do_all_wts name beam_widths nil))

  (define do_all_wts (name weight_vals weight)
      (let ((w nil))
          (if (not weight_vals)
          (prog (test_score name weight))
          (for (set w (cdr (car weight_vals))) w (set w (cdr w))
              (let ((nw (cons (copy (car (car weight_vals)))
              (cons (copy (car w)) nil))))


      (let ((nws (cons nw (copy weight))))
          (do_all_wts name (cdr weight_vals) nws)
          (free_val 'nws)))))))

;;; Function for listening to "good" weightings
  (define test_weights (name w)
"Call with a fileid and a list of weights, will synth the utterance with
 dumb+ concatenation in utt_seg."
      (let ((weights (mapc mk_pairs weight_names w)))
      (Parameter Concat_Method NONE)
```

```
      (set cep_no_cep_synth 't)
      (set nus_params
         (append (append (copy weights) (copy beam_widths))
                 (list (list (copy 'exclude_list) (copy name)))))
      (load_segs name)
      (Synth)
      (Save UdbSelectInfo "-")
      (Parameter Concat_Method DUMB+)
      (load_segs name)
      (free_val 'nus_params)
      (free_val 'weights)
  'ok))

  (define cep_it (name w)
      (let ((weights (mapc mk_pairs weight_names w)))
      (Parameter Concat_Method NUUCEP)
      (set cep_no_cep_synth nil)
      (set nus_params
         (append (append (copy weights) (copy beam_widths))
                 (list (list (copy 'exclude_list) (copy name)))))
      (load_segs name)
      (Save UdbSelectInfo "-")
      (free_val 'nus_params)
      (free_val 'weights)
  'ok))

  (define load_segs (name)
"Load a segment file (natural utterance description and synth it"
    (let ((fullname (strcat db_data_dir seg_dir name ".seg.ch")))
    (set utt_seg (load fullname))
    (free fullname)
    (Synth utt_seg)))


;;;  For batch running uncomment the next line (need better way to do this)
  (do_all "MHT_SD_E43")

'ok
```

mkscore.awk の内容

```
BEGIN{i=1;Str=$0;}
{
if (NF==7) {a=$3-$2; b=$5-$4; c=$7-$6;
     A+=a*a; B+=b*b; C+=c*c;
     i++;
     }
else {printf("%s %f %f %f\n", Str, A/i, B/i, C/i);
            A=B=C=i=0;
            Str = $0;}
}
END{printf("%s %f %f %f\n", Str, A/i, B/i, C/i)}
```

fil.awk の内容

```
{
if (NF==3)
    printf("%s",$0);
else
    printf("%s\n",$0);
}
```

A.2 得られたデータ、図

3.2 節

○データファイル

/home/as68/satoshi/work/opt-param/MHT/VarCandBeam/
MHT_SD_E43.varCandBeam.sData

○図

/home/as68/satoshi/work/opt-param/MHT/VarCandBeam/
MHT_SD_E43.varCandBeam.sData.S.ps

3.3 節

○データファイル

/home/as68/satoshi/work/opt-param/MHT/Cepstrum/
MHT_SD_E43.25_25_100.sData

3.4 節

○データファイル

/home/as68/satoshi/work/opt-param/MHT/Cepstrum/
MHT_SD_E43.25_25_100.ver2.sData

○図

/home/as68/satoshi/work/opt-param/MHT/Cepstrum/
mhtSdE43bw25cw25cm100ver2_dur.S.V.ps
mhtSdE43bw25cw25cm100ver2_join.S.V.ps
mhtSdE43bw25cw25cm100ver2_lcontext.S.V.ps
mhtSdE43bw25cw25cm100ver2_pitch.S.V.ps
mhtSdE43bw25cw25cm100ver2_prosContext.S.V.ps
mhtSdE43bw25cw25cm100ver2_unit.S.V.ps

4.2 節

○データファイル

/home/as68/satoshi/work/opt-param/MHT/UdbSelectInfo/
MHT_SD_E43.25_25_100.ver2.selData.sData

○図

/home/as68/satoshi/work/opt-param/MHT/UdbSelectInfo/
MHT_SD_E43.25_25_100.ver2.dur.S.ps
MHT_SD_E43.25_25_100.ver2.join.S.ps
MHT_SD_E43.25_25_100.ver2.lcontext.S.ps
MHT_SD_E43.25_25_100.ver2.pitch.S.ps
MHT_SD_E43.25_25_100.ver2.prosContext.S.ps
MHT_SD_E43.25_25_100.ver2.unit.S.ps

5.2 節

○データファイル

/home/as68/satoshi/work/opt-param/FIA/FIAajs/

ang*.selData.sData

joy*.selData.sData

sad*.selData.sData

○図

/home/as68/satoshi/work/opt-param/FIA/FIAajs/

PS/ang*_dur.S.ps

PS/ang*_join.S.ps

PS/ang*_lcontext.S.ps

PS/ang*_pitch.S.ps

PS/ang*_prosContext.S.ps

PS/ang*_unit.S.ps

PS/joy*_dur.S.ps

PS/joy*_join.S.ps

PS/joy*_lcontext.S.ps

PS/joy*_pitch.S.ps

PS/joy*_prosContext.S.ps

PS/joy*_unit.S.ps

PS/sad*_dur.S.ps

PS/sad*_join.S.ps

PS/sad*_lcontext.S.ps

PS/sad*_pitch.S.ps

PS/sad*_prosContext.S.ps

PS/sad*_unit.S.ps