言語音声における超分節素の特性制御と知覚実験のためのツール

佐藤 大和
東京外国語大学 ｱｼﾞｱ・ｱﾌﾘｶ言語文化研究所

益子 幸江
東京外国大学 大学院総合国際学研究院

言語音声におけるアクセント，声調，イントネーションなどの超分節的特性を研究するためのソフトウェア・ツールを作成した．このツールは，発話音声を音響分析し，そのピッチ周波数や音韻持続時間特性を変更あるいは新たに生成して，その音声を再合成するツールと，得られた合成音を使って聴知覚実験を実施するツールから成る．両者とも，表計算ソフト（エクセル）を利用したインタフェースにするなど，利用し易いよう工夫されている．音声合成は，１ピッチ波形を指定されたピッチ周期で重ね合わせる手法と，音声スペクトルの調波構造を正弦波重畳モデルによって実現する手法の二つの方法が利用可能となっている．

A software tool for supra-segmental characteristic control and perceptual experiment in spoken languages

Hirokazu Sato
Research Institute for Languages and Cultures of Asia and Africa

Yukie Masuko
Faculty of the Graduate School of Global Studies

Tokyo University of Foreign Studies

A software tool for studies of supra-segmental characteristics including accent, tone and intonation in spoken language is proposed. The tool is composed of two functional components. One is speech resynthesis based on edited or newly produced pitch frequency and durational characteristics, where both pitch-synchronous overlap method of 1-picth waveforms and sinusoidal wave superposition method to create harmonic structure of speech spectrum are available. The other is conduct of auditory experiments using synthesized speech. These are designed to realize easy-to-use interface by applying widely used spread sheet software (Microsoft Excel).

１．はじめに

音声および音声言語の研究は，英語や日本語など使用人口が多い世界の主要言語においては，膨大なコーパスの構築が進み，これに基づく研究が進展している．また一方，少数言語や方言などは，フィールド調査によってあらたな発見などがあり，その保存や記述の努力がなされている．これらの研究は，どちらかというと語彙，文法の研究や，音声言語においては音韻の分節的側面からの研究が主体であり，アクセントや声調，あるいはイントネーションなどの超分節的特徴に関する研究は十分なされてきたとは言いがたい．

我々は，日本語アクセントやタイ語，ビルマ語，ベトナム語など東南アジア諸言語の声調に関して，音声学的，音響学的分析研究を進めてきた．これらの研究を通じて明らかになってきたことは，ピッチ周波数曲線（以後ピッチ曲線と呼ぶ）等の音響分析特性を見ただけでは，アクセント素や声調素といった超分節素との対応関係を明らかにするには限界があるということである．ピッチ曲線などの特性が超分節素の知覚に及ぼす影響や役割を調べることによってはじめて“言葉”と音響特性との関連を明らかにすることができる．

上記の点から，アクセントや声調，イントネーションなど，言語音声の超分節特性の研究に資することを目的として，音声のピッチ曲線や持続時間を変更／生成したデータに基づき，新たな音声を再合成するとともに，これらの音声を使用して聞き取り実験を行う韻律制御実験ソフトウェア・ツール（Spitツール）を作成した．本報告はその内容に関して述べたものである．

２．ねらい

言語音声の分析ばかりでなく，音声を再合成できるソフトウェア・ツールとして praat [1]が広く利用されている．このソフトでは，分析されたピッチ周波数や持続時間の変更は，原則としてグラフィッ

ク・インタフェースでなされる．そのため，ピッチや持続時間の変更に手間がかかるという使いにくさがあった．また，ピッチ変形に伴う音声合成の手法も１ピッチ波形の重畳方式のみであった．

Spit ツールでは，以下の点に留意して作成された．

（１）スプレッドシート　インタフェース

数値や関数でデータ更新を可能とするため，音声分析結果は一定周期のフレームデータとして表計算ソフト：エクセルに表示するインタフェースとした．

（２）2 方式の音声合成

１ピッチ波形の重畳・加算による合成方式に加えて，スペクトルの調波構造を実現する音声合成方式の２方式の選択を可能とした．

（３）聴取実験のエクセル・インプリメント

音声の聴取実験も，使いやすいツールとするため，表計算ソフト：エクセル上で実施できる仕組みとした．

本ツールの研究上の適用領域は，以下に示すような項目の研究を想定している．

・日本語などアクセント言語におけるアクセント知覚
・中国語や東南アジア諸言語の声調知覚
・文構造とイントネーション
・疑問，命令，意思，強調，対人モダリティ等に伴うイントネーション
・感情音声

などである．

３．Spit ツールの概要

Spit ツールは，韻律制御エディタ（SpitEditor）と刺激音提示・集計ツール（SpitPlayer）から構成されている．SpitEditor は，音声のピッチ周波数と持続時間を変形して，種々の超分節特性を有する音声を合成するソフトウェア・ツールであり，SpitPlayer は，こうして作られた合成音声や自然音声を用いて，種々の聴取実験を実施するとともに，その結果を集計するソフトウェア・ツールである．（現在，これらのツールは，Microsoft Windows 7 で動作している．）

４．音声合成の手法

Spit ツールの内容に関して説明する前に，SpitEditor において変更されたピッチ周波数や持続時間に基づいて音声を再合成する二つの方法について述べる．通常，音声合成としては，ホルマント合成や線形予測分析合成（LPC）などのように，声帯振動の音源と声道の伝達特性を分離した“音源－声道モデル”が用いられてきたが，聴知覚実験に用いるには音声品質上の問題があるため，以下，音声の波形領域での合成とスペクトル構造の合成の二通りの方式を利用することとした．

4.1 ピッチ同期重畳加算方式（PSOLA）

PSOLA とは，Pitch Synchronous Overlap and Add の略であり，E. Moulines らによってそれまでの“音源－声道”モデルにかわって，波形を直接処理することによって合成を可能にした方法である[2,3]．

PSOLA では，1 ピッチ波形の最大値，もしくはこの波形の最初の零交差点（ピッチマークと呼ぶ）を中心として，窓関数 W(n) によって２ピッチ分の時間長（N サンプル）で切り出し，得られた音声波形要素を，順次新たなピッチ周期に合わせてずらしつつ重畳して音声を再合成する手法である．窓関数としては以下のハニング窓が使用される．

$$W(n) = 0.5 - 0.5\cos(2\pi n/(N-1)), \quad (n=0 \sim N-1)$$

この方法では，ピッチ周波数の変更が現音声と大きく異ならない限り，原音声の音質に近い音声が得られる．

4.2 正弦波重畳方式（SWS）

SWS は，Sinusoidal Wave Superposition の略であり，音声スペクトルの調波構造を正弦波の重畳で実現することによって，所望の音声を得る方法である．PSOLA の場合と同様，零交差点間隔から得られたピッチ波形を FFT にて周波数領域に変換するとともに，線形予測分析によりスペクトル包絡を求める．音声の基本周波数成分とその倍音成分を，正弦波として表現し，それらを重畳して目的音声を合成する．波形のピッチ周期の境界では，位相が連続するよう接続される．

SWS では，音声の調波構造を再現するため原音声とはやや音質が異なり，クリアな音質の合成音が得られるが，PSOLA におけるような波形の重畳による音声の歪がないため，ピッチ周波数の変更に伴なう音質の変化は少ないと思われる．実験目的に応じて両者を使い分けることが望ましい．図１－１，図１－２に，音声合成の 2 方法の処理の概要を示す．

５．SpitEditor の内容

SpitEditor は，言語音声の超分節的特徴に及ぼすピッチ周波数や時間情報の役割を検討するためのツールであり，元となる音声からピッチ周波数を抽出し，その周波数や時間長情報の変更を行って，そのデータに基づいて新たな音声を合成するものである．その機能と概要は以下のとおりである．

（１）音声波形窓の表示
（２）音声波形のピッチマークの自動設定と手動修正
（３）有声音部，無声音部，無音部等のラベリング
（４）ピッチ周波数，持続時間データの表示と編集
（５）編集データに基づく音声合成
（６）音声合成データの確認と保存

以下これらの内容に関して述べる．

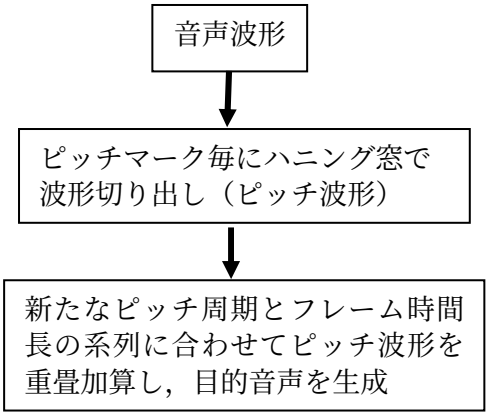


図１－１　PSOLA 法による音声合成

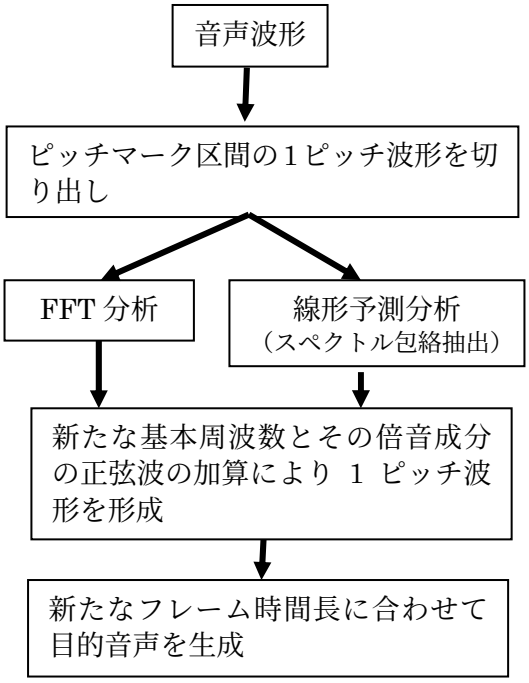


図１－２　SWS 法による音声合成

5.1 音声波形の零交差点の検出と基本周波数

SpitEditor を起動すると，作業用ウィンドウが表示される．画面は，「各種メニュー表示」「音声波形領域」「時間表示領域」「ピッチ周波数領域」「ラベリング領域」および「SWS，PSOLA 合成のためのピッチ周波数と時間情報領域」から成っている．

必要な音声ファイルを読み込み，波形を表示させたのち，ピッチ周波数を分析する．ピッチ周波数は，自己相関法などの手法を直接適用するのではなく，音声波形上の周期から基本周波数を求める手法を採用した．安定的に基本周期を求めるため，波形の零交差位置（ピッチマーク）を探索し，その周期から基本周波数を求める．最初の零交差位置が求まると，自己相関法により次のピッチマークの位置を予測し，その近傍で零交差位置を検出する．これを繰り返し実施することにより，連続的にピッチマークが設定される．一定周期ごとの平均ピッチ周波数を求めたい場合は，一定窓幅におけるピッチマーク周期の平均値から求める．

5.2 ピッチマークの自動設定と修正

SpitEditor 上でのピッチマークの自動設定は，その範囲をマウスのドラッグで定めたのち，「音声」メニューから「ピッチマークの自動設定」を選択すると，波形の零交差位置にピッチマークが表示される．零交差位置は（負→正）に変わる零交差を default としているが，（正→負）などに変更可能である．

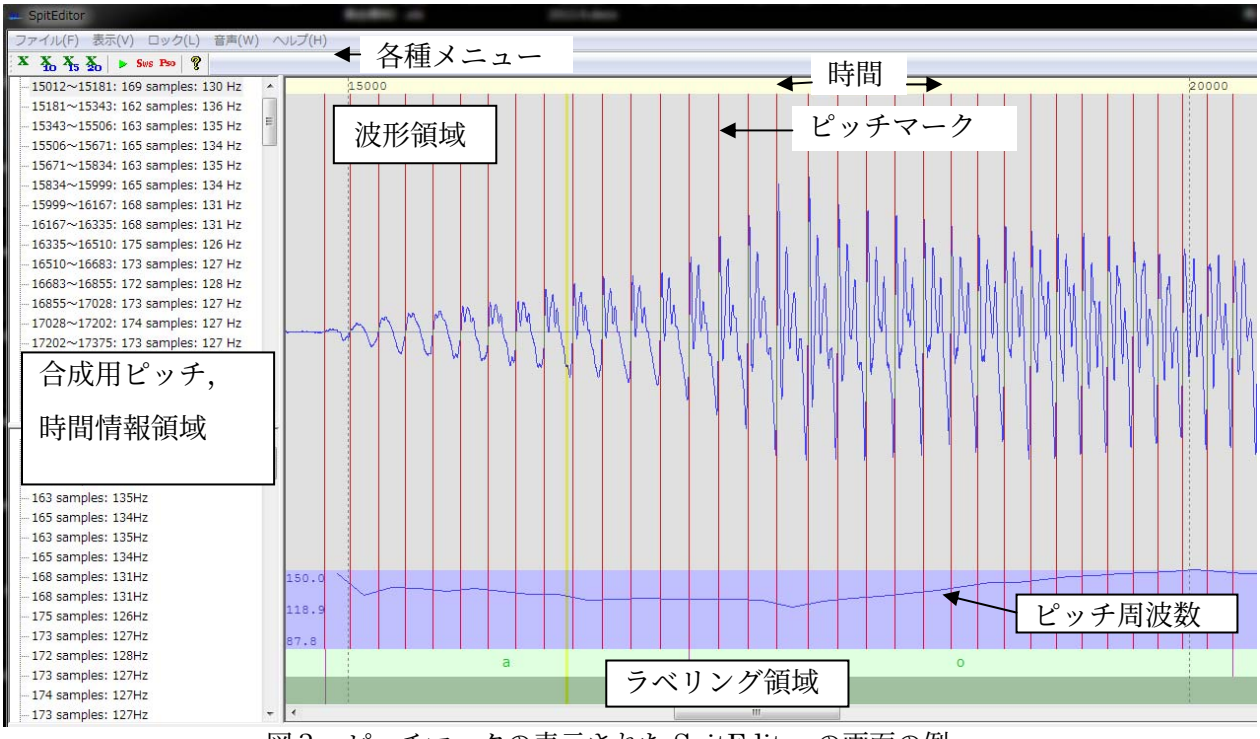


図２　ピッチマークの表示された SpitEditor の画面の例

ピッチマークが設定されると，「ピッチ周波数領域」にピッチ周波数パタンが表示され，また「合成用ピッチ，時間情報領域」にも自動的に情報が書き込まれる．ピッチマークが表示された SpitEditor の画面例を図２に示す．

なお，ピッチマークの設定に当たっては，その探索のため男声や女声などによって，あらかじめピッチ探索の範囲を定めておくと抽出精度がよくなる．「音声」メニューの中の「F0 推定範囲設定」を選択し，ピッチ周波数の上限値と下限値を設定する．

ピッチマークの自動設定は，必ずしもいつも正しい位置に設定されるとは限らない．特に，文末などにおいて著しくピッチが降下した場合などで設定誤りが起こりやすい．そのため，いったん抽出されたピッチマークの位置を修正したり，追加や削除ができるようになっている．

・ピッチマークの移動・修正：マウスのドラッグ
・　　〃　　追加：(Ctrl + 左クリック)
・　　〃　　削除：(Ctrl + 右クリック)

ピッチマークが設定されると，ピッチ周波数と時間情報が得られるので，原音声を二つの合成方式で合成復元することが可能になる．画面上部にある三つのボタン (▸)，(Sws)，(Pso) をクリックすると，それぞれ原音声，SWS 方式合成音，PSOLA 方式合成音による復元音声を聞くことができる．このボタンは，ピッチ周波数などを変更した合成音を聞くときにも利用される．

5.3 ラベリング

SpitEditor 作業画面の下部にラベリング領域があり，(領域１) と (領域２) の２層のラベリング層から成っている．ラベリングの手順は以下のとおりである．

（１）ラベリング境界線の挿入

文や語の境界，必要な音韻境界を区分化する線を挿入する．境界線が定まると，線と線の間でラベルを書き込むことができる．

（２）ラベリング領域１（音素等ラベル）

領域１は，音素，音節など任意のラベルをキーボードから入力することができる．この領域の記号は，音声には利用しない．

（３）ラベリング領域２（セグメントラベル）

領域２は，音声セグメントの素性を入力する領域である．次の２項目は必須項目であり，必ずその記号で入力しなければならない．この２記号は，音声合成の際に利用するからである．

V　：Voiced を表す．ピッチマークが付き，ピッチ可変の音声合成が可能な区間
Si　：無音区間を表す

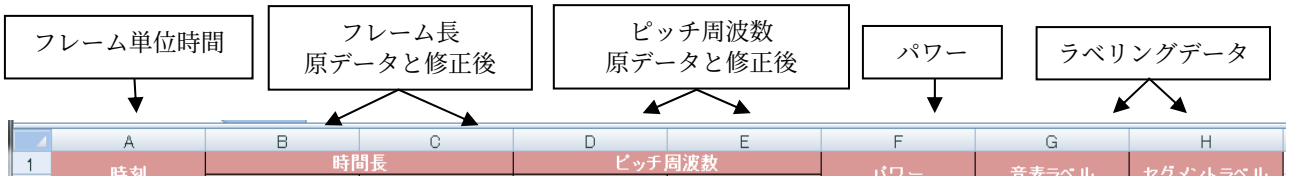
後で示すように，この２種の音声区分では，持続時間長を変えることができる．

無声摩擦音など，その他の音声セグメントに関しては，任意の記号を使ってかまわない．上記２種以外のラベリング区間は，音声合成に関して原音声の波形がそのまま利用されることになる．

5.4 韻律データの編集と音声合成

ピッチマークが設定され，またラベリング・データが指定されると，ピッチ周波数や音声セグメントの持続時間を変更して，新たな合成音を作成することが可能となる．

画面の左上部にある４つのボタン，(X)，(X10)，(X15)，(X20) は，ピッチ周波数や時間，パワーなどの情報をエクセル・シート上に表示させるボタンである．(X) の場合は，ピッチ周期ごとのこれらの情報が表示され，他はそれぞれ 10ms，15ms，20ms の３種類のフレーム周期に相当する音声データがシート上に表示される．図３にフレーム周期 10ms の場合の例を示す．



	A	B	C	D	E	F	G	H
1	時刻	時間長		ピッチ周波数		パワー	音素ラベル	セグメントラベル
2		編集前	編集後	編集前	編集後			
3	0	10	13	131.780	105.424	1605.91	a	v
4	10	10	13	135.695	108.556	2410.10	a	v
5	20	10	13	134.052	107.241	2811.85	a	v
6	30	10	13	134.829	107.863	3313.49	a	v
7	40	10	13	132.389	105.911	3748.12	a	v
8	50	10	13	131.250	105.000	4364.24	a	v
9	60	10	13	126.303	101.043	4924.47	a	v
10	70	10	13	127.770	102.216	5321.26	a	v
11	80	10	13	127.723	102.178	5903.84	a	v
12	90	10	10	126.880	101.504	7033.44	o	v
13	100	10	10	127.042	101.633	8393.53	o	v
14	110	10	10	123.621	98.897	9901.30	o	v
15	120	10	10	123.895	99.116	10577.46	o	v
16	130	10	10	127.932	102.346	9822.36	o	v
17	140	10	10	130.444	104.355	9162.39	o	v
18	150	10	10	133.386	106.709	8894.72	o	v
19	160	10	10	136.775	109.420	8755.58	o	v

図３　エクセル・シートによる音声パラメータの表示と変更・編集の例

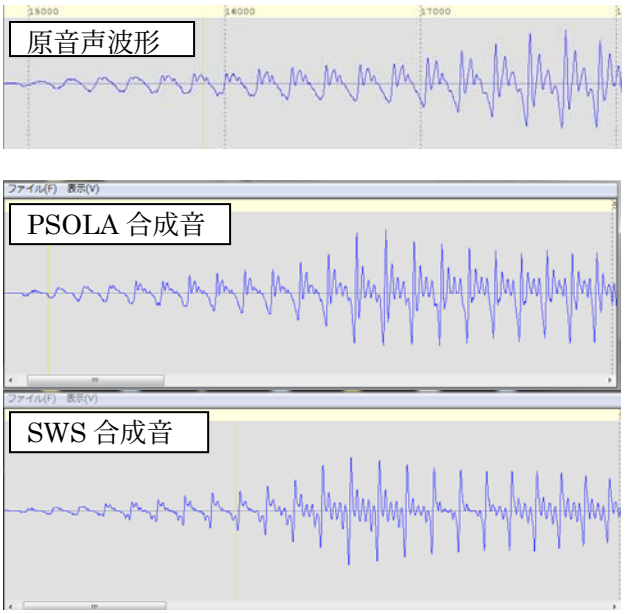


図４　原音声と合成音波形例（「青い波だ」の冒頭）（合成音は，いずれもピッチ周波数が原音声の 1.3 倍になっている）

図３に示すように，C 列の時間長と E 列のピッチ周波数が書き換え可能となっている．これらのセルを，直接または関数によって新たなパタンを生成するなどして書き換え，音声合成ボタンに(Sws), (Pso)によって合成音を受聴することができる．合成音波形表示機能で結果をチェックするなどして，必要ならば保存する．

なお，SWS 合成においては，FFT と線形予測分析のパラメータ（次数やデータ数）を，「音声」メニューの「合成パラメータの設定」で変更することが出来る．音声の帯域，標本化周波数に合わせて調整する．また，PSOLA 合成においては，ここではピッチマークが零交差位置に設定されるが，この合成を波形のピークの位置に合わせてピッチ波形を切り出して合成したい場合は，すでに述べたピッチマーク位置の修正手法によってその位置を変更して合成する．

PSOLA 法と SWS 法で合成した音声の波形例を，図４に示す．

６．SpitPlayer の内容

刺激音提示・集計ツールは，表計算ソフト Microsoft Excel 2010 上に実装された．このツールは，(聴き取り実験の設定)，(使用する音声ファイルの設定)，(提示順がランダム化された回答マスタシートの生成)，(被験者の回答シートの生成と聴取実験の実施)，(回答の集計シートの生成)，など実験の一連の流れが，エクセル・シート上で実行できるように作られている．以下，それぞれの内容に関して述べる．

ツールの本体 SpitPlayer.xlsm を起動すると，「実験計画」シート（表紙）と「音声リスト」シートからなるエクセル・ファイルが表示される．

（１）「実験計画」シート

聴取実験の枠組みを設定するシートであり，実験名，被験者数，聴き直しの可／不可，再判断の可／不可，音声を聴取したあとの判断の選択肢の記号，信号音の設定，再生開始時間，などの設定項目と，新たなシート生成のためのボタン類からなる．

図５に「実験計画」シートを示す．

判断の選択肢は２～５まで指定可能である．なお，聴き直しが（不可）の場合には，再判断を（可）とする（聴取と判断をやり直す）ことは許されていない．

（２）「音声リスト」シート

聴取実験で提示する刺激音声のファイル名を入力する．刺激音声の提示は，1 音声提示，2 音声提示，3 音声提示の 3 方法が可能となっており，対応する音声リスト・シートを選択し，入力する．

・1 音声リスト：刺激音は１つづつ提示する場合
・２音声リスト：刺激音は２つづつ提示する場合
（AX 法などで利用）
・3 音声リスト：刺激音は３つづつ提示する場合
（ABX 法などで利用）

（３）「回答マスタ」シート

「音声リスト」に使用する音声ファイルを記載後，「実験計画」シート上で「回答マスタ作成」のボタンをクリックすると，音声リストがシャッフル化された「回答マスタ」シートが生成される．

（４）「回答」シート

「実験計画」シート上で被験者数を記入し，「回答シート作成」のボタンによって被験者分の「回答」シートが作成される．

「回答」シートには，実験の開始ボタンが付いており，これをクリックすることにより実験が実施される．「開始」ボタンのクリックによって，エクセル・ウインドウはグレー表示になり，「実験計画」シートの設定に従った実験画面が表示される．これは，音声再生と判断ボタン等からなっており，これらをクリックしつつ実験を進めることができる．

実験用の画面例を図６（１）（２）に示す．（１）は，1 音声提示で判断は A または B の 2 選択の場合であり，（２）は 2 音声提示で，判断は A，B または？の 3 選択，再判断を可とするケースである．

実験が終了すると，結果は「回答」シート上に記載される．なお，聴き直しが（可）の場合，その音声を何回聞いたかも「回答」シートに記載される．（但し，集計はされない）

（５）「集計」シート

全ての被験者による実験の終了後，「回答の集計」のボタンによって，被験者全員の結果が集計された「集計」シートが作成される．

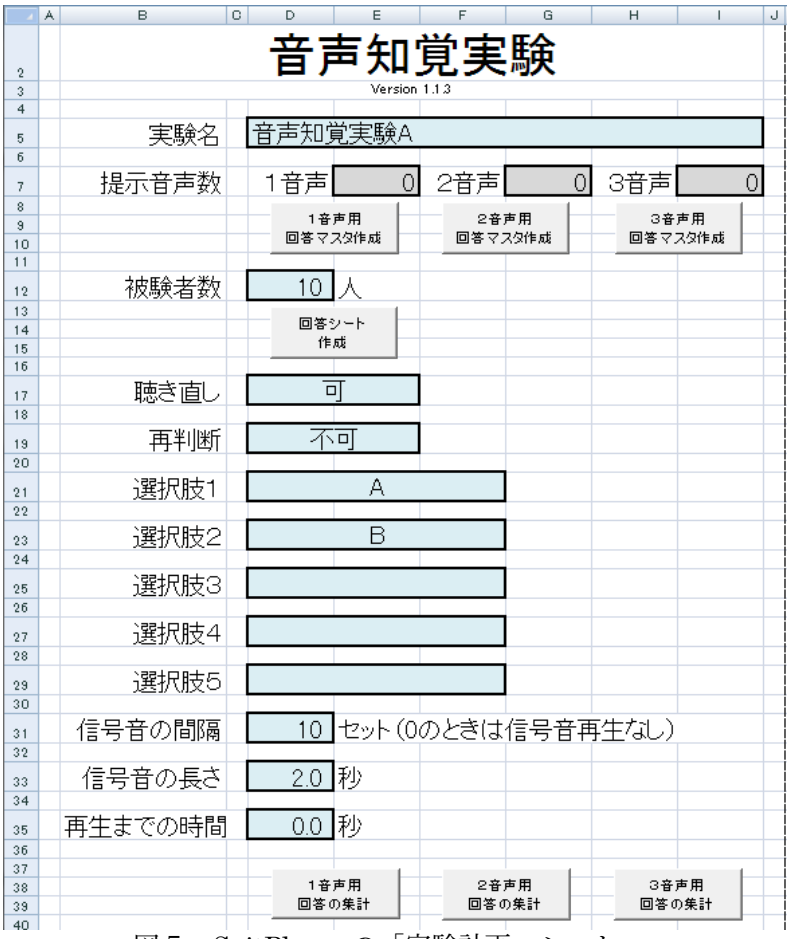


図５　SpitPlayer の「実験計画」シート

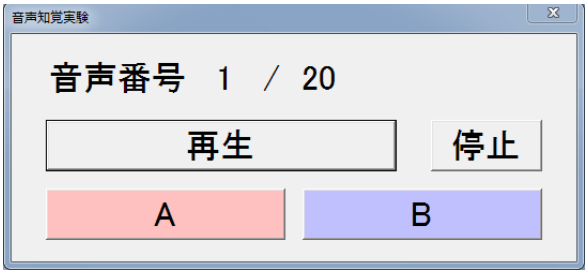


図６（１）　実験画面例（1 音声提示，2 選択肢）

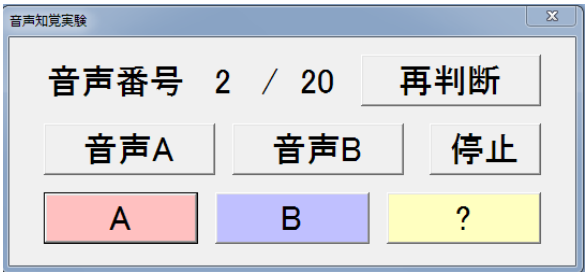


図６（２）　実験画面例
（2 音声提示，3 選択肢，再判断：可）

７．まとめ

アクセントや声調、イントネーションなど言語音声の韻律的特性を制御し、聴知覚実験を行うためのツールを作成した。表計算ソフトを利用して、使いやすいものとなるよう留意されている。今後は、具体的に日本語アクセントの知覚や声調言語における声調パタンの研究に適用していく予定である。

本研究は，科学研究費（基盤 B)「東アジアと東南アジア言語における超分節特性の比較対照に関する研究」（課題番号：23300093）及び（基盤 C)「東南アジア諸言語と日本語のイントネーションの音響音声学的研究」（同：23520457）の補助金によってなされた.

参考文献

1）Boersma, P. and Weenink, D.:　praat version: 5.3.41 (URL: http://www.fon.hum.uva.nl/praat/)

2）Moulines, E. and Charpentier, F. :"Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphone," Speech Communication 9, pp.453-467 (1990)

3) Hirokawa, T., Ito, K.. and Sato, H. :"High quality speech synthesis based on wavelet compilation of phoneme segments", Proceeding of ICSLP 92, pp.567-570 (1992)