

## 「見る・聞く」技術

# 音声合成

音声合成とは、人間が話す声をコンピュータで作り出すことです。

### ご利用にあたっての注意

この講座は、2009年当時の情報です。予告なしに更新、あるいは掲載を終了することがあります。あらかじめご了承ください。

最終更新日 2009年7月27日

## もくじ

- ↓ 人の声の出る仕組み
- ↓ コンピュータで声を作る流れ
- ↓ 音声を作る方式を3つ紹介
- ↓ 嘩り方の特徴を表す韻律（いんりつ）とは
- ↓ 最新の（動画）圧縮技術
- ↓ 韵律を変えると色んな声になるよ、聞いてみよう
- ↓ 小話
- ↓ 関連ページへのリンク

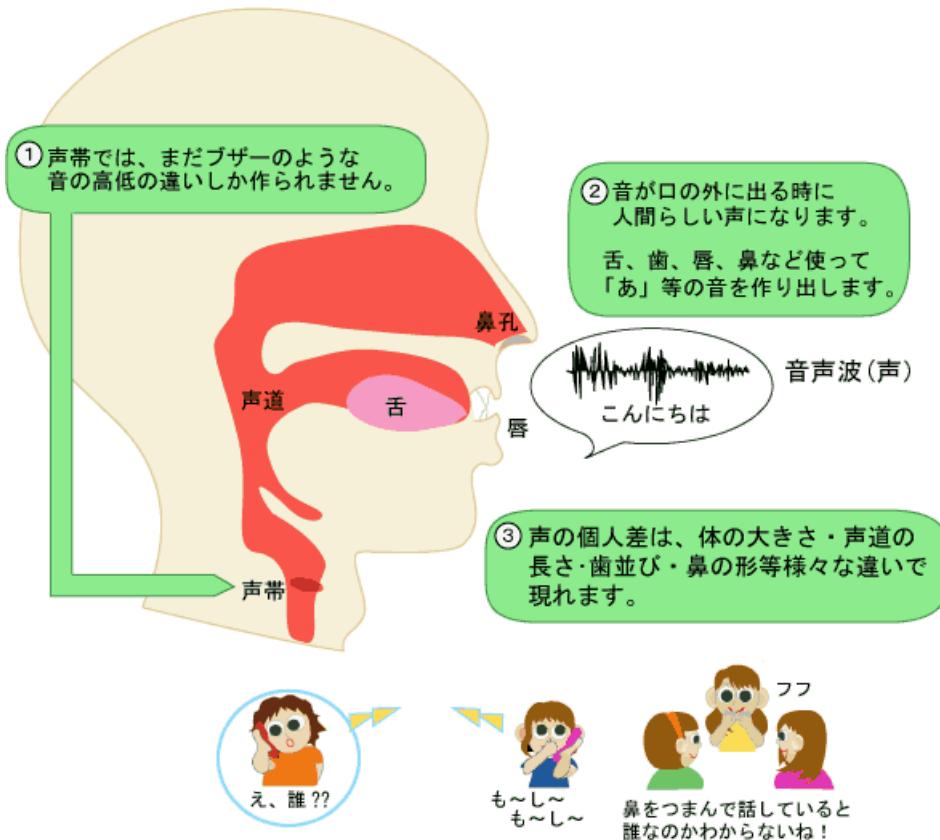
(講座メモ)

▶ ボタンを押すとサンプル音声が聞けます。



# 人の声の出る仕組み

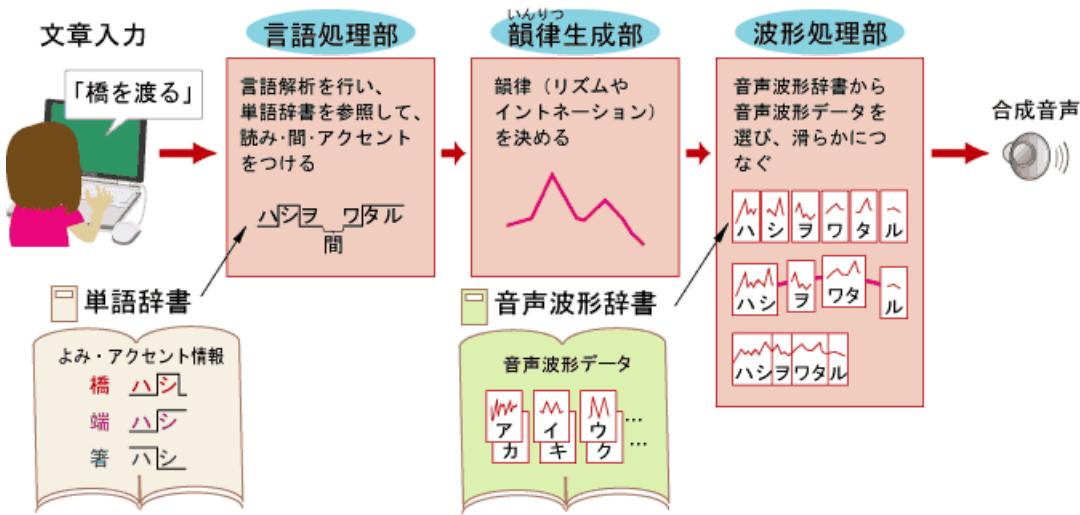
声はどのようにして作られるのでしょうか



## コンピュータで声を作る流れ

コンピュータでは、言語処理、韻律（いんりつ）生成、波形処理をして、音声が合成されます。

(韻律とは、声の高さ、イントネーション、リズム、ポーズなどの喋り方の特徴です)



## 音声合成用の発音

言語処理部では、漢字やひらがなの混じった文章を音声合成用の発音に変換します。この発音は、私達が普段使っている漢字を読むための「ふりがな」とは違います。例を挙げてみましょう。

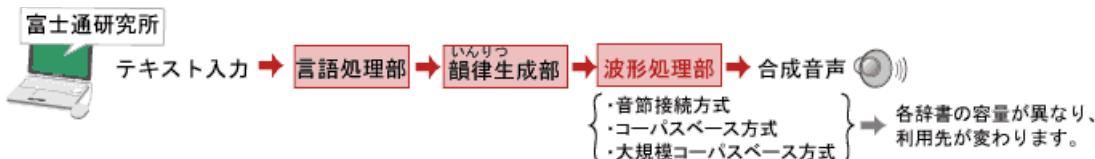


## 音声を作成する方式を3つ紹介

富士通の音声合成の波形処理部には、主な方式が3つあります。

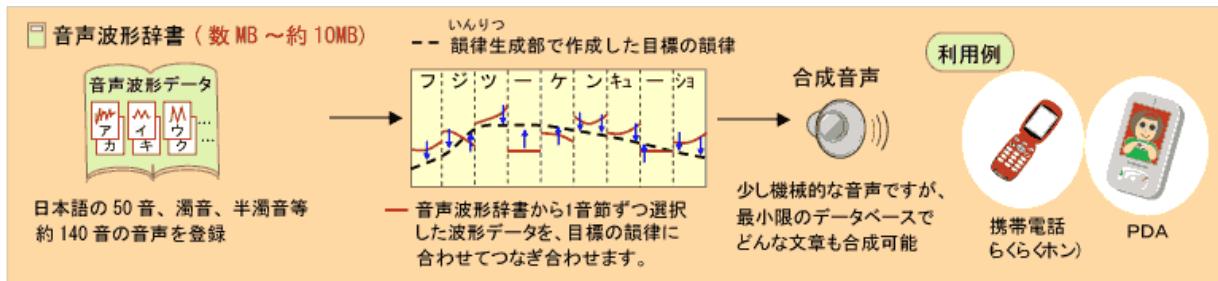
「音節接続方式」、「コーパスベース方式」、「大規模コーパスベース方式」

それぞれ音声波形辞書の容量が異なるため、携帯電話やカーナビや音声サービスなど音声合成機能が組み込まれる製品の特徴に応じて使い分けられています。



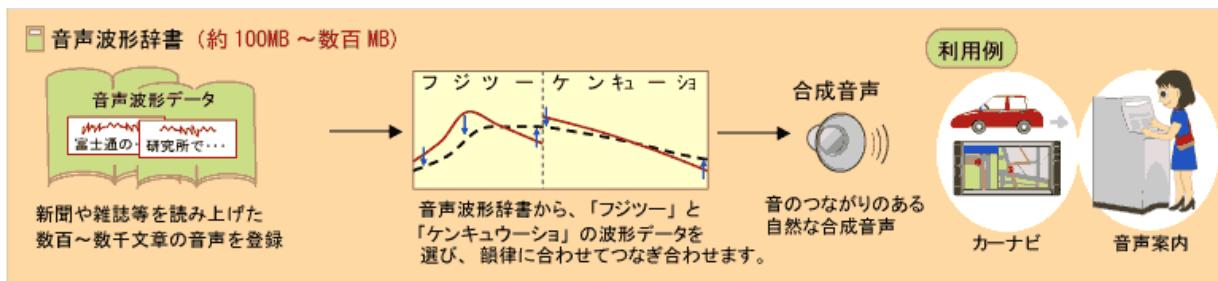
## 音節接続方式

- 音声波形辞書：数MBから約10MB（日本語の50音、濁音、半濁音等約140音の音声を登録）
- 特長：少し機械的な音ですが、最小限の音声波形データベースで、どんな文章も合成可能
- 利用例：携帯電話（らくらくホン）、PDAなど



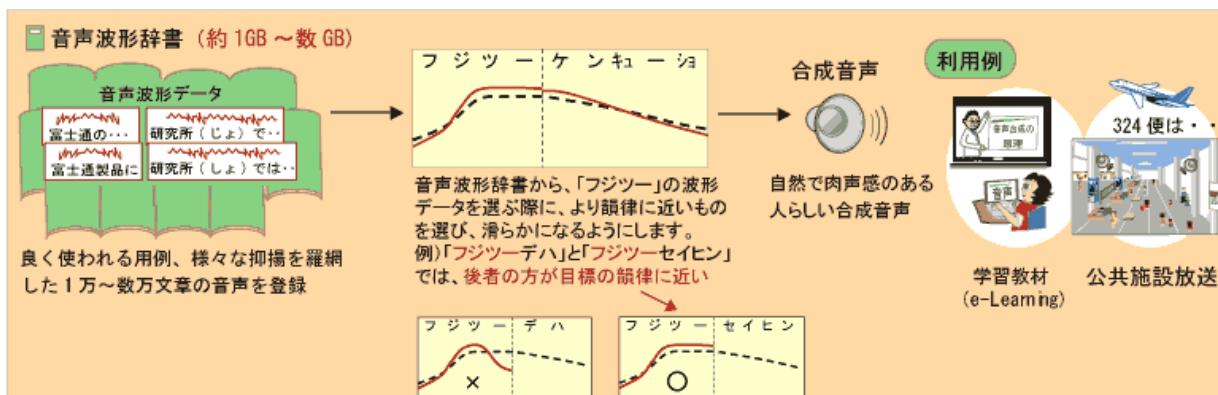
## コーパスベース方式

- 音声波形辞書：約100MBから数100MB（新聞や雑誌等の数千文章の音声を登録）
- 特長：音のつながりのある自然な合成音
- 利用例：カーナビ、音声案内など



## 大規模コーパスベース方式

- 音声波形辞書：約1GBから数GB（様々な抑揚を含む数万文章の音声を登録）
- 特長：自然で肉声感のある人らしい合成音声
- 利用例：学習教材（e-learning）、公共施設放送など



## それぞれの合成音を聞き比べて下さい



音節接続方式で聞く



コーパスベース方式で聞く



大規模コーパスベース方式で聞く

## 従来方式と新方式の音声合成を聞き比べてみよう

従来方式（コーパスベース方式）と新方式（大規模コーパスベース方式）で合成した音声のサンプルです。（当社従来比）

### ニュース

近畿地方の天気概況をお知らせします。今日の近畿各地は、日本海を中心を持つ高気圧に覆われて、中部、南部では晴れるでしょう。北部は寒気の影響で、雲が広がりやすいでしょう。空気が乾燥し、火災が発生しやすい状態が続きます。火の取り扱いには十分注意して下さい。



従来方式で聞く



新方式で聞く

### 音声案内

お電話ありがとうございます。こちらは、テクニカルサポートセンターです。お知りになりたい情報の番号を、プッシュボタンで押して下さい。なお、音声案内中のプッシュボタン操作も可能です。

1. 製品の使用方法。2. 製品の故障などのトラブル。3. 製品についての技術的な問い合わせ。もう一度お聞きになりたい場合は、9を押して下さい。



従来方式で聞く



新方式で聞く

### 館内放送

本日は、当店をご利用頂き、誠にありがとうございます。本館5階、レストラン街におきまして、イタリアンフェアを開催しております。本場イタリアのシェフが腕を振るった本格パスタなどをご用意し、皆様のお越しをお待ちしております。この機会に、是非ともご賞味下さいませ。



従来方式で聞く



新方式で聞く

## 電車内の放送

お客様にお願い申し上げます。発車間際の駆け込み乗車は大変危険ですのでおやめ下さい。

また、車内では携帯電話をマナーモードに設定して頂き、通話は周りの皆様のご迷惑となりますのでご遠慮下さい。皆様のご協力をお願い致します。

▶ 従来方式で聞く

▶ 新方式で聞く

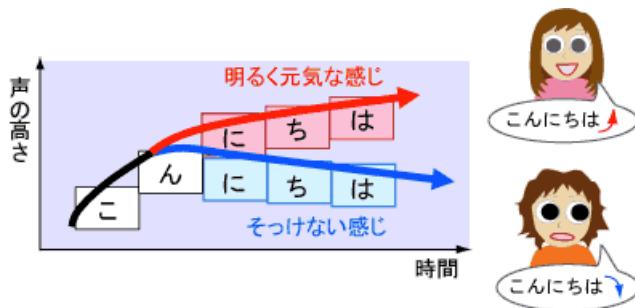
## 喋り方の特徴を表す韻律（いんりつ）とは

### 韻律ってなんだろう

韻律（いんりつ）とは、声の高さ、イントネーション、リズム、ポーズなどの喋り方の特徴です。韻律の違いによって、例えば同じ言葉でも喋る人の「気持ち」を伝えることができます。

気持ちを伝える例をひとつ挙げてみましょう。

「こんにちは」という言葉の場合、後半の声の高さが上がると明るく元気な感じを受けますが、反対に後半の声の高さが下がると、そつけない感じを受けることもあります。



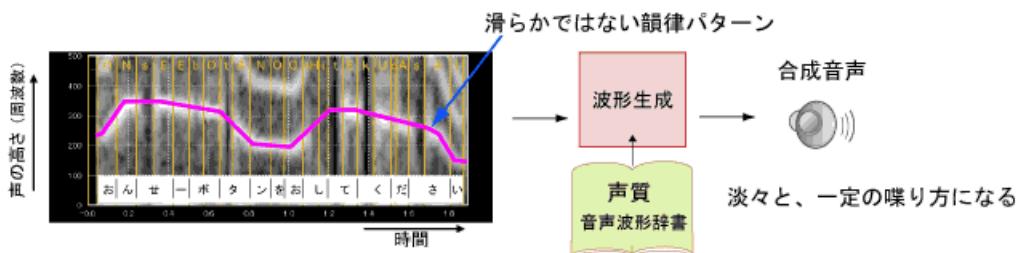
### 肉声の韻律を使った合成

富士通研究所は、より豊かな表現力を持つ韻律生成方法を開発しました。それは、人間の肉声から自動的に韻律を抽出する技術によって可能となりました。

ゼロから人工的に合成した韻律パターンではなく、沢山の肉声の韻律パターンの中から適切なものをピックアップし、組み合わせることによって滑らかな韻律を生成します。

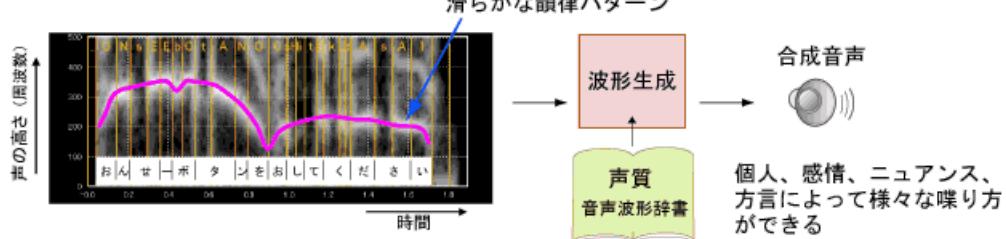
## 従来技術

人間の喋り方を真似て人工的に生成した韻律パターン



## 肉声韻律合成

肉声から抽出した韻律パターン



## 音声合成サンプル

従来技術と肉声韻律合成技術の音声合成サンプルを聞き比べて下さい。例文は「音声ボタンを押して下さい」です。

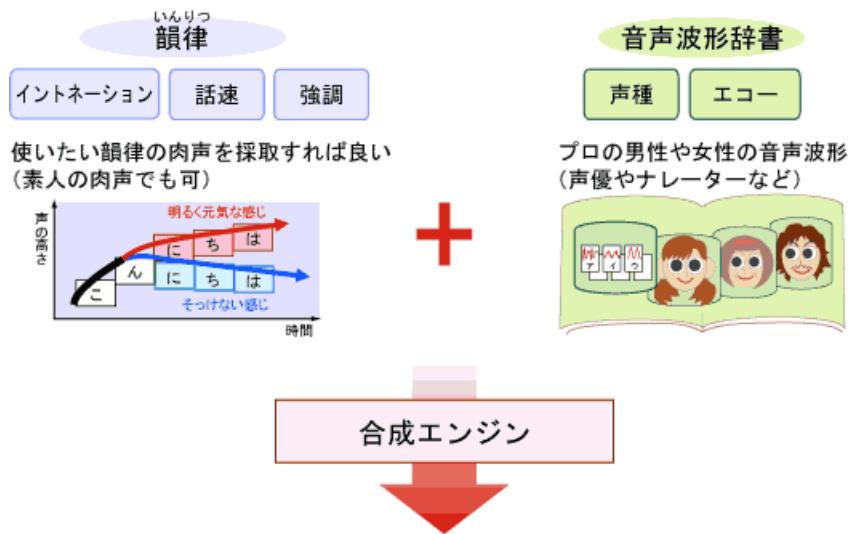
従来音声を聞く

肉声韻律音声を聞く

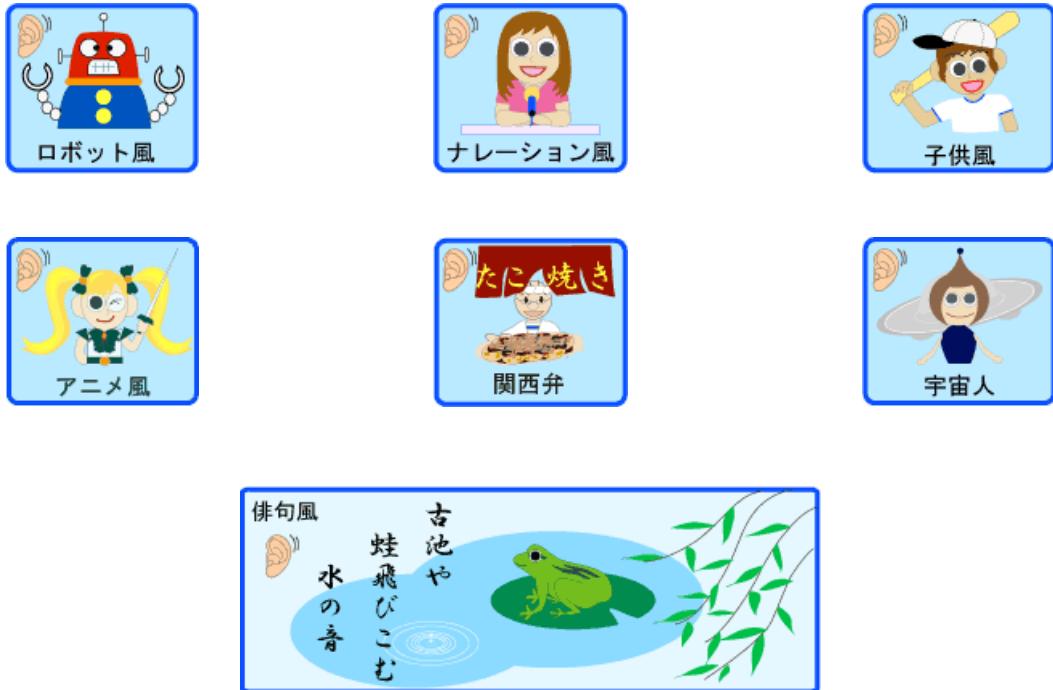
**韻律を変えると色んな声になるよ、色々聞いてみよう**

## 韻律制御により表現をコントロール

様々な韻律パターンを用意することによって、さらに豊かな表現力を持つ音声合成を実現することができます。色々な韻律パターンとプロのナレーターの声から作った音声波形辞書をあわせて、色々な特徴のある声を合成したり、エコーやビブラートを効かせる事によって歌っているような声や機械的な声などを表現することもできます。



様々な音声を作ることができます。  
その一例を下に挙げます。



## 適用事例

この技術は、2007年4月から2008年3月まで、NHK番組「爆笑問題のニッポンの教養」のナレーションに採用されました。

肉声のナレーションとどこか雰囲気の違う「人工音声」のナレーションが、番組自体を一種の異次元空間に見立てるという効果を演出しました。

## 小話

# 声の研究

毎日、声の研究をしている研究員は、声にはとても敏感になっているそうです。テレビ番組を見ていても、ついついナレーションの声質や深み、韻律（イントネーション、リズム、ポーズ）などを意識してしまいます。ナレーションだけを聞いて、その声の持ち主を特定できることも多いようです。研究員ではない筆者は、ナレーションを聞く時、どんな人の声かということまで注目していないし、その声の持ち主を特定することはできません。

また、それぞれの研究員によって、好みの声質があるそうです。例えば声のサンプルを探る際に、たくさんのナレーターの候補の中から選考する際、ある程度までは候補者を絞り込めるのですが、最終選考になるとみんなの意見がわかれてしまいます。そこで、最終選考に残った人の合成サンプルを作成してみて、一番きれいに合成することができた人に決定するそうです。

## 男性の声と女性の声はどこが違うのでしょうか

男性と女性の声はどのように違うのでしょうか。

女性のほうが高い声だというのは想像できますが、その他に声道の形の違いが影響します。女性の声道は、男性の約3分の2しかないので、声道での音の響き方が変わり、その結果として、音声の周波数成分が違ってきます。

音声は、色々な周波数の音が混ざり合ってできています。それぞれの周波数の強さの違いを周波数成分と呼びます。音質の違いは、この周波数成分の違いとして表すことができます。犯罪捜査に使われる声紋は、この周波数成分を表したものです。

合成音声では音の高さを自由に変えることができます。でも男性の声を高くしてあっても女性の声にはなりません。やっぱり男性が無理して高い声で喋っている声になります。それに男性と同じイントネーションで女性の声を合成しても、淡淡と情報を伝えているだけで、女性らしい声にはなりません。

男性らしさ、女性らしさというものは、音の高さだけではなく、普段何気なく話している会話の全体に現れるものようです。



## 関連リンク

### プレスリリース

- › プロのナレーションに迫る高品質な音声合成を実現（2009年7月27日）

### その他

