てくのろじぃ
# 解体新書

このコーナーでは東芝製品を支える
優れた技術や最新の研究成果を紹介します

2008年7月

# 音声合成技術

**ニャンダロー：**夏本番、車で出かける機会も多くなったけど、ルートはもっぱらカーナビ任せ。声で案内してくれるからとっても便利だニャ。

**籠嶋先生：**多くのカーナビゲーションシステムには、東芝の音声合成技術が使われているんだよ。

**ニャ：**最近のカーナビの音声案内は、本当に人がしゃべっているみたいに聞こえますね。

**先生：**カーナビ以外にも、PCの読み上げソフトやゲームなど、さまざまな機械で音声合成技術が使われているんだ。高音質でなおかつメモリ容量も少なくて済む安価な技術が可能になったからなんだよ。

**ニャ：**音声のもとになる音をあらかじめ覚えこませておくんですか？日本語なら50音で済むのかニャ？

**先生：**そう簡単ではないんだ。人間の声は、話す単語や文により高さもイントネーションも違うよね。単純に音をつないで言葉にしても、自然な人の声にはならないんだ。音のもと（音声素片）をあらかじめ準備した辞書から選び出し、発音やイントネーションの情報と照合して、声の高さ（ピッチ）を変えてつなぎ合わせる。音は空気を伝わる波だから、その波形をばらばらにしたりつないだり、伸び縮みさせたりして処理するんだ。問題は、ピッチを変えるときに音質が変わってしまう（劣化）ので、鼻にかかったような声になってしまうことなんだ。

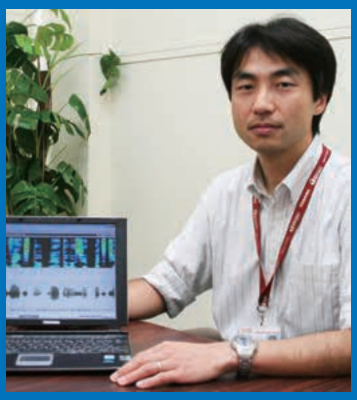
**ニャ：**それじゃ、高さの違うたくさんの音声素片を辞書に入れておけばいいんですか？

**先生：**それも一つの方法だけど、メモリ容量が膨大になってしまうので手軽に使えなくなってしまうよね。そこで、ピッチを変えても劣化しにくい音声素片を見つけることにしたんだ。

**ニャ：**でも、それって大変な作業ですよね。

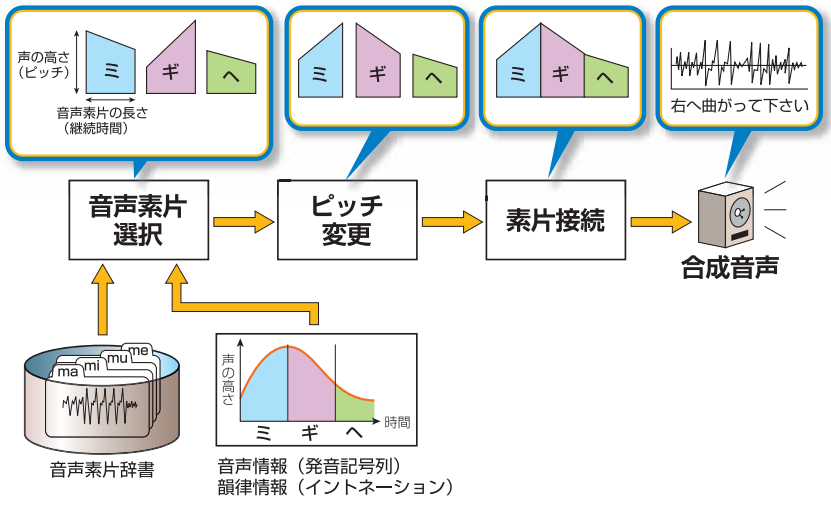
**先生：**人が耳で聞き比べて作業するならね。東芝では、これを計算機で評価できる方法を開発したんだ（閉ループ学習法）。ナレーターが発声した大量の音声の中から、同じ音の波形（音声素片の候補）を取り出し、その音のピッチをナレーターの発声を参照しながら（ピッチ分析）、いろいろな言葉の候補に合わせて変更する。その時、ナレーター自

今月の先生



研究開発センター
**籠嶋岳彦**さん

**音声合成の基本的仕組み**



身がその言葉を発生したときの声（教師データ）に一番近いものを一つだけ選んで辞書に登録するんだ。その選ぶ時の基準になる音色の差を、一定の式にあてはめて評価できるようにしたんだよ（誤差評価）。日本語の場合、子音と母音の組み合わせを単位とすると、300種類くらいの音をもっていれば、すべての音が出せるんだ。計算機に自動的に評価・選択させることで、短期間にこの300の音それぞれについて劣化が最小となる音声素片を選び出すことができたんだ。これで、高音質と省メモリを両立したんだよ。

**ニャ：**それはすごいニャ。

**先生：**300の音といっても、一つの音にはさまざまなバリエーションがあって、合成する言葉の候補もさまざまなんだ。その膨大な組み合わせの中から最適な一つを選ぶんだから、定式化できたことは大変な進歩なんだ。

**ニャ：**これは、日本語だけですか？

**先生：**発音の仕方が違うけど、あらかじめ音声素片のデータと言葉の候補を決めてしまえば、処理の仕方は同じだよ。すでに9カ国語に対応できているんだ。

**ニャ：**なにげなく聞いている合成音声だけど、人の声に近づけるにはものすごい苦労があったんだニャ。これからは、渋滞してもカーナビに八つ当たりしません。籠嶋先生、ありがとうございました。

※本技術は平成20年度全国発明表彰「内閣総理大臣賞」を受賞しました。