

統計的機械学習問題としての音声合成

徳田 恵一^{1,a)}

概要: 任意のテキストから音声波形を生成するテキスト音声合成（以下、音声合成）の問題は、しばしば「泥臭い」問題として扱われてきた。本講演では、音声合成の問題をいかにして、統計的機械学習の枠組みで定式できるかについて議論する。音声合成の基本問題は次のように記述することができる：

音声データベース、つまり音声波形と対応するテキストの組の集合がある。それらとは別に与えられた任意のテキストに対応する音声波形を求めよ。

これまでの音声合成システムは、テキスト解析、音響モデリング、音声特徴抽出/波形生成等、実現可能な部分問題に分解することにより実現されてきた。部分問題のひとつが“統計的パラメトリック音声合成”であり、隠れマルコフモデルを統計モデルとして用いる場合には“HMM 音声合成”と呼ばれる。これらの部分問題を統合し、ひとつの確率モデルとみることにより、上記基本問題を直接解くことが可能となる。本講演では、統計的アプローチによる音声合成研究の今後の展望についても議論する。

キーワード: 統計的機械学習, 隠れマルコフモデル, HMM 音声合成, HTS

Speech synthesis as a statistical machine learning problem

Abstract: Text-to-speech synthesis, which is a technique for generating natural-sounding artificial speech for a given text, is often regarded as a *messy* problem. In this talk, I would like to discuss how we can formulate the problem of speech synthesis in a unified statistical machine learning framework. The basic problem of speech synthesis can be stated as follows:

We have a speech database, i.e., a set of speech waveforms and corresponding texts. Given a text to be synthesized, what is the speech waveform corresponding to the text?

Conventionally text-to-speech systems were realized by decomposing the problem into subproblems, such as text processing, acoustic modeling, speech feature extraction/waveform reconstruction, etc. One of subproblems is statistical parametric speech synthesis, which is called “HMM-based speech synthesis” when we use hidden Markov models (HMMs) as statistical models. By combining all these processes and regarding them as one probabilistic model, we can understand the whole speech synthesis process in a unified statistical framework. I also discuss future challenges and the direction in speech synthesis research.

Keywords: statistical machine learning, hidden Markov model, HMM-based speech synthesis, HTS

¹ 名古屋工業大学
Nagoya Institute of Technology, Gokiso-cho, Showa-ku,
Nagoya 466-8555, Japan

a) tokuda@nitech.ac.jp