Title	パーティクルフィルタを用いた雑音に頑健な音声スペクトル上の複数ローカルピーク推定に関する研究
Author(s)	友池，誠二
Citation	
Issue Date	2008-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/4312
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

Estimation of local peaks of speech spectrum based on particle filter in noisy environments

Seiji Tomoike (0610061)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 7, 2008

Keywords: Estimation of local peaks, Non-stationary noise, Particle filter, Multi-dimensional likelihood.

As one main characteristic of speech, harmonics play an important role for speech recognition, fundamental frequency estimation, speech enhancement, as so on. The harmonics are closely related to the local peaks of speech spectra in the frequency domain. A highly accurate speech harmonics is given by estimating the local peaks of the speech spectrum. Therefore, the estimation of local peaks on the speech spectrum plays an important role for various speech processing.

For the local peaks estimation, many algorithms have been reported so far. There are two main types of the local peaks estimations. One is the methods based on extraction of every local peaks. The other is the methods based on extraction of harmonics. However, conventional methods often overestimate or underestimate the number of peaks. For the local peaks estimation on the speech spectrum, the conventional local peaks estimation methods do not use the estimated peak knowledge in the previous frames. These methods estimate the local peaks only in the current frame. Therefore, the conventional methods have a drawback that the noises greatly influence the accuracy of local peaks estimation in the current frame. The sudden incidence of peak candidates those are not appropriate for local peaks might be estimated in noisy environments. Thus, these methods have no robustness for noises. Since the harmonics vary gradually, the frequency positions of peaks in the previous frames are important for estimating the peaks in the current frame. Learning position from the previous frames offers benefits to estimation of local peaks in the current frame in noisy environments.

In this thesis, the author proposes a local peaks estimation method on the speech spectrum in noisy environments. The proposed method aims to estimate local peaks even in non-stationary noisy environments that conventional methods are hard to deal with. For the solution, the author uses the particle filter which is estimation of parameters from the current frame and the previous frames. The feature of the particle filter is to approximate the accurate posterior probability distribution using a lot of particles, i.e. discrete values. The posterior probability distribution is represented according to the density of the population of the particles. As the number of particles increases toward infinity, the approximation approaches the true posterior probability. The number of local peaks is unknown and the harmonics fluctuate in the high frequencies and the transition of each local peaks is independent. Therefore, the number of local peaks should not be given for the local peaks estimation on the speech spectrum. The proposed method introduces the likelihood which is able to estimate multi local peaks on the speech spectrum. The likelihood enables to estimate peaks present probability simultaneously. The local peaks might exist in the frequency ranges with the high peak presence probabilities to a high degree around the current frame. Therefore, the proposed method with the particle filter is effective estimation for the local peaks estimation that the number of local peaks is unknown.

The proposed method consists of a two-step estimation. The first step is to estimate the peak presence probability based on the spectral envelope of the cepstrum. The local peaks are able to be simultaneously estimated using the likelihood with the same peak presence probability in the high probability regions. The likelihood which describes the spectral envelope brings a criterion for determination whether peaks are present or not. No transition model is needed because the position of peaks in the next frame can be estimated by the peak presence probability. The proposed method estimates the peak presence probability by representing posterior probability distribution with a lot of particles even the number of local peaks is unknown or the transition of each local peaks is independent. The second step is to extract peaks from the candidates of the peaks based on the peak presence probability. The frequency bands which have maximal posterior peak presence probability become a candidate of peak. The position of peaks is obtained by extracting one point by extracting the point with the maximal peak presence probability.

In order to evaluate the accuracy of peaks estimation, two experiments are carried out under various conditions. In experiment 1, we synthesized the noisy speech by adding the pink noise from the first frame to end frame, and the comparisons between the proposed method and the conventional methods are carried out. In experiment 2, we use the synthetic speech by adding the narrowband noise whose duration is set to two frames from the fifth frame as non-stationary noise, and the comparisons between the proposed method and the conventional methods are carried out. To evaluate the performance quantitatively, we use the synthetic speech with the pre-set positions of peaks. These methods are evaluated with two measures, the number of the pre-set peaks and the frequency distance between estimated peaks and pre-set peaks. The number of correct local peaks derives accuracy of the number of local peaks. And, the distance between correct local peaks and estimated local peaks derives the accuracy of frequency distance. It is important to improve the accuracy of these two measures.

There are little differences between fundamental frequencies, the type of speakers or the contents of vowels. Therefore, the author averaged the results which have the similar tendency. In experiment 1, the results show that the proposed method is superior to the conventional methods in term of the frequency distance and the number of correct peaks even in the non-stationary noisy environment. This is because the proposed method learns the peak presence probability from the previous frames with a clue in slight evidence. In experiment 2, the proposed method is superior to the conventional methods in term of the frequency distance and the number of correct peaks under the condition of positive input SNR or narrowband noise. For the robustness of narrowband noise, the proposed method estimates local peaks in the conditions that the duration of narrowband noise within two frames. This is because the learning of the peak presence probability prevents from picking the local peaks come from the noises. However, in the conditions that the input SNR is negative and the added noise is pink noise or white noise, the proposed method estimates not well. This is because the proposed method learns the local peaks cannot be taken it as local peaks produced by a noise. Therefore, the proposed method made improvements of robustness for noises that were the drawback of the conventional method for practical condition.