

Nano-projet AMEA : Traitement de la parole avec PSOLA

Thomas Hézard

thomas.hezard@ircam.fr - 01 44 78 43 02

PSOLA (Pitch Synchronous OverLap and Add) est une technique de traitement du signal numérique utilisée pour le traitement de la parole. Cette technique permet, entre autres, d'effectuer des modifications de la prosodie (évolution de la fréquence instantanée du signal) et de la durée du signal. Nous présentons ici les grandes lignes de la méthode, on pourra se référer à [1] pour plus de détails.

L'algorithme PSOLA se découpe en trois grandes étapes :

- A étiquetage des périodes sur le signal d'origine et découpage en trames d'analyse courtes synchronisées sur les marqueurs
- B réorganisation temporelle des marqueurs et association d'une trame d'analyse à chaque marqueur de synthèse afin de modifier la durée du signal et/ou la prosodie,
- C re-synthèse par addition/recouvrement (overlap-add) des trames de synthèse.

L'étape A consiste à détecter les périodes dans le signal d'origine et à construire un marqueur au centre de chacune d'elle. Plusieurs méthodes existent afin de détecter la périodicité (variable au cours du temps) d'un signal de parole, nous nous proposons ici d'utiliser celle proposée par H. Kawahara dans [2]. On associe ensuite à chaque marqueur une trame centrée sur le marqueur et dont la longueur est proportionnelle à la période instantanée (à l'instant correspondant au marqueur).

L'étape B consiste à modifier l'organisation temporelle des marqueurs construits à l'étape A. On pourra déplacer, dupliquer ou supprimer des trames afin de modifier la fréquence instantanée et/ou la durée du signal. La construction des trames de synthèse se fait en deux temps : calcul de la position des marqueurs de synthèse à partir des fréquences instantanées (éventuellement modifiées) puis appariement trame de synthèse / trame d'analyse. Par exemple, pour allonger la durée on dupliquera les trames (association de plusieurs trames de synthèse à la même trame d'analyse), et pour baisser la fréquence instantanée on construira des marqueurs de synthèse plus écartés que les marqueurs d'analyse.

L'étape C consiste à construire le signal de synthèse par addition-recouvrement des trames de synthèse, en portant une attention particulière aux fluctuations apportées par l'addition-recouvrement des fenêtres (nécessité d'une normalisation).

L'objectif de ce projet est de réaliser un programme de modification de la parole avec l'algorithme PSOLA en Matlab.

Le programme devra être capable de transposer le signal d'origine par un coefficient constant et de rallonger ou raccourcir le signal d'origine par un coefficient constant. Une trame est proposée pour le programme Matlab (fonction `modif_PSOLA.m`) et une implémentation simple de l'étiquetage selon la méthode présentée dans [2] est fournie (fonction `etiquet.m`). Une fois les premières fonctionnalités validées, on pourra éventuellement améliorer la fonction d'étiquetage (gestion des zones non-voisées notamment) et/ou ajouter des fonctionnalités au programme (générer un signal de synthèse à pitch constant ou dont la courbe prosodique est fixée par l'utilisateur par exemple). Les articles de référence sont fournis.

Références

- [1] E. Moulines et F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones.", *Speech Communication*, 9, 5-6 (décembre 1990), pp. 453-467
- [2] H. Kawahara, Y. Atake and P. Zolfaghari, "Accurate vocal event detection method based on a fixed-point to weighted average group delay", *ICSLP-2000*, Beijing, pp.664-667 (2000)