

Pitch-Synchronous Overlap-Add (PSOLA)

由 Tom Bäckström 创建, 最后修改于九月 09, 2019

Many speech applications require from the ability to modify the fundamental frequency. For a classic but marginal application, think of the [auto-tune](#) function often used in post-processing of singing voices. With such tools it is possible to change the fundamental frequency of a speaker's or singer's voice without changing the phoneme or timbre of the sound. One of the more popular tools developed for this purpose is pitch-synchronous overlap-add (PSOLA). Like the name suggests, it is closely related to the overlap-add method used in the [short-time Fourier transform](#) algorithm. It allows changing the pitch of a speech sound without modifying or with only minor influence on other characteristics of the signal, such as vowel-identity. In addition to auto-tune, an important application of PSOLA is [speech synthesis](#), where we want to be able generate speech with any reasonable pitch contour. Voice conversion is another application, where the objective is to convert the speech of one person, such that it sounds like speech of another person.

The basic idea of PSOLA is to decompose the signal into individual pitch-periods, such that we can move the pitch-periods to change the effective length of those periods. That is, the fundamental frequency of a signal is expressed as a periodic structure of the time-signal. If we cut the signal into segment corresponding to the length of such periodic structures, then we can shift their positions as desired and then add them back together, like in the overlap-add process (see STFT). Since short-term correlations in the signal are not changed, that is, signal inside the windows/segments is not changed, then the spectral envelope of the signal is not changed.

The concept is best illustrated by example.
TBC