Synthesis and Control of Synthesis Using a Generalized Diphone Method

Xavier Rodet and Adrien Lefèvre

IRCAM, 1 place Stravinsky, 75004 Paris, FRANCE, rod, lefevre@ircam.fr

Abstract: Generalized Diphone Control is a powerful means of building a musical phrase from dictionaries of analysed sound units by building sequences of units and concatenating and articulating them. Through a graphical user interface on Macintosh, the Diphone 2.0 software provides analysis, control and synthesis according to various models, such as the Sinusoidal Additive model and the Chant model. A large experience has been obtained about its musical usage after two years of work with musicians and composers.

In speech synthesis, a diphone is a segment of sound extending from the middle of a phone to the middle of the next phone. More precisely, in order to allow for time and pitch modifications, a diphone is the set of synthesis parameters representing this segment in some given synthesis technique. All the diphones necessary for the synthesis of a language are recorded, analysed and stored in a dictionary. At the synthesis stage, a speech sentence is built by concatenating diphones found in the dictionary. We have extended the concept of diphone for the synthesis of musical sounds. Generalized Diphone Control is a powerful means of building a musical phrase from dictionaries of sound segments by concatenating and articulating them. Sound segments stored in a dictionary can comprise any recorded sound as well as instrumental sounds. Any analysis-synthesis method such as source-filter synthesis, additive synthesis or physical model can be controlled in terms of Generalized Diphones.

In this paper, we present several important improvements to diphone control and synthesis implemented in a software named "Diphone", such as a hierarchical structure of sequences and sub-sequences of diphones, variable speed of transitions, various smoothing functions at diphone boundaries, etc. To date, Sinusoidal Additive synthesis, Source-Filter synthesis and Formant-Waveform synthesis (Chant) have been implemented in Diphone and other methods are being ported. The graphical user interface of Diphone provides facilities for searching dictionaries, displaying their content and modifying them (Fig. 1) in addition to building, modifying and

```
Fig. 1. Management of a dictionary of instruments, segments and basic segments
```

79
synthesizing sequences of diphones and so on. Sequences and diphones are represented as graphical objects (Fig. 2) directly displaying their characteristics (duration, center, non-interpolation zones, loudness, etc.). Click-and-drag allows easy change of these characteristics in a WYSIWYG style. Parameter evolution stored in diphones or calculated from a sequence, are displayed in graphical windows. Modify, cut, copy and paste are fully supported on sequences and on parameter evolution. The combination of the Generalized Diphone control and graphical interface proved itself a very flexible and inspiring tool for composition and synthesis, which composers are using at IRCAM and outside of the institution. The novel characteristics of the Diphone software have opened interesting perspectives in vocal and instrumental sound processing. In particular, musical materials which were difficult to conciliate (e.g. discrete structures versus continuous ones), can now be exploited in a very flexible way. Additive synthesis, for instance, is known to be powerful but difficult to use and control for musical purposes. The Diphone program largely facilitates the use of additive synthesis. Moreover, the possibility of controlling different synthesis methods from the same program allows a more global control of musical aspects of sound synthesis.

REFERENCES