CHAPTER 7

MODELING AND EVALUATION OF PROSODY FEATURES

7.1 INTRODUCTION

The main components of Prosody which governs the design of a Tamil Speech Synthesizer are phrasing, duration and intonation. Phrasing is done at syntactic and semantic aspects. Duration modeling is carried out using CART. Standard Tones and Break Indices (ToBI) are there for languages to accomplish Intonation modeling.

7.1.1 Modeling Pause Duration Between Syllables

The pause duration that occurs between syllables is modeled to allow syllables to join in the same way as they occur in natural speech. To analyze and predict the pauses that occur in natural speech between syllables, an arbitrary text is manually parsed and hand labelled. The process is explained as follows:

(i) The pronunciations for the syllable pattern CV at the beginning and end of words are depending on how they appear.

(ii) For CVC and CVCC patterns, boundaries are marked at the consonants (LER).

(iii) Extract VC patterns label them. Generally, VCs occur at start and middle of the phrases.

(iv) Mark the transition periods between the syllables.

The pause duration between two adjacent syllables depends on the nature of articulation of the boundary sounds of the syllables. Consider the following example:

'/vEl koNdu/'

Here,

(i) *'/vEl/'* is a CVC syllable, let us label it as C1V1C2

(ii) *'/koN/'* is a CVC syllable, let us label it as C3V2C4

(iii) *'/du/'* is a CV syllable, let us label it as C5V3

The transition period or pause duration between the first two adjacent syllables, C1 V1 C2 and C3 V2 C4 is dependent on the manner of articulation of the consonants C2 and C3. i.e., the co articulation information of C2 and C3 plays the role. To analyze these characteristics, the sound units found in the speech database are classified. The sounds appear in their transliterated English form.

The paradigm analyzes the database for the transition periods between individual consonants. A set of 350 Tamil sentences employed for the analysis does this task. A class id is assigned for each speech unit of the acoustic class Proposed by Samuel Thomas (2007) to which its boundary sound belongs. Syllable database alone makes use of this classification. Table 7.1 gives out the assigned class ids for various classes of phonemes.

Table 7.1 Assigning Class Ids for various manners of articulations in Tamil (Syllables)

S.No.	Basic sounds	Class	Class Id
1	*k, ng, gn, kh (ayutha)*	Velar	S1
2	*c, ch, j*	Palatal	S2
3	*t, d, dh, n, nh*	Retroflex	S3
4	*th, ndh,*	Dental	S4
5	*p, b, m*	Labial	S5
6	*y, r, l, v, zh, rh*	Approximant	S6
7	*sh s h sr*	Fricative	S7
8	*a, e, u, ae, i, o, ou*	Vowel	S8

Diphone Database does not need such an ample description and classification of sounds. Still, there are some co articulation differences for various pronunciations at the merging of a vowel and a consonant. The manner of articulation for the existing classification of consonants viz., vallinam, mellinam and idaiyinam are used. Exceptional manners of articulation such as kutriyalugaram, kutriyaligaram, etc., are ignored having the size of the database in mind. Table 7.2 gives out the assigned class ids for various classes of phonemes for diphone database.

Table 7.2 Assigning Class Ids for various manners of articulations in Tamil (Diphones)

S.No.	Basic sounds	Class	Class Id
1	*k, s, ch, d, t, p, b, r, rh*	Vallinam	D1
2	*ng, gn, n, nh, m, ndh*	Mellinam	D2
3	*y, r, l, v, zh, rh*	Idaiyinam	D3
4	*a, e, u, ae, i, o, ou*	Vowels	D4
5	*kh (ayutha)*	Specials	D5

The thesis gives an analysis of the pause duration between syllables based on these classes in the following subsections. In each of the subsections, an analysis is done concerning the sounds of the boundary syllables of two consecutive speech units. As the co articulation information of the diphone databases is self-explanatory, we do not discuss them exquisitely.

7.1.2 Pause Duration Between Velar Sounds and Other Sound Units

Analyzing the examples of speech units where the boundary sound belongs to the velar class, gives the following pause duration values. Consider the example: *‘/pakkam/’*. The pause duration takes an average of 35 milliseconds for most of the classes. If the adjacent sound is a dental sound, the pause duration is more. Figure 7.1 gives the average pause duration between velar sounds and other sound units.

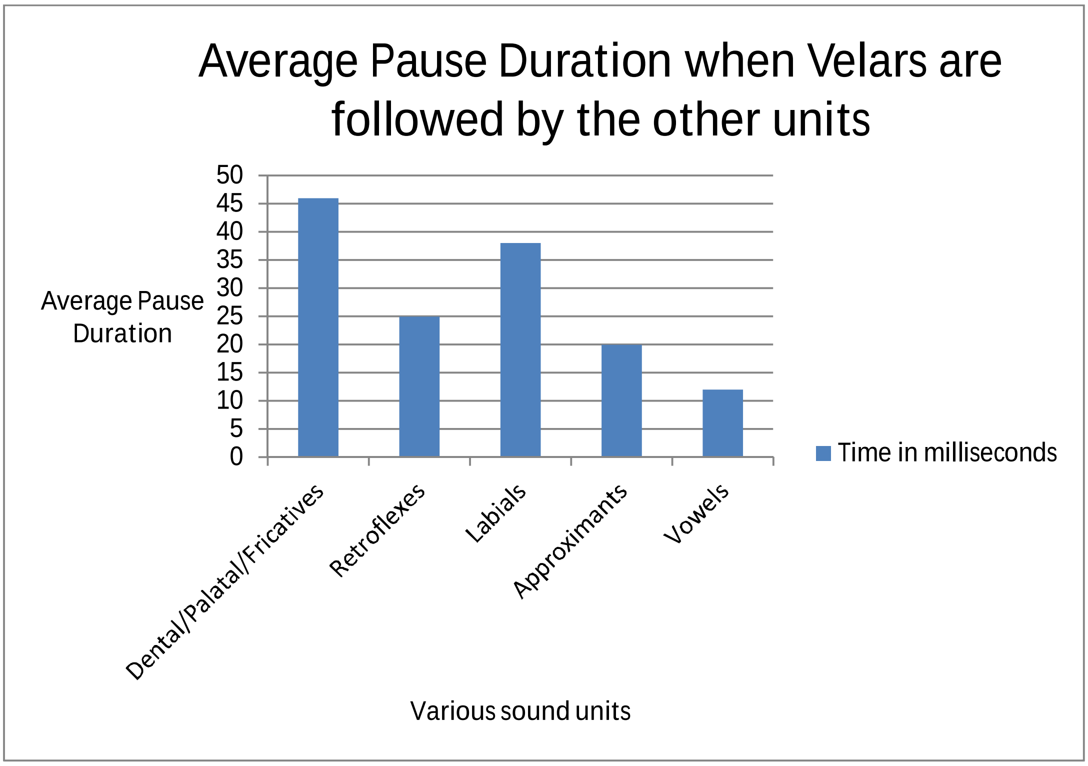

Figure 7.1 Average Pause Duration between velar sounds and other sound units

7.1.3 Pause Duration Between Palatal Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the palatal class, gives the following pause duration values. Consider the example: *'/kAtchi/'*. Unlike the case of velar sounds, the range of values for different classes is lower and is between 12 seconds and 45 milliseconds. Figure 7.2 gives the average pause duration between paltal sounds and other sound units.

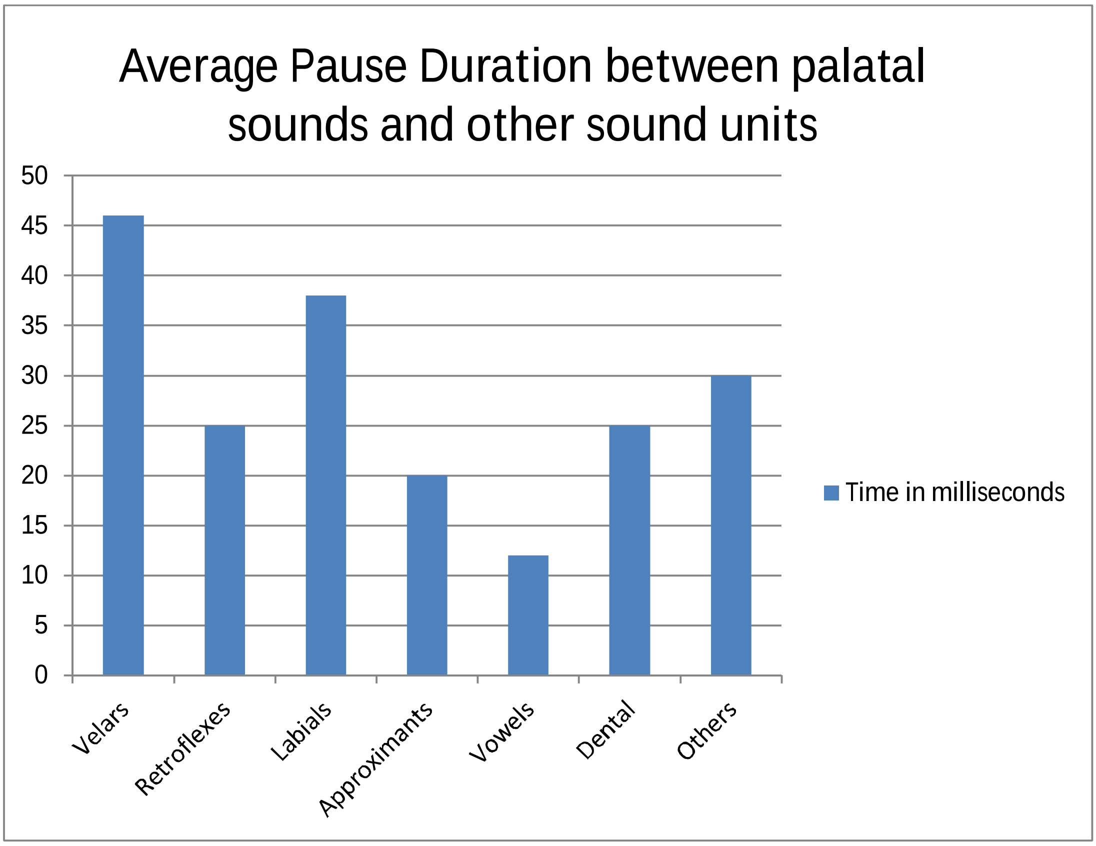

Figure 7.2 Average Pause Duration between palatal sounds and other sound units

7.1.4 Pause Duration Between Retroflex Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the retroflex class, gives the following pause duration values.

Consider the example: *'/pattu/'*. The average pause duration between retroflex, labial, approximant and fricative sounds shows a narrow range probably because of the similarity in the place of articulation. The range of the pause duration is larger, in this case, extending from 12 to 50 milliseconds. Figure 7.3 gives the average pause duration between Retroflex sounds and the other sound units.

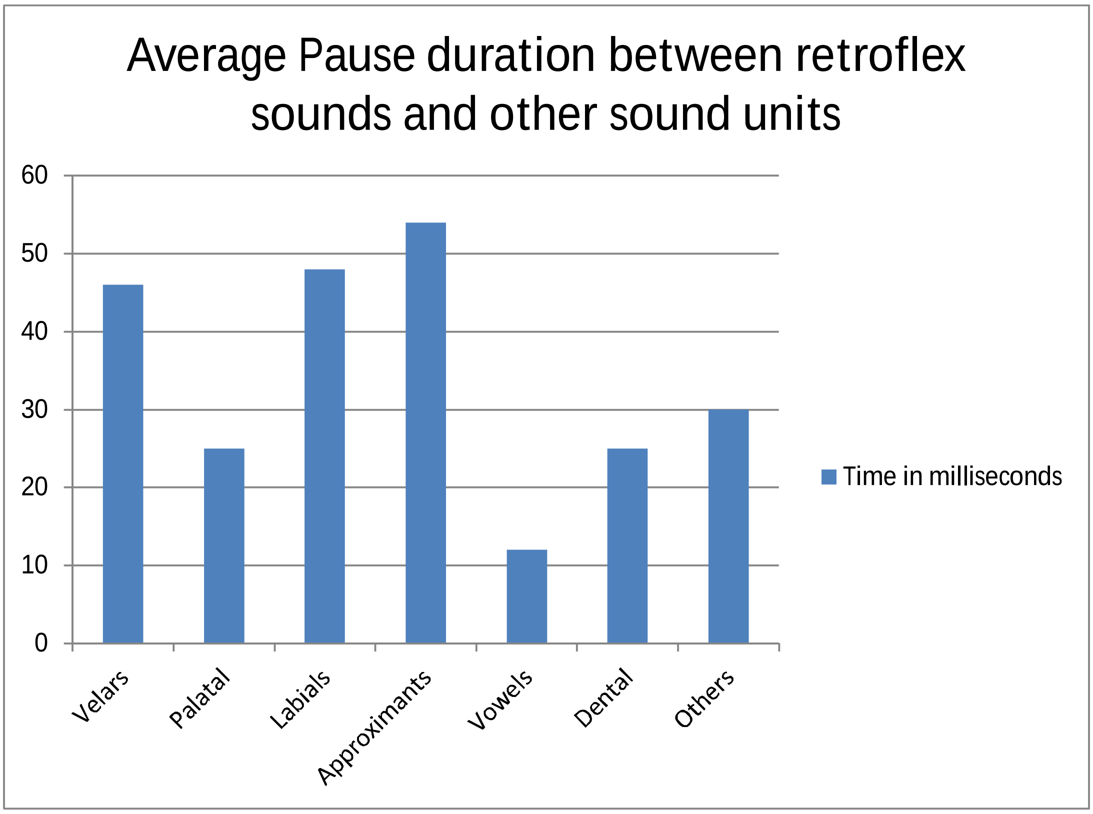

Figure 7.3 Average Pause Duration between Retroflex sounds and other sound units

7.1.5 Pause Duration Between Dental Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the dental class, gives the following pause duration values. Consider the example: *'/athu pOl/.'* The adjacent boundary sounds for dental sounds are either retroflex or labial. Two distinct ranges for transition periods are transpire in this case. The range of the pause duration is between 30 to 70 milliseconds. Figure 7.4 gives the average pause duration between dental sounds and the other sound units.

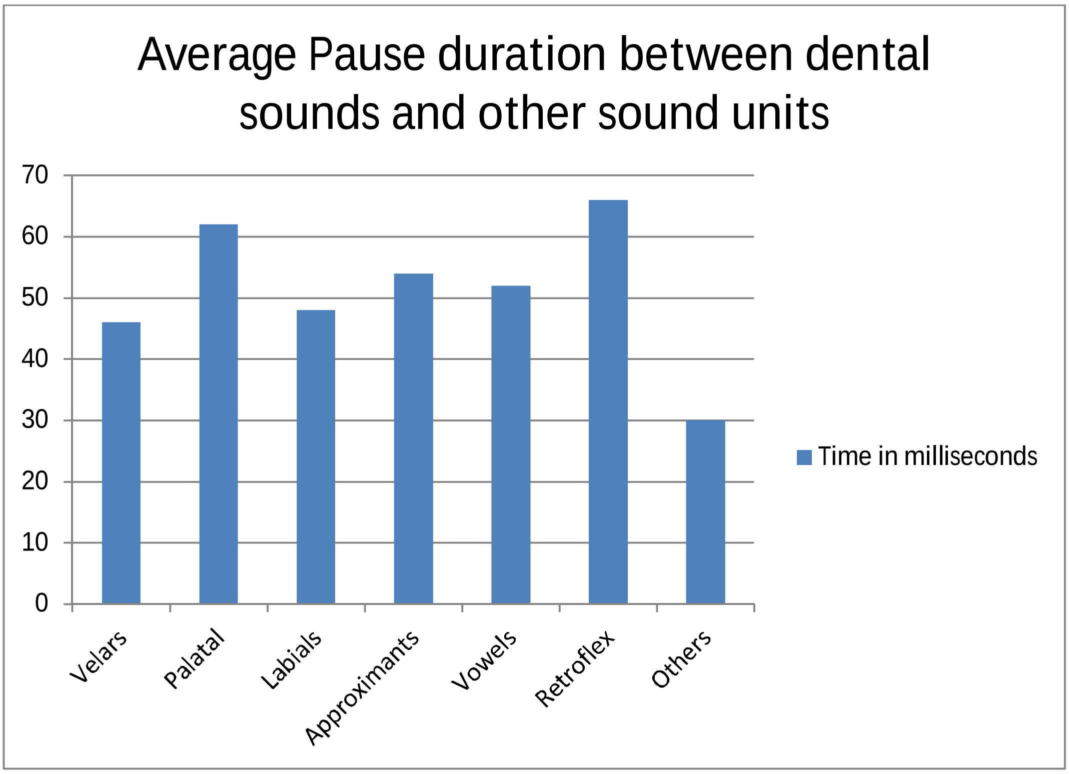

Figure 7.4 Average Pause Duration between Dental sounds and other sound units

7.1.6 Pause Duration Between Labial Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the labial class, gives the following pause duration values. Consider the example: *'/pagal/.'* When compared to other classes described above, the transition period in this class is higher and has a maximum duration range between 15 and 40 milliseconds. Figure 7.5 gives the average pause duration between labial sounds and the other sound units.

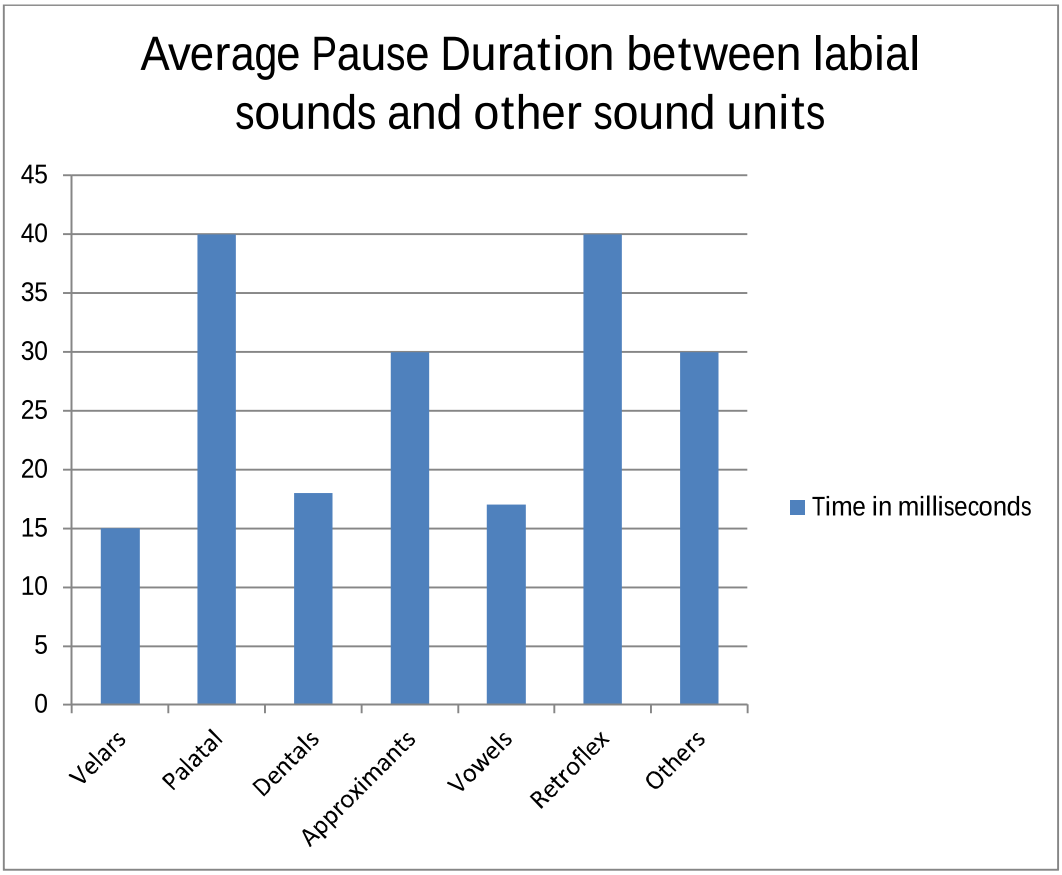

Figure 7.5 Average Pause Duration between labial sounds and the other sound units

7.1.7 Pause Duration Between Approximant, Fricative Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the approximant and fricative classes, gives the following pause duration values. Consider the examples: *'/tamil mozhi/'*, and *'/varusham/'*. The average duration of these classes are similar, and all of them have the lowest transition periods between adjacent sounds. In both the groups, the average pause duration is 20 milliseconds. Figure 7.6 gives the average pause duration between approximants, fricative sounds, and the other sound units.

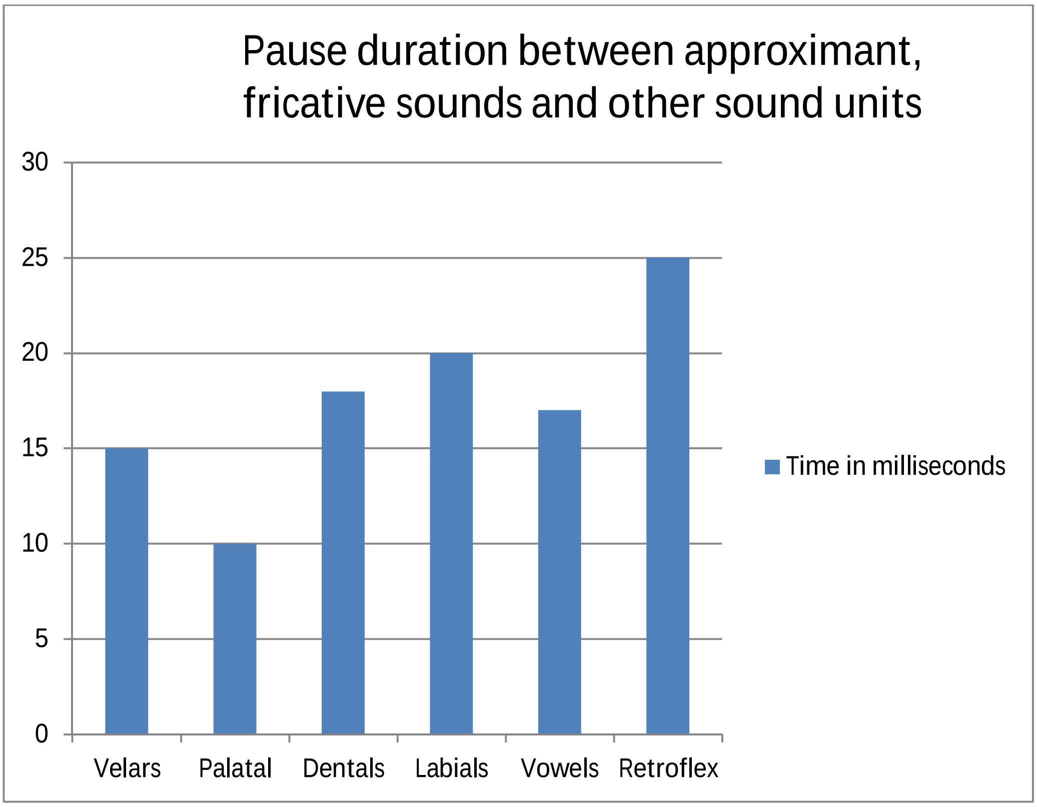

Figure 7.6 Average Pause Duration between approximants, fricative sounds, and the other sound units

7.1.8 Pause Duration Between Vowel Sounds and Other Classes

Analyzing the examples of speech units where the boundary sound belongs to the vowels, gives the following pause duration values. Consider the example: *'/ithayam/'*.It is vivacious that since syllables have been split at consonant C boundaries, no word splits such that a vowel V appears in the middle of a word. Two Vs appear adjacent to each other only when the last syllable of a word is a CV, and the first syllable of next word is a VC. The pause between two Vowels is always like a pause between two words. That is why the average pause duration of vowel sounds is between 25 and 72 milliseconds. Figure 7.7 gives the average pause duration between approximants, fricative sounds, and the other sound units.

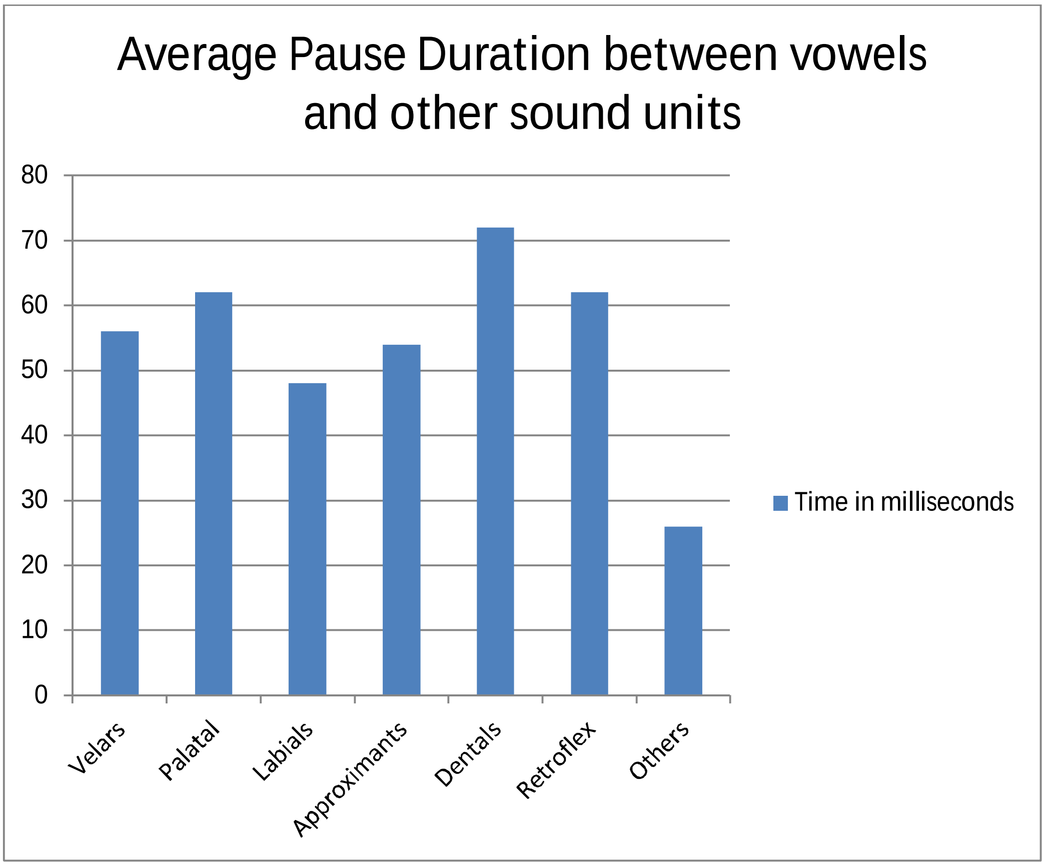

Figure 7.7 Average Pause Duration between vowels and the other sound units

7.1.9 Pause Duration Between Syllables and Diphones

Analyzing the examples of speech units where there is a need for diphone synthesis, we have found that there needs extra coarticulation constructs. The boundary sound may belong to any class of consonants or vowels, a syllable should naturally blend with a diphone. Here, the pause duration value depends also upon the time taken by the synthesizer to switch between databases. Therefore, the time taken by the synthesizer for context-switching should be neglected to model the pause duration. Though the synthesizer neglects the system-bound time, the existence is still there in the synthesis phase. Paying to the naturalness of the synthesized speech, one should bear with this processor-bound time delay. The following pause duration values are obtained while analyzing the exact time taken for diphone synthesis. Consider the example: 154, which is synthesized with diphones as *'/onru ainthu naanku/.'* The transition period in this case is different and has a duration range between 20 and 35 milliseconds based on the acoustic properties of the word. Figure 7.8 gives the pause durations taken by syllables and diphones. The elevated duration is the time taken for context-switching.

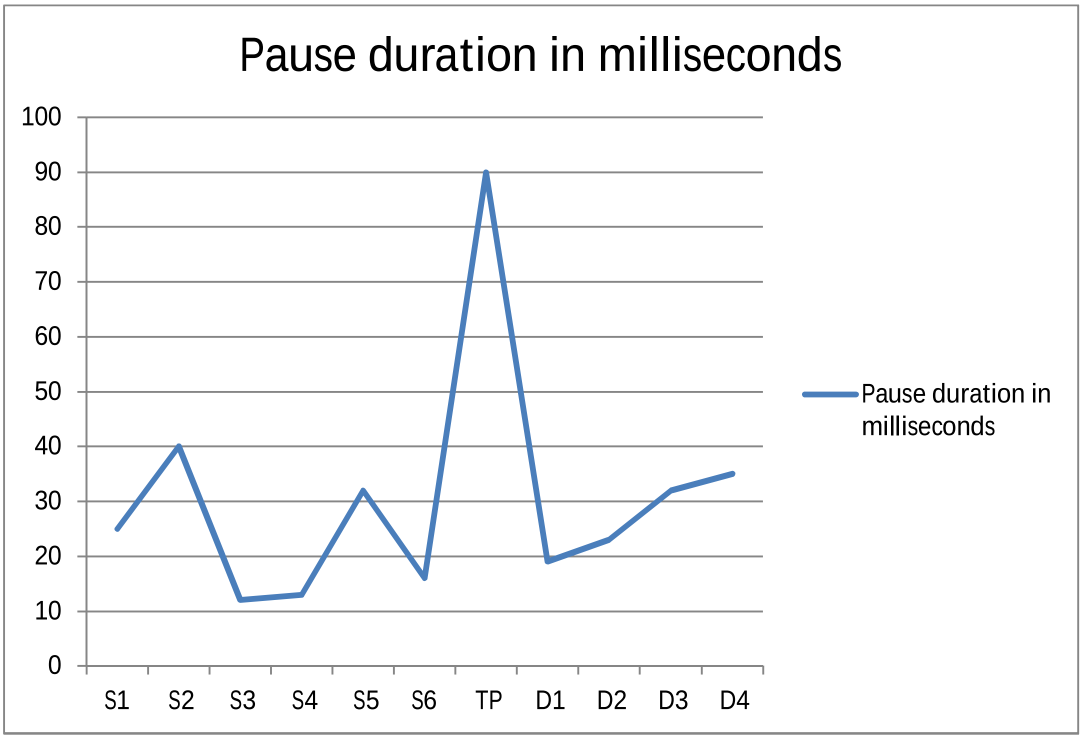

Figure 7.8 Pause Duration taken by syllables and Diphones

7.2 CART MODEL FOR PAUSE DURATIONS

Pause durations differ based on which class a boundary sounds belong. The co articulation of these classes is modeled using Classification and Regression Trees (CART). The CART model is then used to predict the transition period between two syllables.

A CART model is built using examples of different speech unit combinations obtained from hand labeling 350 Tamil sentences. Hand labelling uses the class Id information shown in Tables 7.1 for syllables and 7.2 for diphones. The advantage of following this format is that, when the transition period between a pair of speech units needs a prediction, either the class id information or the name of the speech units can be used.

7.3 SUMMARY

In this chapter, we have discussed the steps followed to build a Tamil Text to Speech synthesizer.

The thesis has also presented the evaluations of the speech units. It has showcased the analysis of various synthesizers. The results of the evaluation show the efficiencies of the speech units. We have discussed two important issues in using syllables and diphones for speech synthesis. At least three realizations of a speech unit must be present to represent a speech unit in the database. Investigations are carried out to check the transition periods between the syllables. As syllables are the better predictors of prosody, the acoustic boundaries themselves have their specific join costs. The transition periods are modelled using a CART tree.