Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
2(1), 2014, 175-179

ISSN 2347 – 7393

Original Research Article

GALO TEXT TO SPEECH USING BASQUE INTONATION MODEL

Laba Kr. Thakuria, Tulshi Patar, Akalpita Das, Purnendu Acharjee, P.H. Thakdar

Department of Instrumentation and USIC
Gauhati University

Abstract

The main objective of this paper is to design in text to speech conversion using Basque intonation. Using this f0 curves of parameterized process are designed according to the intonation model of Fujisaki. Here in the experiment the f0 curves are estimated with 1 ms precision and then decimated to have 5 ms precision. In our experiments, we have created the phrase commands of the model which are described and we have also analyzed the outcomes of the experiments. A good model of intonation is essential to get a high quality text to speech (TTS) conversion. In the observation using speech synthesis the database is manually labeled at word and phone labels which are automatically generated. The database is basically parameterized in three different conditions that are the root means squared error (RMSE,) which is obtained in each case is calculated and compared. Sentences and accent groups are automatically generated using the information provided by the linguistic analysis part of text to speech (TTS) conversion system for Basque. The experimental analysis of the parameters which are obtained using regression trees and classification are described in this research paper. Whatever the evaluated values that we have obtained in our study analyzed and explained.

Keywords: Galo, Basque intonation, f0, TTS, RMSE, speech synthesis, regression tree.

Introduction

A good model of intonation is needed to create a high quality text to speech (TTS). The declination rules and the modeling of different types of sentences are very basic. To increase the quality and the accuracy of the synthesis, we have selected Fujisaki's model of intonation. Fujisaki's model is already used for the synthesis of intonation in many different languages. This model has been adapted for standard Basque language in this work and then introduced into GaloTTS. The basic layout of the paper is as follows: in part 2 we have described the speech material utilized in the

For Correspondence:
thakurialabaATgmail.com
Received on: February 2014
Accepted after revision: February 2014
Downloaded from: www.johronline.com

modeling of intonation and then, in part 3 the way of intonation labeling is made is explained. In part 4 we have discussed the statistical intonation parameters and in part 5 and in part 6 we have discussed the result and conclusion respectively.

A) Galo

The Galo language (Agom) is inherited from the Tibeto-Burman family of languages. This language is mostly spoken by the Galo people. of Arunachal Pradesh. The Galo is one of the major tribes of Arunachal Pradesh. Around 95% of Galo people learn Galo as a first language, although most are also bilingual and borrow frequently from Assamese, Hindi and English (The major languages of Indian subcontinent). In the Arunachal Pradesh there are total 25 major tribes and almost 110 sub- tribes. There is high degree of mutual intelligibility among the different languages of Arunachal Pradesh like language spoken by the Adis, Apatanis, Galos, the Hill Miris, the Nyishis and the Tagins. Moreover they share many characteristic features in their cultural code and trace their ancestry from a common forefather, namely Abotani. Hence, the language spoken by them can rightly be given generic name –Tani language. The languages in Arunachal Pradesh can broadly be classified into two groups: namely Abotani group and Non-Abotani (Buddhism). The majority of the tribes of the state belong to Abo Tani group and Galo belong to Tani group. These language groups are very close both in syntax and semantics. Galo people can understand each other when speaking with different Galo dilects .From region to region, village to village, and clan to clan, Galo people speak slightly differently in pronunciation and vocabulary. Sometimes differences are in pronunciation, sometimes in the actual words used, sometimes in the meaning of those words, and sometimes in the way they are used (i.e., the grammar). The major Galo dialects are Pugo, spoken around the district capital (Itanagar), Aalo and Lare spoken in the south of Aalo, and Subdialects are numerous, and often correspond to regional or clan groupings. The Galo have decided to adopt the Roman script with certain additions of two vowels and two consonants whose phonetics are common in international phonetic alphabets (IPA). The most common additions are the Roman symbol v,w,q and x, which are not use for writing Galo. These symbols now represent IPA ə, ɨ, ŋ and ɲ/ñ respectively. The majority of the tribes of the state belong to Abo-Tani group. The evolution of language can be represented chronologically as shown below:

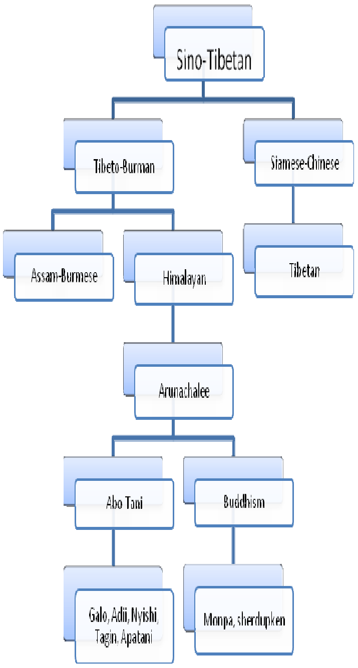

Figure 1: Chronological representation of Arunachal language

Speech Material

Fujisaki's intonation model for Basque is initially proven using a very small dialectal database. To achieve the desired level of quality for text to speech (TTS) synthesis in Basque, a more extended and complex database is required. So, in the experiment a new and more complete corpus database is designed with this goal. This corpus is read by a native male Basque speaker in standard Basque in a laboratory environment. The resulting database is called GU_Galo. Our database comprises 3000 isolated sentences with various syntactic structures, different lengths and diverse levels of complexity. The vocabulary used is very rich, with 4000 different words out of 5000. This database also includes special particles that could have a distinctive effect on intonation. The main characteristics and the distribution of sentences in this database are shown in table 1.

Table 1: Main characteristics of GU_Galo database.

	Value
Size of recording	64 Mb
Sentences	344
Declarative sentences	238
Question sentences	80
Exclamation sentences	26
Prosodic phrases	630
Words	2398
Different words	1380

In the research experiment mainly the database is manually labeled at word level and then phone labels are automatically generated using speech synthesis and a dynamic time warping algorithm. Sentence and accent group labels are automatically generated using the information provided by the linguistic analysis module of the text to speech conversion system for Basque GaloTTS. f0 curves were calculated with 1 ms precision, applying a method based in [6] and then decimated to have 5 ms precision. The resulting distribution of breaks is shown in table 2. According to this distribution most of the breaks must be modeled with a phrase command (79%).

Table2: Distribution of breaks in Galo database

	Orthographic break	Non-Orthographic break
Total	169	117
With PC	140	83
Without PC	29	34

Labeling Intonation

Here according to Fujisaki's intonation model, the f0 curves are automatically parameterized with an algorithm based on analysis-by-synthesis. These combinations of parameters are selected according to certain linguistic constraints detailed in [5]. The accent commands are placed related with the position of the corresponding accent group and its duration is limited to vary in a certain range depending on the position of the stressed syllable within the accent group. One accent command is used to model each accent group, except for the last accent group of questions and exclamation sentences, where an extra command has been added to model the final rise in the intonation curve.

A). Labeling experiments

The whole database is parameterized three times varying the locations and number of phrase commands: □□In the first, phrase commands are placed only in the breaks that are orthographically marked. □□□In the second, consisted in placing a phrase command at every break uttered by the speaker. □□In the last experiment phrase commands are placed only in the breaks labeled as needing a phrase command. This indicates that both a break insertion model and a break classification algorithm are available. Figure 1 shows the synthetic intonation curves corresponding to these three labeling experiments, related to the natural one for one of the sentences of the database.

B). Labeling results

After the whole database is parameterized under the three conditions, the root mean squared error (RMSE) obtained in each case is calculated and compared. Figure 2 shows the RMSE in all cases: we have obtained the biggest error (12.09%) corresponds to the first experiment, where phrase commands are placed only in breaks indicated in text. The other two cases have smaller error, being the difference between them no meaningful (11.78% and 11.67%).

For the synthesis of intonation in Basque language using Fujisaki's model, a good model of insertion of breaks is highly needed, but accurate classification of these breaks is not important. Considering these results, when synthesizing intonation one phrase command will be introduced for each prosodic phrase.

Parameter Estimation From Text

To get the appropriate synthetic intonation curve, the accent and phrase command parameters have to be related with characteristics extracted from input text. The values of intonation parameters has been made using classification and regression trees (CARTs), because they are able to manage variables of discrete and continuous nature, they automatically select the factors that have the greatest influence in the prediction of the target variable.

A).Prediction variables for accent commands

The information that are used for predicting accent command parameters is mainly related to the accent group. The variables provided to the trees are: □□Initial position, final position and duration of each accent group, measured in ms. and normalized to the duration of accent group and prosodic phrase. □□Order position of the accent group in the prosodic phrase, expressed both as an absolute quantity and relative to the total number of accent groups in the prosodic phrase, sentence and utterance. □□Total number of accent groups in the prosodic phrase, sentence and in the whole utterance. □□Total number of syllables in the current prosodic phrase, sentence and utterance. □□Position of the stressed syllable of each accent group, measured in ms. from the beginning of the prosodic phrase and from the start of the accent group. □□Type of sentence which in the case of Bodo database can have the values of declarative, question, exclamation or pause sentence. □□Type of accent, which depends on the position of the stressed syllable within the accent group. □□Type of accent command. Accent commands have been classified into three groups: last command of a question or exclamation sentence, penultimate command of a question or exclamation sentence and any other command. □□Index of accent command, which indicates the number of accent commands that are left until the end of the utterance. Another aspect that has to be considered is the variable the trees are predicting: depending on the distribution of this variable, efficiency of the prediction varies. For the pulse parameters, amplitude is predicted without any transformation, but duration is normalized to the duration of accent group, and position is given relative to the beginning of the accent group and normalized to its duration.

B).Prediction variables for phrase commands

For the phrase command amplitude prediction, the information used is related with the prosodic phrase whose intonation has to be modeled by this command. The variables provided to the tree are: □□Order position of the prosodic phrase in the sentence, given both as an absolute number and relative to the number of prosodic phrases in the sentence. □□Total number of prosodic phrases in the corresponding sentence and utterance. □□Duration of the prosodic phrase, measured in ms. □□Duration of the utterance measured in ms. □□Total number of accent groups and syllables in the prosodic phrase, the sentence and in the whole utterance. The number of accent groups in the prosodic phrase is also given normalized to the total number of accent groups in the sentence and the utterance. □□Position of the first stressed syllable of the prosodic phrase, indicated from the beginning of the corresponding prosodic phrase, expressed as an absolute quantity and relative to the duration of the prosodic phrase. □□Type of sentence and utterance.

Discussion

For each parameter of accent commands i.e. amplitude, position and duration, a binary regression tree has been built. For the phrase commands only one tree is built to predict their amplitude. In all figures PP stands for prosodic phrase and AG for accent group. The importance given by the tree to the different variables provided for accent command parameter prediction, when estimating accent command amplitudes. In this case, the type of sentence is the most influential factor, being the amplitude of commands bigger in questions and exclamations than in declaratives and pause sentences. Then type and index of accent command are considered with greater amplitudes assigned to ultimate and penultimate commands of questions and exclamations than to the others.

The importance of the variables when predicting accent command duration. The most important variable is number of syllables in the prosodic phrase and the number of syllables in the sentence has also great influence, being the predicted command longer for long phrases.

The importance of the variables provided to predict accent command position. The most important variable is type of accent command: last commands of questions and exclamations are predicted farther from the beginning of the accent group. The rest of variables have little influence compared to this one.

The importance of the variables when predicting phrase command amplitude is displayed. In this case, the most important variable is the order position of the prosodic phrase, being the command bigger for the first and second prosodic phrases than for the rest. The following three variables are related with the length of the prosodic phrase and indicate that the shorter the prosodic phrase is, the smaller the amplitude of the phrase command will be.

Conclusions

Standard Basque intonation has been modeled according to Fujisaki’s intonation model. Parameter of this model is automatically calculated for a new corpus specially designed for the purpose of modeling intonation. These parameters are related to linguistic characteristics of the corpus and the model has been introduced in our TTS system BodoTTS.

Acknowledgements

We acknowledge Prof. P. H. Talukdar , Professor Department of Instrumentation & USIC for continuous co-operation to complete this paper and we also thankful to Gauhati University for giving us the opportunity to do the research.

References

[1] Hernáez, I.; Navas, E.; Murugarren, J.L.; Etxebarria, B., 2001. Description of the AhoTTS System for Basque Language. *4th ISCA Tutorial & Research Workshop on Speech Synthesis*.

[2] Fujisaki, H.; Hirose, K., 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of Acoustic Society. Jpn*. (E) 5, 4.

[3] Mixdorff, H.,1998. *Intonation patterns of German-modelbased quantitative analysis and synthesis of F0 contours.* PhD Thesis. Technische Universitat Dresden.

[4] Wang, Ch.; Fujisaki, H.; Ohno, S.; Kodama, T., 1999. Analysis and synthesis of the four tones in connected speech of the standard Chinese based on a commandresponse model. In *Proceedings of Eurospeech'99,* Budapest, pp. 1655-1658.

[5] Navas, E.; Hernáez, I.; Armenta, A.; Etxebarria, B.; Salaberria, J., 2000. Modelling Basque intonation using Fujisaki's model and CARTs. *State of the art in speech synthesis digest, 3/1-3/6.*

[6] Griffin, D.; Lim, J. S., 1988. Multiband excitation vocoder.*IEEE Trans. ASSP*. Vol 36, N 8.

[7] Breiman, L.; Friedman, J.H.; Olsen, R.A.; Stone, C. J., 1984. Classification and Regression Trees. *Chapman & Hall.*