INFORMATICA, 2005, Vol. 16, No. 2, 193–202

Diphone Databases for Lithuanian Text-to-Speech Synthesis

Pijus KASPARAITIS

Department of Computer Science II, Faculty of Mathematics and Informatics, Vilnius University
Naugarduko 24, 03225 Vilnius, Lithuania
e-mail: pkasparaitis@yahoo.com

Received: May 2004

Abstract. One of the components of the text-to-speech synthesis system is the database of sounds. Two Lithuanian diphone databases in the MBROLA format are presented in this paper. The list of phonemes and the list of diphones necessary for Lithuanian text-to-speech synthesis are described. The problem of phoneme combinations that are not used in the Lithuanian language is dealt with in the work. Also, the article is concerned with transcribing a Lithuanian text.

Key words: text-to-speech synthesis, diphone.

Introduction

In his previous articles (Kasparaitis, 2001a; Kasparaitis, 2001b) the present author described the Lithuanian text-to-speech synthesizer “Aistis” in which 480 phonetic units of various lengths were used: parts of phonemes, allophones, diphthongs and mixed diphthongs, however, most often consonants with the beginnings of vowels and allophones of vowels are made use of. This method has some disadvantages, the most important one being the problems related to changing the length of sounds. It is most convenient to remove or add a piece of the sound from/to the most stable part of the sound, however, this leads to new connections of sounds (connections in the stable part plus connections of the phonetic units). Moreover, when putting together a consonant having the beginning of a vowel and a vowel, the exact length of the vowel is not known because information about the transition points is not available. The use of units of a different length makes the formulation of rules defining the duration of sounds more complicated. The transcription rules are rather complicated too.

One of the ways to avoid the above-mentioned problems is to change the principles of creating the database of phonetic units. Syllables, half-syllables, diphones, triphones, subphonemes or combinations of them can be used in text-to-speech synthesis (Bhaskararao, 1994). Diphones (segments taken from the middle of one sound to the middle of another one) were first proposed in (Peterson *et al.*, 1958) and half-syllables in (Fujimura and Lovins, 1978). At the moment the unit-selection method (Hunt and Black, 1996) that uses a large number of variants of the same phonetic unit is very popular but the methods

where the single example of the phonetic unit is used are still popular. Seeking to modify the sounds and to make their connections smoother various methods can be applied: TD-PSOLA (Moulines and Charpentier, 1990), MBROLA (Dutoit, 1997) and so on. Creating the database of diphones allow to avoid the problems mentioned above. In this case a piece of a signal can be removed or added only from the end or the beginning of the phonetic unit, which precludes the appearance of new discontinuities. The transition points between the sounds are stored in the database, therefore the duration of sounds can be calculated exactly.

1. List of Phonemes in Lithuanian

Seeking to create the diphone database of a certain language, first and foremost, it is necessary to draw up a list of the phonemes in that language. There are 58 phonemes in the Lithuanian language (Girdenis, 1995):

/a/, /a:/, /e/, /e:/, /ė:/, /i/, /i:/, /ie/, /o/, /o:/, /u/, /u:/, /uo/, /b/, /b"/, /d/, /d"/, /g/, /g"/, /k/, /k"/, /p/, /p"/, /t/, /t"/, /c/, /c"/, /č/, /č"/, /dz/, /dz"/, /dž/, /dž"/, /f/, /f"/, /x/, /x"/, /h/, /h"/, /s/, /s"/, /š/, /š"/, /z/, /z"/, /ž/, /ž"/, /j"/, /l/, /l"/, /m/, /m"/, /n/, /n"/, /r/, /r"/, /v/, /v"/.

Long vowels are marked with a colon, soft consonants – with a double quote.

In Lithuanian short syllables can be stressed and unstressed, while the long syllables – unstressed, stressed with the falling and with the rising accent. The vowel within the nucleus of a syllable changes due to stressing. The stressed sounds can be created from the unstressed ones during synthesis (by modifying their duration and the fundamental frequency), or the stressed sounds can be stored separately. The synthesised speech should sound more natural when using the second method. Furthermore, the use of the unstressed phonemes only is problematic when creating the database of sounds. During the creation of the database one can experiment only with the cut out, however, still unmodified sounds. Therefore it is not clear whether the unstressed sound could be cut out in such a manner that upon modifying it a stressed sound of a high quality could be obtained. Thus the following 29 units have been added to the list of phonemes:

/a`/, /a:´/, /a:˜/, /e`/, /e:´/, /e:˜/, /ė:´/, /ė:˜/, /i`/, /i:´/, /i:˜/, /i´e/, /ie˜/, /o`/, /o:´/, /o:˜/, /u`/, /u:´/, /u:˜/, /u´o/, /uo˜/, /l˜/, /l"˜/, /m˜/, /m"˜/, /n˜/, /n"˜/, /r˜/, /r"˜/.

The short stressed vowels are marked with the character ’`’, the character ’´’ means the falling accent and the character ’˜’ means the rising accent.

Though it has been stated in literature (Girdenis, 1995) that Lithuanian compound diphthongs from the phonological point of view are built from a vowel and the consonant /j/ or /v/, e. g., /ai/=/a/+/j/, /ei/=/e/+/j/, /au/=/a/+/v/, etc. it has been established by experiments that the right sides of these diphthongs differ slightly from the said consonants. Also, these right sides of the diphthongs are the same before a soft and a hard consonant, however, they differ in the stressed and the unstressed syllable. Therefore the following 4 units were added to the list: /j/, /j˜/, /w/, /w˜/.

If a pause has been added to the above-mentioned list, the latter would contain 92 units (phonemes).

2. Naming Conventions of Sounds

Seeking to avoid problems related to coding of the Lithuanian letters, as well as to bring the notations closer to those used in the international projects, e.g., SAMPA (`http://www.phon.ucl.ac.uk/home/sampa/index.html`), a list of names of phonemes (proposed by prof. A. Girdenis) containing only capital and small Latin letters, the underscore and a double quote has been drawn up. The final list of 92 phonemes is as follows:

/_/, /a/, /e/, /i/, /o/, /u/, /A/, /E/, /I/, /O/, /U/, /aa/, /ea/, /ee/, /ie/, /ii/, /oo/, /uo/, /uu/, /Aa/, /Ea/, /Ee/, /Ie/, /Ii/, /Oo/, /Uo/, /Uu/, /aA/, /eA/, /eE/, /iE/, /iI/, /oO/, /uO/, /uU/, /p/, /p"/, /t/, /t"/, /k/, /k"/, /b/, /b"/, /d/, /d"/, /g/, /g"/, /s/, /s"/, /z/, /z"/, /S/, /S"/, /Z/, /Z"/, /ts/, /ts"/, /tS/, /tS"/, /dz/, /dz"/, /dZ/, /dZ"/, /x/, /x"/, /h/, /h"/, /f/, /f"/, /v/, /v"/, /w/, /W/, /j"/, /j/, /J/, /l/, /l"/, /L/, /L"/, /m/, /m"/, /M/, /M"/, /n/, /n"/, /N/, /N"/, /r/, /r"/, /R/, /R"/.

Here the short unstressed vowels are denoted with a single small letter, e.g., /a/, the short stressed vowels are denoted with a single capital letter, e.g., /A/, the long unstressed vowels are denoted with a double small letter, e.g., /aa/, the long vowels stressed with the falling accent are denoted with a capital and a small letter, e.g., /Aa/, and the long vowels stressed with the rising accent are denoted with a small and a capital letter, e.g., /aA/. The long vowels /e:/ are denoted with a combination of two letters /ea/, whereas the long vowel /ė:/ – with /ee/. Soft (palatalised) consonants are marked with a double quote, e.g., /b"/, the consonants /š/ and /ž/ are denoted with /S/ and /Z/, respectively, and the affricates /c/ and /č/ are denoted with /ts/ and /tS/, respectively.

The second part of diphthongs, e.g., /ai/ or /au/, is denoted with /j/ and /w/ when it is unstressed and with /J/ and /W/ when stressed. The second part of the mixed diphthongs is denoted with a capital letter when stressed, e. g., /L/, /M/, /N/, /R/.

3. List of Diphones in Lithuanian

8464 combinations may be built on the basis of 92 phonemes, however, it is impossible to have all the diphones made from these combinations in the Lithuanian language. E.g., there are 28 fricatives and stop consonants (14 voiced and 14 unvoiced ones). The voiced consonants can precede only the voiced ones. The pause and the unvoiced consonants can be preceded by the unvoiced consonants only. So we have only $14 \times 14 + 14 \times (14 + 1) = 406$ combinations instead of $28 \times (28 + 1) = 812$.

The situation is similar when talking about hard and soft consonants. There are more impossible combinations allowing the number of diphones to be reduced, however, it should not be forgotten that the combinations, which are impossible in the middle of a word, can appear at the boundary of two words, e. g., /k/-/f/ in the phrase *kiek fotografų*, /U/-/U/ in the phrase *du ubagus*, and so on. Sometimes it is difficult to find a word (or a phrase) containing a certain diphone and to determine whether such a diphone is possible on the whole. Therefore in examining the list two directions were followed, that is, having found a word containing a diphone, the diphone was transferred from a general list to the

list of possible diphones, or rules were formulated, and the diphones that obeyed those rules were transferred to the list of the impossible diphones. For the diphones that were left on a general list either words containing these diphones or new rules were formulated.

The following rules were formulated:

1. A hard consonant cannot precede a soft consonant or a front vowel (e.g., /e/). A soft consonant cannot precede a hard consonant, a pause or a back low vowel (e. g., /a/). On the basis of this rule 1936 diphones have been rejected.
2. Only certain combinations of vowels and consonants form diphthongs and mixed diphthongs. The combinations that cannot form diphthongs and mixed diphthongs should be excluded. (818 diphones have been rejected).
3. A voiced consonant cannot precede an unvoiced consonant and vice versa (290 diphones have been rejected).
4. A word cannot end in a soft consonant before a back high vowel (e.g., /u/). Moreover, a vowel or a diphthong cannot precede the vowel /ea/ (95 diphones have been rejected).
5. A word cannot end in a vowel stressed with the falling accent (176 diphones have been rejected).
6. A sibilant consonant cannot precede a hushing consonant (16 diphones).
7. The vowels /Aa/, /Ea/, /Uu/ cannot precede some vowels and consonants (86 diphones).
8. Some other rules.

Some diphones follow even several rules. The rules were applied in turn and the diphone that has been rejected on the basis of one rule was not checked by other rules.

4. Problem of Non-existent Diphones in Text-to-Speech Synthesis

A list containing 5003 diphones was drawn up and the 3461 combinations were assumed to be impossible. However, to apply the diphone database to text-to-speech synthesis it is necessary to know what to do in case the transcription program should demand a non-existent diphone. The synthesizer must be able to synthesise, however, a person listening to a synthesised speech must understand that something is wrong. This might happen due to several reasons:

1) the transcription program can make mistakes;
2) transcription can be done manually. The man can make a mistake or he can deliberately introduce an impossible combination wishing to listen to how such a combination sounds;
3) any sequence of letters can be put into the input of the transcription program. Sometimes it is impossible to transcribe this sequence producing only existing diphones.

Usually this problem is solved by replacing a non-existent diphone with an existing one.

5. List of Replacements of Diphones

Each of 3461 non-existent diphones must be assigned one existing diphone for replacement. Often a non-existent diphone can be replaced with several existing diphons, e.g., /b-b"/ can be replaced with /b-b/ and with /b"-b"/. One diphone, which seems more appropriate, must be chosen.

Hence, a list of replacements has been drawn up following the below-presented rules:

1. A hard consonant before a soft consonant and a front vowel (e.g., /e/) must be replaced with a soft consonant, e.g., diphone /b-b"/ must be replaced with /b"-b"/.
2. A soft consonant before a hard consonant, a pause and a back low vowel (e.g., /a/) must be replaced with a hard consonant, e.g., /b"-b/ with /b-b/.
3. The second part of a diphthong or a mixed diphthong following a pause must be replaced with an appropriate consonant, e.g., /_-w/ with /_-v/, /_-L/ with /_-l/.
4. The stressed second part of a mixed diphthong (consonant) following the vowel, which cannot be used for building a mixed diphthong is replaced with an appropriate consonant, e.g., /uo-L/ with /uo-l/. The second part of a diphthong following a vowel, if such a combination cannot build a diphthong, is replaced with an appropriate vowel, e.g., /uo-W/ with /uo-U/. If such a combination could be used to built a diphthong, any part of a diphone can be replaced so as to build an existing diphthong, e.g., /o-W/ with /O-w/.
5. The second part of a diphthong or a mixed diphthong following a consonant is replaced with an appropriate vowel or a consonant, e.g., /b-w/ with /b-u/, /b-L/ with /b-l/. In case the left part of a diphone is also the second part of a diphthong or a mixed diphthong, it is replaced with an appropriate vowel or a consonant too, e.g., /L-L/ with /l-l/.
6. An unvoiced consonant preceding a voiced consonant is replaced with a voiced one, e.g., /p-b/ replaced with /b-b/.
7. A voiced consonant preceding an unvoiced consonant and a pause is replaced with an unvoiced one, e.g., /b-p/ with /p-p/.
8. The vowel /ea/ following a vowel is replaced with /e/, e.g., /aa-ea/ with /aa-e/.
9. The stressed soft second part of a mixed diphthong (consonant) preceding a back high vowel is replaced with a hard one, e.g., /L"-oo/ with /L-oo/.
10. The falling accent of a long vowel at the end of a word (preceding a pause or a stressed vowel) is changed to the rising one, e.g., /Aa-Aa/ replaced with /aA-Aa/. Also, the accent of the vowels /Aa/, /Ea/, /Uu/ preceding some consonants is changed to the rising accent. The same is true for the vowels /Ea/, /Uu/ preceding certain unstressed vowels, e.g., /Aa-p/ replaced with /aA-p/.
11. A sibilant consonant preceding a hushing consonant is replaced with a hushing one, e.g., /s-S/ with /S-S/.
12. The syllabic consonant /j"/ preceding a consonant is replaced with the non-syllabic consonant /j/, e.g., /j"-b/ with /j-b/.

This list of replacements of diphones ensures that the synthesizer is able to pronounce any sequence of phonemes produced by the transcription algorithm.

6. Lithuanian Diphone Databases

On the basis of the above-mentioned list containing 5003 diphones, two databases of Lithuanian diphones were created. First and foremost, texts containing all necessary diphones were prepared for this purpose. Texts were read out loud, and speech signals were recorded by a computer in a digital form. A special program "Diphone Studio" (`http://www.fluency.nl/dstudio/dstudio.zip`) designed to create diphone databases was used to cut the speech segments out of the recordings. Having put the list of phonemes into the input of the program, it generates a list of diphones. Then a word (or a phrase) can be recorded (or a prepared record can be taken) for each diphone. The program enables the beginning and the end of the diphone, as well as transition from one phoneme to another to be marked by means of a mouse. The program also makes it possible to listen to the diphone that has already been marked, and when a sufficient number of diphones has already been marked, to write a sequence of the names of diphones and to synthesise a word or a phrase according to it and to listen to them. The texts were prepared and the cutting of the diphones was made by prof. A. Girdenis. The voice of prof. A. Girdenis was used in creating one database, whereas the other one was set up of the recordings of the voice of announcer G. Deksnys.

Such databases can already be used in a simple speech synthesizer. Attention should be drawn to the fact that after a text has been transcribed into phonemes the names of phonemes should be replaced with the names of diphones, e.g., the list of phonemes *l, a, b, a, s* should be replaced with the list of diphones *_-l, l-a, a-b, b-a, a-s, s-_* and then the non-existent diphones should be replaced with the existing ones. The signal can be synthesised simply by putting diphones together or a certain smoothing algorithm can be used. The problems might arise in case after the replacement of diphones it becomes necessary to join a voiced and unvoiced segment in the middle of the phoneme.

Seeking to achieve that the sounds present in these databases could be modified (the duration and the fundamental frequency be changed) and thus to create conditions for experimenting with the various models of duration and intonation, the decision was made to join the MBROLA project (`http://tcts.fpms.ac.be/synthesis/mbrola.html`).

7. The Aim of MBROLA Project

The initiator of the project is the TCTS Laboratory of the Faculté Polytechnique de Mons (Belgium). The Project is aimed at creating a set of speech synthesizers and accumulating databases of speech sounds of as many languages as possible. These databases are distributed free of charge for a non-commercial and non-military application. Thereby it is sought to encourage research into speech synthesis, and particularly research into prosody (duration and intonation of sounds) generation.

8. Operation of MBROLA Synthesizer

The MBROLA synthesizer based on the concatenation of diphones (original technology worked out at TCTS laboratory) was created within the framework of the MBROLA project. It takes a list of phonemes as input, together with durations (in milliseconds) of phonemes and a piecewise linear description of pitch. The latter are defined as a set of pairs of numbers where the first denotes the distance from the beginning of the sound in per cent and the second – the height of the fundamental frequency (in hertz). E.g.,

```
_ 100
r 87 0 120 80 122
a 123 60 96
s 150 10 90
_ 101
```

Here the symbols "_," "r", "a", "s" are names of sounds, figures 100, 87, 123, 150 and 101 define the duration of sounds in milliseconds, figures 0, 80, 60 and 10 define the distance from the beginning of the sound in percent and figures 120, 122, 96, 90 – the height of the fundamental frequency. The synthesizer produces speech signals of 16 bits, the sampling rate of which is determined by the sampling rate of the diphone database used.

9. Lithuanian MBROLA Databases

In October 2003 the first two Lithuanian MBROLA diphone databases were published: "lt1" (the voice of prof. A. Girdenis) and "lt2" (the voice of G. Deksnys). These databases together with the necessary software can be downloaded from the Internet (`http://tcts.fpms.ac.be/synthesis/mbrola/mbrcopybin.html`). After installing the software the databases can be registered by means of the program "Control Panel", and the program "Mbroli" can be used to open the files having the format ".pho" and to listen to the speech synthesised using the data from these files. Three samples are provided with the Lithuanian databases.

Having properly generated the parameters of the duration and fundamental frequency curves, by means of these databases it is possible to synthesise a sound of a sufficiently high quality. This can be testified using the examples presented together with these databases. Therefore these databases alongside the MBROLA software can be used as a module for signal generation in the Lithuanian speech synthesizer of a high quality.

10. Transcription

In his previous work (Kasparaitis, 1999) the author presented the algorithm for transcribing a Lithuanian text intended for the database containing 480 phonetic units. When the database created from diphones is used, transcription can be performed in two stages: a text is transcribed into phonemes and then names of the phonemes are replaced with

the names of diphones. This enables the transcription process to be greatly simplified. In many cases it is not necessary to take into account a rather broad context, and the problems related to the context are solved by diphones themselves. E.g., the vowels /a/, /e/, /i/, /u/ preceding soft and hard consonants were stored as different phonetic units in the previous database. Now this problem is being solved at the stage of selecting a diphone rather than at the stage of transcribing, that is, in one case a diphone corresponding to the end of a vowel and the beginning of a hard consonant is taken, whereas in a second case – a diphone corresponding to the end of a vowel and the beginning of a soft consonant. Let us assume that differences in the beginning of a vowel preceding a soft and a hard consonant may be ignored. Similarly, the consonants /m/ and /n/ preceding a stressed and unstressed vowel, the consonant /n/ preceding the consonants {/g/, /k/} and all other consonants were stored as different phonemes in the previous database. All these problems can be solved at the stage of selecting the diphones rather than at the stage of transcription.

However, in using the previous database no problem of non-existent diphones arose. As has already been mentioned, the problem of non-existent diphones can be solved by means of replacements though this problem can also be solved at the stage of transcription. Suppose we need to transcribe the string of letters “ptga”. By means of the first method it is transcribed into the following sequence of phonemes: /_/, /p/, /t/, /g/, /a/, /_/, and then the phonemes are replaced with the diphones: /_-p/, /p-t/, /t-g/, /g-a/, /a-_/. Finally the non-existent diphone /t-g/ is replaced with the existing one /d-g/. By means of the second method the voicing is taken into account, therefore the result of transcription is as follows: /_/, /b/, /d/, /g/, /a/, /_/, with no non-existent diphones appearing. However, in this case voicing can involve not only two but also an unlimited number of phonemes, therefore such features as voicing, unvoicing and hushing should be transferred into the features of the current letter. This entails the reduction of the context analysed in the rules but the calculation of the features of the current letter becomes longer.

When transcribing (as in earlier cases) the boundaries of syllables, stressing, softness/hardness and the context are to be taken into account. The boundaries of syllables are important now only between vowels, since earlier they were also important between the vowels and the sonorants. Possible values of the stressing feature are as follows: *unstressed, stressed short, stressed with the falling accent, stressed with the rising accent, preceding the stressed* (used in diphthongs). The value *inside the stressed syllable* has been rejected. Now softness/hardness is important only to the consonants, earlier it was of significance to the vowels too. Earlier the context covering one letter to the left and three letters to the right was analysed, whereas now it is suffice to take one letter to the left and one letter to the right. During the context analysis it is enough to recognise the following situations only: the diphthongs, the vowel “i” playing the role of the sign of softness, the combinations of the consonants “ch”, “dz” and “dž”.

Simplification of transcription leads to reduction of the number of transcription rules. Their number has been reduced from 740 to 157 rules.

11. Application of Diphone Databases in Speech Synthesis

From the above-said it certainly became clear that the MBROLA synthesizer is not a speech synthesizer in a proper sense since not only a raw text but also phonemes and prosodic information is presented to it as input. Seeking to synthesise the Lithuanian speech by means of the MBROLA software and databases, the text, first and foremost, is to be stressed. The algorithms presented in (Kasparaitis, 2000; Kasparaitis, 2001c) can be successfully used for this purpose. Further the text is transcribed using the above-described algorithm. Then modules for generation of durations of sounds and intonation should be used. Such modules have not been created for the Lithuanian language thus far, however, the MBROLA software and the databases are a very convenient tool for creating and testing the duration and intonation models. It is worth undertaking it in the immediate future. Only after the said models have been created it is reasonable to assess the intelligibility and acceptability of the speech synthesised on the basis of diphones and to compare it with the Lithuanian speech synthesizers created earlier.

Conclusions

This work presents a list of phonemes necessary for Lithuanian text-to-speech synthesis, describes the principles of drawing up a list of diphones on the basis of the said list and the manner of solving the problem of non-existent diphones. This knowledge can be used in creating new Lithuanian diphone databases. On the basis of the above-described principles two freely distributed diphone databases have been created, which may be used as a module for speech generation when creating text-to-speech synthesizers or as a highly convenient tool when experimenting with the prosodic phenomena of the Lithuanian language. The transcription algorithm adapted to these databases is described. In the immediate future it is worthwhile to create the duration and intonation models of Lithuanian using the above-mentioned databases. This would be a large step forward in improving Lithuanian text-to-speech synthesis.

References

Bhaskararao, P. (1994). Subphonemic segment inventories for concatenative speech synthesis. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art and Future Challenges*, John Willey & Sons, Chichester, New York, Brisbane, Toronto, Singapore. pp. 63–86.

Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. Kluwer Academic Publishers, Dortrecht.

Fujimura, O., J. Lovins (1978). Syllables as concatenative phonetic elements, In A. Bell and J.B. Hooper (Eds.), *Syllables and Segments*, North-Holland, New York. pp. 107–120.

Girdenis, A. (1995). *Teoriniai fonologijos pagrindai*. Vilniaus universitetas, Vilnius.

Hunt, A., A. Black (1996). Unit selection in a concatenative speech synthesis system. In *ICASSP-96*, Atlanta, GA. pp. 373–376.

Kasparaitis, P. (1999). Transcribing of the Lithuanian text using formal rules. *Informatica*, **10**(4), 367–376.

Kasparaitis, P. (2000). Automatic stressing of the Lithuanian text on the basis of a dictionary. *Informatica*, **11**(1), 19–40.

Kasparaitis, P. (2001a). *Lithuanian Text-to-Speech Synthesis*. Doctoral thesis. Vilnius University, Vilnius.
Kasparaitis, P. (2001b). Lietuvių kalbos kompiuterinis sintezatorius “Aistis”. *Garso korta 2001*. Kaunas, Technologija.
Kasparaitis, P. (2001c). Automatic stressing of the Lithuanian nouns and adjectives on the basis of rules. *Informatica*, **12**(2), 315–336.
Moulines, E., F. Charpentier (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, **9**, 453–467.
Peterson, G., W. Wang, E. Sivertsen (1958). Segmentation techniques in speech synthesis. *J. Acoust. Soc. Am.*, **30**, 739–742.

P. Kasparaitis was born in 1967. In 1991 he graduated from Vilnius University (Faculty of Mathematics). In 1996 he has been admitted as a PhD student in Vilnius University. In 2001 he defended a thesis for a doctoral degree. Current research interests include text-to-speech synthesis and other areas of computer linguistic.

Difonų bazės lietuvių kalbos sintezei

Pijus KASPARAITIS

Vienas iš sintezės pagal tekstą komponentų yra garsų bazė. Šiame darbe pristatomos dvi MBROLA formato lietuvių kalbos difonų garsų bazės. Aprašomas lietuvių kalbos sintezei reikalingų fonemų ir difonų sąrašas. Nagrinėjama lietuvių kalboje nesutinkamų fonemų derinių problema. Taip pat aptariamas lietuviško teksto transkribavimo uždavinys.