# **Object-Oriented Access to the Estonian Phonetic Database**

Einar Meister (1), Arvo Eek (1),

Toomas Altosaar (2), Martti Vainio (3)

(1) Institute of Cybernetics, Tallinn Technical University Akadeemia tee 21, Tallinn 12618, Estonia Einar.Meister@ioc.ee

(2) Acoustics Lab., Helsinki University of Technology Otakaari 5A, SF-02150 Espoo, Finland Toomas.Altosaar@hut.fi

(3) Department of Phonetics, University of Helsinki P.O.Box 35, FIN-0014 Helsinki, Finland Martti.Vainio@helsinki.fi

#### Abstract

The paper introduces the Estonian Phonetic Database developed at the Laboratory of Phonetics and Speech Technology of the Institute of Cybernetics at the Tallinn Technical University, and its integration into QuickSig – an object-oriented speech processing environment developed at the Acoustics Laboratory of the Helsinki University of Technology. Methods of database access are discussed, relations between different speech units – sentences, words, phonemes – are defined, examples of predicate functions are given to perform searches for different contexts, and the advantage of an object-oriented paradigm is demonstrated. The introduced approach has been proven to be a flexible research environment allowing studies to be performed in a more efficient way.

### Introduction

The Estonian Phonetic Database developed within the BABEL-project is the first database of Estonian designed in a well-defined manner and recorded under controlled conditions. The database includes valuable speech material for research on Estonian phonetics and speech technology. To gain flexible access to the phonetic information available within the database different access mechanisms and models for representing speech have been compared (Altosaar, Millar, Vainio, 1999). The object-oriented paradigm implemented in the QuickSig system has shown an advantage over the traditional string-matching paradigm and has therefore been chosen as the experimental environment.

The paper provides an insight into the Estonian Phonetic Database and its integration into the QuickSig object-oriented environment.

#### **Estonian Phonetic Database**

The Estonian Phonetic Database (EPDB) has been developed within the EU COPERNICUS project no. 1304 BABEL: A Multi-Language Database (Roach et al., 1998) from 1995 to 1998. In the project six partners from Western Europe and six partners from Central and Eastern Europe were involved. In parallel with Estonian the databases for Bulgarian, Hungarian, Polish and Romanian were collected.

The major tasks of the project were:

- to compile a text corpus of Estonian;
- to recruit 70 speakers of Standard Estonian according to demographic distribution;
- to establish the recording environment and procedures following the ESPRIT-SAM project;

- to make recordings with selected speakers divided into three subsets – Many Talker Set, Few talker Set, Very Few Talker Set;
- to verify the quality of recorded signals done by a Polish partner;
- to adapt the SAMPA phonemic transcription for Standard Estonian;
- to annotate approx. 15% of the recorded signals;
- to produce CD-ROMs with signal and label files, and documentation.

# **Text corpus**

The Estonian text corpus is based on the design of EUROM-1 (Chan et al., 1995) with minor modifications to the number of blocks, guaranteeing that all main phonologically relevant oppositions will be revealed in the text corpus. It includes:

- CVC-blocks representing word-initial and word-final consonants and interconsonantal short/long vowels;
- Number blocks representing the number system of the language;
- Filler sentences representing the patterns of quantity degrees;
- Passages representing intonational, phrasal and foot structures.

#### Speakers

There are about one million speakers of Estonian speaking different regional variants. However, Standard Estonian is based on the language spoken in northern Estonia, particularly in the capital city Tallinn. All students in secondary school have to use Standard Estonian, whatever their geographical location.

For the recordings, only speakers using Standard Estonian were selected. All speakers were tested for reading ability, articulation, voice quality and pathologies. At the beginning of the recording session every speaker was instructed on reading style and given the opportunity to preview the texts.

Age	Number of speakers	
	Female	Male
16-20	3	4
20-29	11	9
30-39	7	9
40-49	7	8
50-70	7	5

Table 1. Distribution of speakers by age and gender.

The utterances are divided into three major sets:

- Very Few Talker Set: 1 male, 1 female; each to read 9 blocks of CVC syllables, 2 blocks of 50 numbers, 4 blocks of filler sentences, and 40 passages;
- Few Talker Set: 4 males, 4 females; each to read 2-3 blocks of CVC syllables, 1 block of 50 numbers, 1 block of filler sentences, and 10 passages;
- Many Talker Set: 30 males, 30 females; each to read 1 block of CVC syllables, 1 block of 50 numbers, 1 block of filler sentences, and 1-2 passages.

### Recordings

The recordings were carried out in a soundproof room using a Sony ECM 44B microphone and the OROS AU21 board (16 bit) at a sampling rate of 20kHz. The recording procedure has been controlled by the EUROPEC 4.0 software and monitored by an operator. The auditory quality control has been carried out at the end of the recording sessions.

### Annotation

The text corpus is presented in Estonian orthography and in the SAMPA phonemic transcription (Wells, 1995) as well as adapted for Standard Estonian (Eek and Meister, 1998). In order to select from SAMPA a specific machine-readable symbol set for Estonian, a list of phonological units of Standard Estonian has been compiled. The Standard Estonian vowel system comprises 9 vowels:

and the consonant system comprises 17 consonants:

p, t, k, t', f, v, s, S, h, s', m, n, n', l, l', r, j.

Phonemes with // are palatalised.

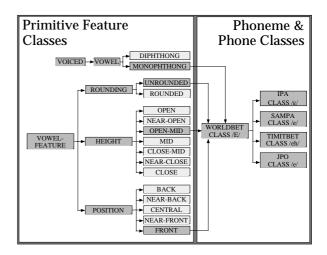
All phonemes can occur single (short) or doubled (long). The latter has been applied for the phonemic (or broad phonetic) labelling of the 80 passages of the Many Talker Set. The labelling has been carried out manually in the SFS environment (Huckvale, 1996) according to the common labelling criteria developed by project partners.

# Dissemination

The database has been stored on three CD-ROMs including in addition to signal and label files the text corpus, speakers' data, and a description of the Estonian sound system. The Estonian Phonetic Database together with the other BABEL databases will be made available soon by the European Language Resources Association (ELRA).

# QuickSig

QuickSig (Karjalainen, Altosaar, Alku, 1988) is an objectoriented DSP programming environment implemented in Common Lisp. QuickSig (QS) consists of object classes and method functions that enable the definition of signal processing algorithms via the class inheritance mechanism of CLOS (Common Lisp Object System). QS is extendible by the user and new algorithms can be defined and tested immediately due to incremental compilation. QS includes an object-oriented database extension which allows for storage and updating of arbitrary data types and relations of speech signals, transcriptions, segments and related information in a flexible manner (Karjalainen and Altosaar. 1993). Through an object-oriented implementation of the Worldbet phonetic alphabet (Hieronymus, 1993) a common formalism for representing the sounds of any language is available.



**Figure 1.** Class hierarchy used to define the Worldbet class /E/. Other derived phonetic alphabets are also shown.

QS is able to process speech databases of differing languages and phonetic alphabets in a generic environment. QS has proven to be a flexible research environment allowing the speech scientist to perform studies in an intuitive, logical and motivating manner (Altosaar et al., 1998).

# **Database Access in QuickSig**

The efficient use of data available in speech databases depends on the methods of database access and available query operations. No definite standards exists for the format of databases but typically, speech databases comprise a collection of signal files and associated files with orthographic or phonetic transcriptions, as well as segment boundaries. Additional files may provide information related to the recording conditions, speaker identity, dialect, etc. Database search and access is typically performed using rigid string-matching paradigms based on a textual representation of speech.

A more advanced approach is based on an object-oriented paradigm where compiled structural object hierarchies are constructed for database utterances and searches performed by object matching functions to seek out desired contexts. This approach is implemented in the QuickSig speech database environment. To gain easier access to the phonetic knowledge within EPDB, the database was converted into an object-oriented database (OODB) within QS. Annotation objects were

created such as segments, phones, words, and sentences for the phonetic domain while linguistic words and sentences were generated from orthographic data.

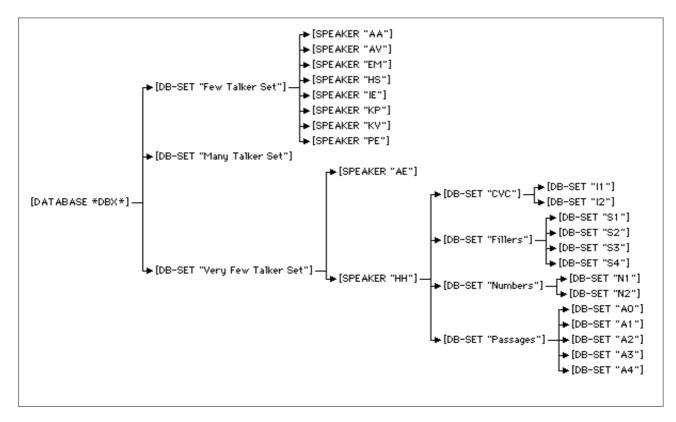


Figure 2. A part of the Estonian Phonetic Database as viewed with the database grapher.

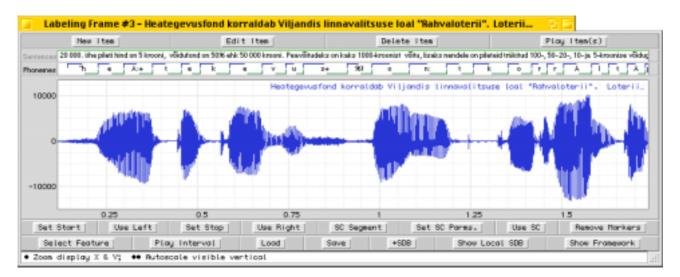


Figure 3. An example of a labelling frame in the QuickSig environment.

Utilising Worldbet's class hierarchy and its derived phonetic classes enables us to perform searches using the efficient class inheritance mechanism of Lisp and CLOS. A library of low level functions, relations, and types is used to construct predicate functions. The following form defines a predicate function to search for short /i/ vowels:

A search returns a list of actual phone objects that match the predicate function, which can be operated upon immediately by a wide variety of signal processing operations available in the QS environment. In a similar way different user defined predicate libraries can easily be defined and applied, depending on the specific needs of concrete research tasks. Searches are efficient and utilise Lisp's powerful type checking infrastructure.

# **Examples of Applications**

# **Phonetic research**

In the following example we use the above predicate to search for matching instances in the EPDB. We apply the predicate to two different female speakers and find that speaker "ES" has 29 matches and speaker "ER" has 23. We then calculate the average auditory spectrum of all occurrences of the short /i/ for both speakers. These spectra are shown in figure 2.

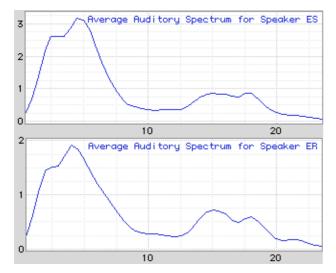
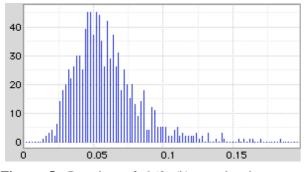


Figure 4. Results of database query. Average auditory spectrum of two speakers constrained to short /i/ context.

The same predicate can also be used to study the duration distribution of the vowel over part of the database. In figure 5 the duration of 960 /i/ vowels is shown as a histogram (occurrence vs. duration).



**Figure 5.** Duration of 960 /i/ vowels shown as a histogram.

### Speech Technology

The database in the QS environment provides a flexible framework for research on speech synthesis, recognition and speaker recognition, as well. Some examples can be found in (Altosaar et al., 1998), the results of the pilot experiments on speaker recognition using neural nets are reported in (Altosaar and Meister, 1995).

# Conclusion

The Estonian Phonetic Database has been developed as the basic resource for studies on Estonian phonetics and for speech technology developments. Integrating the EPDB into the QS signal processing environment enables the researchers to gain flexible access to the phonetic information available within the database and to perform studies in a more efficient way.

# Acknowledgements

The Estonian Phonetic Database has been created within the COPERNICUS project no. 1304, the further developments are supported by the Estonian Science Foundation (grant no. 1454) and by the Estonian national programme "Estonian Language and National Culture".

#### References

- Altosaar, T., Millar, B., Vainio, M. (1999) Relational vs. Object-Oriented Models for Representing Speech: A Comparison Using ANDSOL Data. – Proceedings of EUROSPEECH'99, Budapest, vol. 5: 915-918.
- Roach, P., Arnfield, S., Barry, W., Dimitrova S., Boldea M., Fourcin A., Gonet W., Gubrynowicz R., Hallum E., Lamel L., Marasek K., Marchal A., Meister E., Visci K. (1998) BABEL: A Database of Central and Eastern European Languages. Proceedings of the First International Conference on Languages Resources and Evaluation, Granada, Spain, May 28 -30, 1998, vol. 1: 371-374.
- Fourcin, A., Dolmazon, J.-M. (1991) Speech knowledge, standards and assessment. – Proceedings of XII International Congress of Phonetic Sciences, Aix-en-Provence, vol. 5: 430-433.
- Chan D., Fourcin, A. and others (1995) EUROM A Spoken Language Resource for the EU. – Proceedings of EUROSPEECH-95, Madrid, vol. 1: 867–870.
- Wells, J. C. (1995) Computer-coding the IPA: a proposed extension of SAMPA. University College London.
- Eek, A., Meister, E. (1998) Estonian Speech in the BABEL Multi-Language Database: Phonetic-Phonological Problems Revealed in the Text Corpus. – Workshop on Speech Database Development for Central and Eastern European Languages, Granada, Spain, May 17,1998.
- Huckvale, M. (1996) Speech Filing System: SFS for Users. University College London.
- Karjalainen, M., Altosaar, T., Alku, P. (1988) QuickSig An Object-Oriented Signal Processing Environment. – Proceedings of IEEE ICASSP-88, New York.
- Karjalainen, M., Altosaar, T. (1993) An Object-Oriented Database for Speech Processing. – Proceedings of EUROSPEECH-93, Berlin.
- Hieronymus, J.L. (1993) ASCII Phonetic Symbols for the World's Languages: Worldbet. Bell Labs Technical Memorandum.
- Altosaar, T., Karjalainen, M., Vainio, M., Meister, E. (1998) Finnish and Estonian Speech Applications Developed on an Object-Oriented Speech Processing and Database System. – Workshop on Speech Database Development for Central and Eastern European Languages, Granada, Spain, May 17,1998.
- Altosaar, T., Meister, E. (1995) Speaker Recognition Experiments in Estonian using Multi-Layer Feed-Forward Neural Nets. – Proceedings of EUROSPEECH-95, Madrid, vol.1: 333-336.