



# PURCHASE CONTRACT

between

**ELDA (Evaluations and Language resources  
Distribution Agency)**

and

**University of West Bohemia in Pilsen**

---

KC

## PURCHASE CONTRACT

(hereinafter referred to as the "Contract")

*Entered into pursuant to the Commercial Code, Act no. 513/1991, Coll., Section 409 et sequentes, as amended*

### I.

#### Contracting Parties

**Purchaser:** University of West Bohemia in Pilsen  
Address: Univerzitní 8, 306 14 Plzeň  
Represented by: doc. PaedDr. Ilona Mauritzová, Ph.D., Rector  
Bank: Komerční banka a.s., Plzeň-město  
Account no.: 4811530257/0100  
Identification no.: 497 77 513  
Tax identification no.: CZ49777513  
(hereinafter referred to as the "Purchaser") as one Contracting Party

a

**Seller:** ELDA (Evaluations and Language resources Distribution Agency)  
Address: 55-57 rue Brillat-Savarin, 75013 Paris, France  
Represented by: Dr. Khalid Choukri, Managing Director  
Bank: BNP PARIBAS  
Account no.: 00020064648/14  
Identification no.: 402 781 876 R.C.S. PARIS  
Tax identification no.: FR 53 402781 876  
Registered in the commercial register of "Registre du Commerce et des Sociétés", section  
Greffe du Tribunal de Commerce de Paris, insert 15 November 1995, Number 017099  
(hereinafter referred to as the "Seller") as the other Contracting Party

(together also referred to as the "Contracting Parties")

Enter into the following Contract based on the result of the Tender for the Public Contract entitled  
"Delivery of Speech Databases for the NTIS Project":

KC  
10/1



## II.

### Subject-Matter of the Contract

- 2.1. The Seller undertakes, within the scope, and pursuant to the terms and conditions of the Contract herein, to supply a total of **eight (8) speech databases – language resources** that will enable training of acoustic and language models of a multi-lingual system for automatic speech recognition, on an electronic medium, for example, a CD, including a **Written report for every single database** showing particularly that each database has been subject to a complex validation process that guarantees, for each of the databases, to contain data in the format and structure required by the Purchaser (hereinafter referred to as the “Goods”), to the Purchaser, and to grant the rights to use to the Goods to the Purchaser. Together with the Goods, the Seller undertakes to transfer to the Purchaser the title to exercise the rights given by the permission (license) – granted without any time or territorial limits – to conduct internal research and development using data contained in the Goods, and also the rights to use the said data for creation of products and services that the Purchaser is authorized to further distribute or sell. **Annex no. 1**, which forms an integral part of this Contract, gives specifications of the Goods in full detail.
- 2.2. The Seller’s commitment also includes transportation of the Goods to the place of delivery pursuant to Article III herein (including possible insurance of the Goods during transportation).
- 2.3. The Purchaser undertakes to accept the Goods supplied by the Seller under the terms and conditions specified in Article II herein, and to pay the stipulated purchase price for the Goods to the Seller using the payment method, and by the due date, stipulated herein.
- 2.4. Each and every time the term “Goods“ is used in the Contract herein, it shall also mean all specific goods (the individual databases).

## III.

### Time and Place of Delivery, Use of Software

- 3.1. The Seller undertakes to deliver the Goods to the Purchaser and to fulfil its obligations pursuant the Article II of the Contract herein **in accordance with the schedule below:**

#### November 2013

- database for development of telephone applications for the Slovak language
- database for development of dictation applications for the English language – native speakers, British English
- database for development of dictation applications for the English language – non-native speakers
- database for development of dictation applications for the Russian language
- database for development of dictation applications for the English language – native speakers, American English
- database for development of telephone applications for the English language

#### June 2014

- database for development of telephone applications for the Polish language
- database for development of telephone applications for the Russian language

- 3.2. Simultaneously, the Contracting Parties agree that the terms stated in the schedule contained in Section 3.1 herein are only presumed terms for the delivery of specific Goods – the individual databases; the Purchaser reserves the right to accept the Goods gradually, depending on the availability of its funding, during the period of 2013-2014.
- 3.3. The Contracting Parties have agreed that the partial orders will be executed as follows:



The Purchaser will request, in writing, the Seller to deliver the specific Goods – the individual databases; with the written call containing at least:

- i) the specification of the particular Goods – the individual databases to be delivered
- ii) the deadline for the Goods delivery, which shall not exceed 30 calendar days following the receipt of the written call for delivery by the Seller.

Should the Seller be in delay fulfilling obligations specified in the sentence above, the Purchaser will be entitled to request the Seller to pay a contractual penalty amounting to 0.2% of the purchase price of the specific Goods for every breach of duty, and for every, even incomplete, day of the delay; the right of the Purchaser to request damages will not be prejudiced.

- 3.4. The Contracting Parties have agreed herein to write a **Memorandum of Transfer** for delivery of each specific Goods – a single database; the Memorandum will contain at least a brief description of each database, the date of delivery and acceptance of the specific Goods, and signatures of authorized representatives of both Contracting Parties.
- 3.5. The **place of delivery** is the University of West Bohemia in Pilsen, Univerzitní 22, Plzeň, Czech Republic; or, if appropriate or needed, any other place according to the instructions of the Purchaser.
- 3.6. On the day of transfer of the Goods to the Purchaser, the Seller undertakes, pursuant to the Contract herein, to transfer to the Purchaser the title to exercise the rights given by the **permission (licence)** delivered together with the Goods (each database), within the scope necessary for the proper and full use of the Goods, i.e. permission to conduct internal research and development using the Goods (the data), and also the right to use the said data for creation of products and services that the Purchaser is authorized to further distribute or sell. The licence is granted without any time and territorial limits; the compensation for the granted license is included in the purchase price of the Goods. The Purchaser is not obliged to use this license.
- 3.7. The Seller declares that it is authorized to grant such a license, that there are no limitations imposed by third parties – natural or legal persons – and that the relevant licence is not subject to any rights of third parties pursuant to the Copyright Act no. 121/2000, Coll., as amended, or to any other law. Should the rights of the Purchaser be limited by a third party, or should a third party prevent the Purchaser from proper exercise of these rights, the Seller agrees and undertakes to stop such behaviour at its own expense and to compensate the Purchaser for possible damage.
- 3.8. Should the declaration stated by the Seller in Section 3.7 of the document herein turn out to be untrue, or the license is, contrary to the declaration of the Seller, invalid or insufficient, the Purchaser will be entitled to charge the Seller a contractual penalty of EUR 4,000; this will not prejudice the Purchaser's right to claim damages. In such a case, the Purchaser will also be entitled to request the Seller to obtain the appropriate license additionally; if the Seller fails to meet such an obligation within 30 calendar days after receiving the request from the Purchaser, the Purchaser will be entitled to withdraw from the Contract.

#### IV.

#### Price and Payment Terms

- 4.1. The purchase price for the Goods was stipulated based on an agreement of the Contracting Parties and results from the price bid submitted by the Seller calculated for the purposes of the Tender concerning the delivery of the subject-matter of this contract.
- 4.2. The Purchaser undertakes to pay the Seller for the supplied Goods the agreed **total purchase price for the subject-matter of the delivery (Databases) amounting to:**

**325,700,- € excluding VAT (in words: Three Hundred Twenty Five Thousand Seven Hundred EUR excluding VAT);**



The total purchase price for the subject-matter of the delivery (Databases) is calculated on the basis of the unit purchase price for the delivery of the specific Goods (individual databases); the unit price for each single database is stated in the following table:

Type of Database	No. of units	Unit price in EUR (€) excluding VAT
Database for development of telephone applications for the Slovak language	1 piece	16,000 €
Database for development of telephone applications for the Polish language	1 piece	16,000 €
Database for development of telephone applications for the Russian language	1 piece	25,000 €
Database for development of telephone applications for the English language	1 piece	60,000 €
Database for development of dictation applications for the Russian language	1 piece	67,000 €
Database for development of dictation applications for the English language – native speakers, British English	1 piece	67,000 €
Database for development of dictation applications for the English language – native speakers, American English	1 piece	67,000 €
Database for development of dictation applications for the English language – non-native speakers	1 piece	6,850 €

KC

14/



Extra charges for the delivery of the databases	No. of units	Unit price in EUR (€) excluding VAT
ELRA Membership fee	1 piece	750 €
Shipment fee	1 piece	100 €

- 4.3. The purchase price is stipulated as the highest possible price including all fees and any other costs associated with the supply of the Goods pursuant to this Contract. The price includes especially transportation costs as well as relevant insurance. The purchase price also includes any administrative charges, taxes (except VAT), duties, import and export surcharges, approval procedure costs, required testing, compliance statements, certificates, transfer of rights, insurance, transportation charges, etc.
- 4.4. The purchase price for delivery of Goods under the Contract will be paid to the Seller in EUR on the basis of a tax document (invoice) issued by the Seller for each delivery of specific Goods (individual databases).
- 4.5. The Contracting Parties have agreed that the Purchaser will not make any advance payments for the Goods to the Seller.
- 4.6. The purchase price for delivery of Goods under the Contract will be paid to the Seller gradually, always following a proper delivery (handover) of the specific Goods (individual databases) together with a Memorandum of Transfer, duly signed by both Contracting Parties; also following the due fulfilment of all obligations of the Seller stipulated in Articles 2.1 and 2.2 of the Contract herein in relation to the specific Goods.
- 4.7. In order to meet all requirements under the Contract herein, a copy of a Memorandum of Transfer and Acceptance of the specific Goods, duly signed by the authorized representatives of both Contracting Parties, must be enclosed with the invoice for the delivered Goods.
- 4.8. A tax document (the invoice) must contain all elements of a proper accounting and tax document pursuant to the relevant legislation, namely the Value Added Tax Act no. 235/2004, Coll., as amended. The Purchaser will be entitled to return an invoice missing the relevant elements to the Seller to be amended before its due date, and the Purchaser will not be in delay with the payment in such a case. The payment period will start again after the duly amended or corrected invoice is received by the Purchaser.
- 4.9. The invoice will be due 30 days after it is demonstrably received by the Purchaser. However, the Purchaser reserves the right to unilaterally postpone the due date of the tax document depending on the availability of funding released from the Operational Programme RDI for up to 30 calendar days. Late allocation of funding from the Operational Programme RDI will not be considered as the Purchaser's default in payment of the purchase price pursuant to the Contract herein.
- 4.10. The Contracting Parties agreed on the following procedure for postponing the payment in case the Purchaser lacks sufficient funding due to the reason stated in article 4.9 herein: the Purchaser will be obliged to inform the Seller about the existence of such circumstances without undue delay. The due date of the tax documents will be extended by 30 calendar days from the day when such a notice is delivered to the Seller.



- 4.11. Should the Purchaser be in default with payment of the invoice, the Seller will be entitled to request an interest on late payment of 0.05% of the amount due for every, even incomplete, day of the delay from the Purchaser. The interest on late payment will not be charged if the payment is postponed due to late allocation of funding from the Operational Programme RDI.
- 4.12. The Purchaser will be entitled to off-set any contractual penalties that the Seller is obliged to pay against the invoiced amount.

## V.

### Rights and Obligations of the Contracting Parties

- 5.1. The Seller will be obliged to deliver the Goods in the agreed quantity, quality and design. All the Goods supplied to the Purchaser according to this Contract must meet the qualitative requirements as stipulated herein.
- 5.2. The Seller will be obliged to supply to the Purchaser Goods free of any defects and according to the terms and conditions stipulated herein. The Memorandum of Transfer certifying the handover of Goods may not be signed before a complete delivery of Goods is supplied to the Purchaser, including relevant associated tasks and services stipulated herein. The Purchaser will be entitled to reject defective Goods.
- 5.3. The Seller will be obliged to supply exhaustive technical documentation necessary for the use of the Goods to the Purchaser together with the Goods; the said documentation will be in the English or Czech language, in electronic form.
- 5.4. The Purchaser assumes the right to use the Goods on the day the Goods is properly delivered and accepted by the Purchaser on the grounds of a signed Memorandum of Transfer confirming the receipt of Goods. Risk of damage to the Goods is assumed by the Purchaser on the same day.
- 5.5. The Seller is obliged to immediately inform the Purchaser about possible risk that the deadline will not be met, and about any circumstances that might prevent the Goods from being delivered.
- 5.6. The Seller is not entitled to assign any rights or obligations arising from this Contract to a third party without obtaining a previous written consent of the Purchaser.
- 5.7. The Seller agrees that any receivables that may be claimed from the Purchaser and that come into existence based on this Contract may not be assigned or set off by a unilateral legal act.
- 5.8. The Seller will be liable to the Purchaser for any damage caused by violation of duties stipulated in this Contract or obligations stipulated by generally binding legal regulations.
- 5.9. The Contracting Parties agreed, and the Seller determined, that the person authorized to act on behalf of the Seller in matters relating to the Contract herein and its execution is:

Name: Khalid Choukri

E-mail: [choukri@elda.org](mailto:choukri@elda.org)

Phone: +33 1 43 13 33 33

- 5.10. The parties agreed, and the Purchaser determined, that the person authorized to act on behalf of the Purchaser in matters relating to the Contract herein and its execution is:

Name: Vlasta Radová

Email: [radova@kky.zcu.cz](mailto:radova@kky.zcu.cz)

Phone: +420 377 63 2547



All correspondence, instructions, notices, requests, notes and other documents made out by the Contracting Parties based on this Contract or in relation to it will be made out in writing in the English language and will be delivered either in person or by registered mail, fax or e-mail to the mailing addresses of the personnel authorized pursuant to this Contract and to their attention.

- 5.11.** The Seller acknowledges that, pursuant to the Financial Control in Public Administration Act no. 320/2001, Coll., Section 2 (e), the Seller is a person obliged to cooperate in exercising financial control. A controlling body of the Operational Programme RDI will be authorized to perform an audit and access those parts of the bid, Contracts and related documents that are subject to protection according to special legal regulations (such as trade secrets, proprietary information) for a period of three years after the operational programme is closed, provided that the requirements defined by the law (for example, the State Control Act no. 552/1991, Coll., Section 11 (c) and (d) and Section 12, subsection 2 (f)), are met). The controlling body of the Operational Programme RDI is similarly authorized to perform an audit at the Seller's subcontractors, as the case may be (see Annex no. 2 to the Guidelines for OP RDI Applicants and Beneficiaries – Contractor Selection Guidelines <http://www.msmt.cz/file/14585>).
- 5.12.** The Seller will be obliged to properly maintain all documents relating to the execution of the subject-matter of the Contract, including accounting books, for at least three years after the OP RDI project is closed, i.e. at least until 2022. During the same time period, the Seller will be obliged to provide requested information and documents to the employees or agents of bodies authorized to audit projects performed under OP RDI and will be obliged to create conditions for these personnel to be able to carry out the audit of the project implementation and cooperate in the audit.
- 5.13.** The Seller will be obliged to observe all the commitments that had been stated in its bid submitted to the tender before this Contract was signed.
- 5.14.** The Seller acknowledges and agrees that this Contract will be made public on the Purchaser's website pursuant to the Public Contracts Act, Section 147 (a), and that the actual price paid for the supply of the subject-matter of the Contract will also be made public by the deadlines and in a manner defined by the Public Contracts Act, Section 147 (a). Pursuant to the Public Contracts Act, Section 147a, subsections (4) and (5), the Seller is obliged to provide the Purchaser with a list of subcontractors following the time limits and scopes specified therein. If in breach of their statutory duties stipulated in the Public Contracts Act, Section 147a, subsections (4) and (5), the Contractor will be fully liable to the Purchaser for any damage caused by violation of the duties.

## VI.

### Guarantee on the Goods

- 6.1.** The Seller assumes a guarantee on the Goods for the period of 24 (Twenty-Four) months.. The guarantee period starts on the day the Goods is delivered to the Purchaser, i.e. on the day the Memorandum of Transfer confirming delivery and receipt of the Goods is signed.
- 6.2.** The claims for any defects covered by the guarantee may be raised by the last day of the guarantee period at the latest; also claims mailed on the last day of the guarantee period will be considered duly raised claims.



## VII.

### Validity and Effect of the Contract

- 7.1. The Contract becomes valid and effective on the day it is signed by authorized representatives of the two Contracting Parties.
- 7.2. The Contract can only be terminated for reasons stipulated in the Contract or specified by law.
- 7.3. A Contracting Party affected by the other party's failure to meet its obligation may unilaterally terminate this Contract for a gross violation of the Contract; in particular, the following will be considered a gross violation of this Contract:
- Purchaser's failure to pay the purchase price in accordance with this Contract within 60 days after the due date of a relevant invoice,
  - Seller's failure to duly deliver the Goods, or even part of the Goods, on the agreed date,
  - Seller's failure to supply Goods possessing the characteristics declared by the Seller in this Contract,
  - Seller's failure to provide correct information or documents indicative of the actual situation in the Bid that had been submitted to the Tender before this Contract was signed, which have or could have affected the result of the Tender.
- 7.4. Once the Contract becomes ineffective, all and any obligations of the Contracting Parties arising from this Contract will cease to exist. The ineffectiveness or termination of the Contract will not affect the existence of the right to claim damages and contractual penalties in case of violation of the contractual obligations existing before the Contract becomes ineffective, and obligations of the Contracting Parties that are to survive this Contract by their nature or by law.

## VIII.

### Closing Provisions

- 8.1. The relationship between the Parties will be governed by Czech law. Unless expressly stipulated otherwise in the Contract, the relationships that come into existence based on this Contract and arising from this Contract are governed by the relevant provisions of the Commercial Code, Act no. 513/1991, Coll., as amended, and other applicable laws.
- 8.2. All and any changes and amendments to the Contract may only be made by a written agreement of the Contracting Parties and must respect provisions of the Public Contracts Act, Section 82, subsection 7. Such agreements must be in the form of numbered amendments to the Contract signed by both Parties and dated.
- 8.3. If any of the Parties is unable to duly perform this Contract due to circumstances preventing it from doing so, the Party will be obliged to inform the other Party about the fact without undue delay and initiate a meeting of the Purchaser's and Seller's representatives.
- 8.4. If a reason causing one provision to be invalid only affects a certain provision of the Contract, only this particular provision will be invalid, unless its nature or the circumstances under which it was stipulated indicate that it is not severable from the rest of the Contract.
- 8.5. The Contracting Parties will always strive to reach amicable settlement of possible disputes arising from the Contract. Possible disputes arising from this Contract will be resolved in accordance with the valid laws by local authorities with appropriate competences in the Czech Republic. In accordance with the Civil Procedure Code, Act no. 99/1963, Coll., Section 89(a), as amended, the Contracting Parties have agreed that disputes arising from this Contract or relating to this Contract will be resolved by the district or regional court in Plzeň.



- 8.6. The Contract is made out in four (4) identical copies, each of which is valid as original. Each of the Contracting Parties will receive two (2) identical copies.
- 8.7. Annex no. 1 – Detailed Technical Specification of the Goods forms an integral part of the Contract.
- 8.8. The Contracting Parties declare that they have read the Contract before signing, and agree with its content without any reservations. The Contract expresses their sincere, real, free and serious will. To demonstrate that these statements are authentic and true, the authorized representatives of the Contracting Parties attach signatures in their own hand.

In Pilsen

Date

- 5 - 12 - 2013

In Paris

Date

20 November 2013

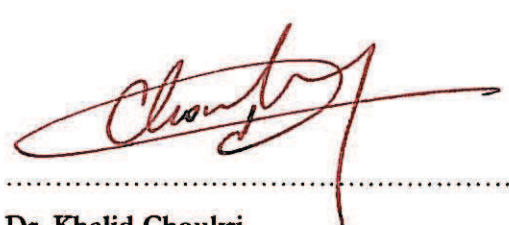
On behalf of the Purchaser:

**University of West Bohemia in Pilsen**

On Behalf of the Seller:

**ELDA (Evaluations and Language  
resources Distribution Agency)**

  
.....  
**doc. PaedDr. Ilona Mauritzová, Ph.D.**  
Rector

  
.....  
**Dr. Khalid Choukri**  
Managing Director





## Annex no. 1 to the Contract – Detailed Technical Specification of the Goods

### Database for development of telephone applications for the Slovak language:

#### ELRA Reference and Name: ELRA-S0095 Slovak SpeechDat(E) Database

This database comprises 1000 Slovak speakers (498 males, 502 females) recorded over the Slovak fixed telephone network. This database is partitioned into 5 CDs.

This speech database was validated to assess its compliance with the format and content specifications.

The speech files are stored as sequences of 8-bit, 8kHz A-law speech files and are not compressed. Each prompt utterance is stored within a separate file and has an accompanying ASCII SAM label file.

Corpus contents:

- 6 application words;
- 1 sequence of 10 isolated digits;
- 4 connected digits: 1 sheet number (5 digits), 1 telephone number (9-11 digits), 1 credit card number (16 digits), 1 PIN code (6 digits);
- 3 dates: 1 spontaneous date (birthday), 1 prompted date (word style), 1 relative and general date expression;
- 1 spotting phrase using an application word (embedded);
- 1 isolated digit;
- 3 spelled-out words (letter sequences): 1 spontaneous e.g. own forename; 1 spelling of directory assistance city name; 1 real/artificial name for coverage;
- 2 currency money amounts: 1 Slovak money amount, 1 International money amount (USD, EURO)
- 1 natural number;
- 6 directory assistance names: 1 spontaneous, e.g. own forename; 1 city of birth / growing up (spontaneous); 1 most frequent city (out of 500); 1 most frequent company/agency (out of 500); 1 "forename surname" (set of 150), 1 "surname" (set of 150)
- 2 questions, including "fuzzy" yes/no: 1 predominantly "yes" question, 1 predominantly "no" question;
- 12 phonetically rich sentences;
- 2 time phrases: 1 time of day (spontaneous), 1 time phrase (word style);
- 4 phonetically rich words.

The following age distribution has been obtained: 39 speakers are below 16 years old, 446 speakers are between 16 and 30, 253 speakers are between 31 and 45, 214 speakers are between 46 and 60, and 48 speakers are over 60.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

### Database for development of telephone applications for the Polish language:

#### ELRA Reference and Name: ELRA-S0090 Polish SpeechDat(E) Database

This database comprises 1000 Polish speakers (488 males, 512 females) recorded over the Polish fixed telephone network. This database is partitioned into 5 CDs, each of which comprises 200 speakers sessions.

This speech database was validated to assess its compliance with the format and content specifications.

The speech files are stored as sequences of 8-bit, 8kHz A-law speech files and are not compressed. Each prompt utterance is stored within a separate file and has an accompanying ASCII SAM label file.

Corpus contents:

- 6 application words;
- 1 sequence of 10 isolated digits;
- 4 connected digits: 1 sheet number (5 digits), 1 telephone number (8-11 digits), 1 credit card number (15-16 digits), 1 PIN code (6 digits);
- 3 dates: 1 spontaneous date (birthday), 1 prompted date (word style), 1 relative and general date expression;
- 1 spotting phrase using an application word (embedded);



- 1 isolated digit;
- 3 spelled-out words (letter sequences): 1 spelling of surname; 1 spelling of directory assistance city name; 1 real/artificial name for coverage;
- 2 currency money amounts: 1 Polish money amount, 1 International money amount (USD, EURO)
- 1 natural number;
- 6 directory assistance names: 1 surname (out of 500); 1 city of birth / growing up (spontaneous); 1 most frequent city (out of 500); 1 most frequent company/agency (out of 500); 1 "forename surname" (set of 150), 1 "surname" (set of 150)
- 2 questions, including "fuzzy" yes/no: 1 predominantly "yes" question, 1 predominantly "no" question;
- 12 phonetically rich sentences;
- 2 time phrases: 1 time of day (spontaneous), 1 time phrase (word style);
- 4 phonetically rich words.

The following age distribution has been obtained: 9 speakers are below 16 years old, 428 speakers are between 16 and 30, 291 speakers are between 31 and 45, 254 speakers are between 46 and 60, and 18 speakers are over 60.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

#### **Database for development of telephone applications for the Russian language:**

##### **ELRA Reference and Name: ELRA-S0099 Russian SpeechDat(E) Database**

This database comprises 2500 Russian speakers (1242 males, 1258 females) recorded over the Russian fixed telephone network. This database is partitioned into 13 CDs.

The speech files are stored as sequences of 8-bit, 8kHz A-law speech files and are not compressed. Each prompt utterance is stored within a separate file and has an accompanying ASCII SAM label file.

Corpus contents:

- 6 application words;
- 1 sequence of 10 isolated digits;
- 4 connected digits: 1 sheet number (5 digits), 1 telephone number (9-10 digits), 1 credit card number (14-16 digits), 1 PIN code (6 digits);
- 3 dates: 1 spontaneous date (birthday), 1 prompted date (word style), 1 relative and general date expression;
- 1 spotting phrase using an application word (embedded);
- 1 isolated digit;
- 3 spelled-out words (letter sequences): 1 spelling of surname; 1 spelling of directory assistance city name; 1 real/artificial name for coverage;
- 2 currency money amounts: 1 Russian money amount, 1 International money amount (USD, EURO)
- 1 natural number;
- 6 directory assistance names: 1 spontaneous (own forename); 1 city of birth / growing up (spontaneous); 1 most frequent city (out of 500); 1 most frequent company/agency (out of 500); 1 "forename surname" (set of 150), 1 "surname" (set of 150)
- 2 questions, including "fuzzy" yes/no: 1 predominantly "yes" question, 1 predominantly "no" question;
- 9 phonetically rich sentences;
- 2 time phrases: 1 time of day (spontaneous), 1 time phrase (word style);
- 4 phonetically rich words.

The following age distribution has been obtained: 10 speakers are below 16 years old, 854 speakers are between 16 and 30, 858 speakers are between 31 and 45, 679 speakers are between 46 and 60, 34 speakers are over 60, and 65 speakers are of unknown age.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

#### **Database for development of telephone applications for the English language:**

##### **ELRA Reference and Name: ELRA-S0242 SALA II US English Database**

This database collected in the United States contains the recordings of 4,090 US English speakers (2,017 males and 2,073 females, including some speakers with Hispanic accents) recorded over the United States mobile telephone network.



The following acoustic conditions were selected as representative of a mobile user's environment (some speakers were recorded in several environments):

- Passenger in moving car, railway, bus, etc. (607 speakers)
- Public place (1,238 speakers)
- Stationary pedestrian by road side (928 speakers)
- Home/office environment (1,188 speakers)
- Passenger in moving car using a hands-free kit (161 speakers)

This database is distributed as 2 DVD-ROMs. The speech files are stored as sequences of 8-bit, 8kHz Mu-law speech files and are not compressed. Each prompt utterance is stored within a separate file and has an accompanying ASCII SAM label file.

This speech database was validated to assess its compliance with the format and content specifications.

Each speaker uttered the following items:

- 6 application words (out of a set of 30)
- 1 sequence of 10 isolated digits
- 4 connected digits (1 sheet number -5+ digits, 1 telephone number -9/11 digits, 1 credit card number -14/16 digits, 1 PIN code -6 digits)
- 3 dates (1 spontaneous date e.g. birthday, 1 word style prompted date, 1 relative and general date expression)
- 1 spotting phrase using an embedded application word
- 2 isolated digits
- 3 spelled words (1 surname, 1 directory assistance city name, 1 real/artificial name for coverage)
- 1 currency money amount
- 1 natural number
- 5 directory assistance names (1 spontaneous, e.g. own surname, 1 city of birth/growing up, 1 most frequent city out of a set of 500, 1 most frequent company/agency out of a set of 500, 1 "forename surname" out of a set of 150)
- 2 yes/no questions (1 predominantly "yes" question, 1 predominantly "no" question, including fuzzy questions)
- 9 phonetically rich sentences
- 2 time phrases (1 spontaneous time of day, 1 word style time phrase)
- 4 phonetically rich words

The following age distribution has been obtained: 129 speakers are under 16, 2,456 speakers are between 16 and 30, 832 speakers are between 31 and 45, 610 speakers are between 46 and 60, and 63 speakers are over 60.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

#### **Database for development of dictation applications for the Russian language:**

##### **ELRA Reference and Name: ELRA-S0161 Russian Speecon Database**

This database is divided into 2 sets:

1. The first set comprises the recordings of 550 adult Russian speakers (271 males, 279 females), recorded over 4 microphone channels in 4 recording environments (office, entertainment, car, public place).
2. The second set comprises the recordings of 50 child Russian speakers (31 boys, 19 girls), recorded over 4 microphone channels in 1 recording environment (children room).

This database is partitioned into 21 DVDs (first set) and 3 DVDs (second set).

This speech database was validated to assess its compliance with the format and content specifications. Each of the four speech channels is recorded at 16 kHz, 16 bit, uncompressed unsigned integers in Intel format (lo-hi byte order). To each signal file corresponds an ASCII SAM label file which contains the relevant descriptive information.

Each speaker uttered the following items:

- Calibration data:
  - o 6 noise recordings
  - o The "silence word" recording
- Free spontaneous items (adults only):
  - o 5 minutes (session time) of free spontaneous, rich context items (story telling) (an open number of



spontaneous topics out of a set of 30 topics)

- 17 Elicited spontaneous items (adults only):

- o 3 dates
- o 2 times
- o 3 proper names
- o 2 city names
- o 1 letter sequence
- o 2 answers to questions
- o 3 telephone numbers
- o 1 language
- Read speech:
  - o 30 phonetically rich sentences uttered by adults and 60 uttered by children
  - o 5 phonetically rich words (adults only)
  - o 4 isolated digits
  - o 1 isolated digit sequence
  - o 4 connected digit sequences
  - o 1 telephone number
  - o 3 natural numbers
  - o 1 money amount
  - o 2 time phrases (T1 : analogue, T2 : digital)
  - o 3 dates (D1 : analogue, D2 : relative and general date, D3 : digital)
  - o 3 letter sequences
  - o 1 proper name
  - o 2 city or street names
  - o 2 questions
  - o 2 special keyboard characters
  - o 1 Web address
  - o 1 email address
  - o 208 application specific words and phrases per session (adults)
  - o 74 toy commands and 48 general commands (children)

The following age distribution has been obtained:

- Adults: 290 speakers are between 15 and 30, 187 speakers are between 31 and 45, 73 speakers are between 46 and 60.

- Children: 28 speakers are between 8 and 10, 22 speakers are between 11 and 14.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

**Database for development of dictation applications for the English language – native speakers, British English:**

**ELRA Reference and Name: ELRA-S0215 UK English Speecon Database**

This database is divided into 2 sets:

1) The first set comprises the recordings of 606 adult UK English speakers (325 males, 281 females), recorded over 4 microphone channels in 4 recording environments (office, entertainment, car, public place).

2) The second set comprises the recordings of 51 child UK English speakers (14 boys, 37 girls), recorded over 4 microphone channels in 1 recording environment (children room).

This database is partitioned into 31 DVDs (first set) and 4 DVDs (second set).

This speech database was validated to assess its compliance with the format and content specifications.

Each of the four speech channels is recorded at 16 kHz, 16 bit, uncompressed unsigned integers in Intel format (lo-hi byte order). To each signal file corresponds an ASCII SAM label file which contains the relevant descriptive information.

Each speaker uttered the following items (over 290 items for adults and over 210 items for children):

- Calibration data:

- o 6 noise recordings
- o The "silence word" recording



- o Free spontaneous items (adults only):
  - o 5 minutes (session time) of free spontaneous, rich context items (story telling) (an open number of spontaneous topics out of a set of 30 topics)
  - 17 Elicited spontaneous items (adults only):  
3 dates, 2 times, 3 proper names, 2 city names, 1 letter sequence, 2 answers to questions, 3 telephone numbers, 1 language
  - Read speech:
    - o 30 phonetically rich sentences uttered by adults and 60 uttered by children
    - o 5 phonetically rich words (adults only)
    - o 4 isolated digits
    - o 1 isolated digit sequence
    - o 4 connected digit sequences
    - o 1 telephone number
    - o 3 natural numbers
    - o 1 money amount
    - o 2 time phrases (T1 : analogue, T2 : digital)
    - o 3 dates (D1 : analogue, D2 : relative and general date, D3 : digital)
    - o 3 letter sequences
    - o 1 proper name
    - o 2 city or street names
    - o 2 questions
    - o 2 special keyboard characters
    - o 1 Web address
    - o 1 email address
  - 208 application specific words and phrases per session (adults)
  - 74 toy commands, 14 phone commands and 34 general commands (children)
- The following age distribution has been obtained:
- Adults: 321 speakers are between 16 and 30, 182 speakers are between 31 and 45, 103 speakers are over 46.
  - Children: All 51 speakers are between 11 and 14.
- A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

**Database for development of dictation applications for the English language – native speakers, American English:**

**ELRA Reference and Name: ELRA-S0233 US English Speecon Database**

This database is divided into 2 sets:

- 1) The first set comprises the recordings of 550 adult US English speakers (286 males, 264 females), recorded over 4 microphone channels in 4 recording environments (office, entertainment, car, public place).
- 2) The second set comprises the recordings of 50 child US English speakers (21 boys, 29 girls), recorded over 4 microphone channels in 1 recording environment (children room).

This database is partitioned into 25 DVDs (first set) and 3 DVDs (second set).

This speech database was validated to assess its compliance with the format and content specifications. Each of the four speech channels is recorded at 16 kHz, 16 bit, uncompressed unsigned integers in Intel format (lo-hi byte order). To each signal file corresponds an ASCII SAM label file which contains the relevant descriptive information.

Each speaker uttered the following items (over 290 items for adults and over 210 items for children):

- Calibration data:
  - o 6 noise recordings
  - o The “silence word” recording
- Free spontaneous items (adults only):  
5 minutes (session time) of free spontaneous, rich context items (story telling) (an open number of



spontaneous topics out of a set of 30 topics)

- 17 Elicited spontaneous items (adults only):

- o 3 dates
- o 2 times
- o 3 proper names
- o 2 city names
- o 1 letter sequence
- o 2 answers to questions
- o 3 telephone numbers
- o 1 language
- Read speech:
  - o 30 phonetically rich sentences uttered by adults and 60 uttered by children
  - o 5 phonetically rich words (adults only)
  - o 4 isolated digits
  - o 1 isolated digit sequence
  - o 4 connected digit sequences
  - o 1 telephone number
  - o 3 natural numbers
  - o 1 money amount
  - o 2 time phrases (T1 : analogue, T2 : digital)
  - o 3 dates (D1 : analogue, D2 : relative and general date, D3 : digital)
  - o 3 letter sequences
  - o 1 proper name
  - o 2 city or street names
  - o 2 questions
  - o 2 special keyboard characters
  - o 1 Web address
  - o 1 email address
  - o 208 application specific words and phrases per session (adults)
  - o 74 toy commands, 14 phone commands and 34 general commands (children)

The following age distribution has been obtained:

- Adults: 217 speakers are between 15 and 30, 204 speakers are between 31 and 45, 129 speakers are over 46.
- Children: 18 speakers are between 8 and 10, and 32 speakers are between 11 and 14.

A pronunciation lexicon with a phonemic transcription in SAMPA is also included.

**Database for development of dictation applications for the English language – non-native speakers:**

**ELRA References and Names:**

- ELRA-S0249 TC-STAR English Training Corpora for ASR: Recordings of EPPS Speech
- ELRA-S0251 TC-STAR English Training Corpora for ASR: Transcriptions of EPPS Speech

This corpus consists of the recordings of around 290 hours from EPPS (European Parliament Plenary Sessions) speeches held or interpreted in European English (a mixture of native and non-native English), 92 hours of which were annotated (transcribed). The recordings were obtained from Europe by Satellite (<http://europa.eu.it/comm/ebs>) from May 2004 until May 2006.

The speech signals were submitted by EbS via internet in Real Media format and via satellite in MPEG1-layer2 format. The signals were decoded, resampled and are stored in WAVE RIFF (Resource Interchange File Format). Each file contains a single channel with 16-bit resolution at a sample rate of 16kHz.

The transcribed corpus consists of 63 transcriptions files. The transcription files are stored in Transcriber XML file format.

This speech database was validated to assess its compliance with the format and content specifications.