

SPEAKER RECRUITMENT METHODS AND SPEAKER COVERAGE – EXPERIENCES FROM A LARGE MULTILINGUAL SPEECH DATABASE COLLECTION

Børge Lindberg (1), Robrecht Comeyne (2), Christoph Draxler (3) & Francesco Senia (4)

(1) CPK, Aalborg University, Denmark [lindberg@cpk.auc.dk]

(2) Lernout & Hauspie Speech Products NV, Belgium [Robrecht.Comeyne@lhs.be]

(3) IPSK, University of Munich, Germany [draxler@phonetik.uni-muenchen.de]

(4) CSELT, Italy [Francesco.Senia@cslt.it]

ABSTRACT

With the globalisation and evolving technology of voice-driven man-machine interfaces there is a growing demand for acquisition of spoken language resources in a number of speaker populations being representative for a number of languages and countries.

In this paper experience from work within a large consortium in creating large multilingual speech databases for tele-services are reported. In particular the methods and experiences in recruiting speakers for such recordings are reported across a number of participating partners. The reporting is from the SpeechDat project (<http://speechdat.phonetik.uni-muenchen.de>).

1. INTRODUCTION

Within the European Union (EU) funded SpeechDat project the objectives are to create large-scale speech databases for voice driven teleservices. Eleven European languages are covered as well as some major dialectal variants and minority languages (Höge et al., 1997; Draxler et al., 1998). In total, 28 databases are being collected: 20 databases are recorded over the fixed telephone network (FDB), 5 databases over the mobile network (MDB), and 3 databases are designed for speaker verification via telephone (SDB). The recordings of the FDB and MDB databases cover between 500 and 5000 calls by different speakers.

The databases are intended to be used today and in the near future to develop a number of applications such as information services (e.g. timetable information), transaction services (e.g. home shopping, home banking) and other call processing services (e.g. voice mail handling, call centre systems).

A number of criteria are defined for the databases in order to be accepted within the consortium and a strict and thorough validation procedure is carried out by a neutral institution in order to guarantee that the spoken language resources are all of a standardised quality (Heuvel, 1997).

Speakers being recruited for a SpeechDat database have to fulfil several requirements which are related to

- speaker specific characteristics (e.g. gender, age, weight and height, socio-economic factors)
- regional and dialectal factors

- environment specific characteristics (e.g. location of call, type of handset, type of network)

Considerable efforts are therefore assigned to the non-trivial task of recruiting speakers.

In this paper the design of the SpeechDat databases is briefly presented followed by a short description of the demographic criteria for the speakers recorded in the databases. Then the various methods of speaker recruitment are presented followed by a summary of the experience obtained concerning their relative merits.

1. DATABASE DESIGN AND CREATION

Three different types of SpeechDat databases are specified for the FDB, the MDB and the SDB. These share a core of roughly 40 items (Winski, 1997; Kordi 1996; van Velden *et al.*, 1996) as listed in Table 1.

#Speakers		Utterance description
500+	4000+	
2	2	isolated digit items
4	4	digit/number strings
1+	1+	natural number
1	1	money amounts
2	2	yes/no questions
3+	3+	dates
2	2	times
3	6	application keywords/keyphrases
1	1	word spotting phrase using embedded application words
5	5	directory assistance names
3	3	spellings
4+	4+	phonetically rich words
9	9	phonetically rich sentences
40+	43+	TOTAL

Table 1: SpeechDat FDB corpus contents definition

The number of items in each category is determined by an estimated minimum number of items needed for the training of speech recognisers. For the phonetically rich sentences and words a maximum number of item repetitions is specified to obtain a high number of phoneme contexts for a good coverage of di- and triphones.

Supplementary to the common core vocabulary, partners are free to record additional material for their own purposes.

1.1 Technical set-up

All SpeechDat databases are recorded on telephone servers connected to either base rate or primary rate ISDN interfaces enabling the possibility to call a toll-free number. Each speaker is provided with a prompt sheet containing the items of Table 1. Most items will result in read speech; typically six items will result in spontaneous speech as answers to short prompted questions.

It is required that all FDB calls are unique. For the MDBs of some languages, callers are allowed to call from a maximum of four different environments. Obviously, the purpose of SDBs is to acquire multiple calls from the same speaker.

2. DEMOGRAPHIC SPECIFICATIONS

The following demographic criteria are specified to ensure a good coverage of the speaker population (Senia et al., 1997):

- 50% ($\pm 5\%$) male and female speakers
- all accent regions have to be covered proportionally
- a minimum of 20% of the speakers covering the age groups 16–30 and 31–45 years, and 15% in the age group 46–60 years
- 2% of the FDB calls have to be from a public place
- 25% of the MDB calls have to be either from home, a public place, at a street or in a moving vehicle environment.

To meet these specifications either an over-sampling approach or a close monitoring of incoming calls and returned data sheets is employed within the project.

3. SPEAKER RECRUITMENT

Speaker recruitment was the most difficult task for all partners, and it took longer than expected for most partners (Draxler et al., 1998; Moreno et al., 1998).

Five different recruitment schemes are used:

- hierarchical recruitment,
- snowball recruitment,
- direct mail or contact,
- general newspaper or magazine advertisements, and
- local radio and newspaper calls for participation.

In hierarchical recruitment, a mediator, e.g. a department level manager, distributes prompt sheets among colleagues, employees, etc.

In snowball recruitment, an already recorded speaker (the recruiter) is asked to name further people, e.g. friends or

relatives that might be willing to participate. Prompt sheets are then sent to the addresses provided by the recruiter, together with a letter explaining by whom their address was provided. An extra incentive is offered to the recruiter, e.g. a proportional number of lottery tickets, or phone cards.

In direct mail or contact, addresses are selected from a given source, e.g. a telephone directory available on a CD-ROM, and letters explaining the project together with prompt sheets are sent to these addresses. Alternatively, contacts are made via telephone calls to the persons available in the source.

General newspaper or magazine advertisements motivate readers to request prompt sheets either by phone, by letter or fax, or via the Internet. Usually, there is no attempt to control the demographics of the callers in this method. For Internet requests, prompt sheets are provided either online as WWW pages, or as PDF documents to be printed locally.

Finally, in local newspaper or radio calls for participation, local radio stations and newspapers with a regional distribution are asked to present the project and to provide an address and phone number for requesting prompt sheets.

Hierarchical and direct mail recruitment schemes are *direct*, i.e. the prompt sheets are distributed directly to potential speakers; all other schemes are *indirect*, i.e. people request sheets which are then sent to them.

Direct schemes have very little overhead and their cost is determined by the cost of the material distributed. Indirect schemes may require considerable administration overhead – keeping track of who recruited whom, who requested a prompt sheet, etc.; the cost of indirect schemes is determined by the direct cost of the material distributed, and administration cost (personnel, software, storage, space, etc.).

Table 2 shows the different recruitment schemes adopted for every language and database type recorded.

It is seen that for 7 out of the 28 databases recruitment via a market research company is used. This method seems to be the most efficient as it has not to be supplemented with other means of recruitment. Unfortunately, for many partners recruitment via a market research company on the other hand seems to be the most expensive means of getting speakers. Generally, market research companies do not guarantee a number of callers, but only the number of contacts. Further, a major problem when using a market research company is that its demographic criteria may not match those of the project (e.g. accent classification). Finally, for MDB databases there may be a too long chain of communication: SpeechDat specs are conveyed to the market research company, who are not technical specialists; they pass the information to the individual market researchers; they instruct the callers (many of whom are "technophobes"). As a result of this, many calls have missing items or are made from the wrong environment.

Usually, independent of the recruitment principle, there is an incentive to participate, e.g. a lottery is organised or a donation to charity is arranged. Alternatively phone cards, money amounts, lottery tickets or gifts are offered.

Language (variant)	DB-type and #spk.	#c	Recruitment Method
Belgian French	FDB, 1000	1	Direct contact; All speakers are recruited by tele-recruiters based on telephone directories. Reminding calls are made.
Flemish	FDB, 1000	1	
Luxemb. German	FDB, 500	1	
Luxemb. French	FDB, 500	1	
Danish	FDB, 4000	1	Local newspaper calls for participation seeding a subsequent snowball recruitment.
English	FDB, 4000	1	A market research company.
Finnish	FDB, 4000	1	Hierarchical recruitment among academia, schools, libraries and business partners. A WWW-based recruitment tool is used extensively.
Finnish Swedish	FDB, 1000	1	
French (*)	FDB, 5000	1	Direct contact using internal telephone directory. Indirect methods used: Internal newspapers and internal WWW. Direct mail/e-mail is also used.
German	FDB, 4000	1	Hierarchical recruitment, snowball recruitment and local radio and newspaper calls for participation.
Greek (*)	FDB, 5000	1	Hierarchical recruitment in large companies, public organisations, schools, universities and even from door to door.
Italian	FDB, 3000	1	Market research company
Norwegian	FDB, 1000	1	Direct mailing to addresses drawn at random from a telephone directory by a marketing agency.
Portuguese	FDB, 4000	1	Direct mailing to employees of Portugal Telecom (about 20,000) representing the required demographic distribution. Also e-mail broad-casting is used.
Slovenian	FDB, 1000	1	Direct mailing to employees of the Post of Slovenia distributed according to the predefined demographic criteria.
Spanish	FDB, 4000	1	Hierarchical recruitment among 26 universities at which a person is in charge of the recruitment of a prescribed number of speakers among students and their relatives.
Swedish	FDB, 5000	1	Direct mail based on bought addresses to people randomly spread all over Sweden. Hierarchical recruitment at phonetic institutions as well as WWW-based recruitment. Snowball recruitment is the main recruitment scheme adopted.
Swiss German	FDB, 1000	1	Direct telephone contact, using a marketing company.
Swiss French	FDB, 2000	1	
Welsh	FDB, 2000	1	Local advertisements at secondary schools and charities. WWW-announcement, and snowball recruitment. Direct mail to university employees. Face-to-face recruitment at festival and at visited companies having a significant proportion of Welsh-speaking employees.
Dutch	MDB, 250	4	Recruitment among friends, family and colleagues of institute. Snowball, advertisements and WWW-recruitment (linked to a popular WWW-page). Direct contact to 150 people who have done mobile recordings before. Reminding calls are made.
English	MDB, 1000	1	A market research company recruit speakers, primarily via "on-the-street" recruitment and in accordance to the environmental constraints. Reminding calls are made.
German	MDB, 1000	1	Snowball, direct mail, advertisements and local newspaper calls for participation.
Italian (*)	MDB, 250	4	Same as for the Italian FDB.
Swedish	MDB, 1000	1	FDB callers are asked to participate in the MDB recording as well. Also snowball recruitment is used. Gifts are used to control the environmental distribution.
English	SDB, 120	20	Same as for the English FDB, except higher amount as caller incentive.
French (*)	SDB, 120	20	Same as for the French FDB
Swiss French	SDB, 20	50	Institute members and closest collaborators are recruited.

Table 2 Speaker recruitment methods applied in the SpeechDat databases. #c denotes number of calls per speaker. (*) indicates that only preliminary information is available (for a subset of the database)

4. RESULTS

In this chapter explicit success rates (number of complete recordings acquired relative to the number of persons being recruited) are listed for the recruitment schemes, where available.

4.1 FDB Recruitment

Upon completion of a call, speakers are usually requested to fill in a reply-form and return it in a post-paid envelope. This was needed to be able to supply speaker information with the database and also to keep track of prompt sheets actually used. Experience showed a remarkable high return rate once callers

had completed: 95% (Norway), 96% (Sweden) and 95% (Denmark). Where prompt sheet tracking was used (as opposed to over-sampling) the number of different prompt sheets was the main factor limiting the recording speed. Once used, the sheets could not be sent out again to a new subject before the deadline to call back for the first subject had passed.

Direct mail. Table 2 shows that a number of partners used the direct mail principle. The success rate of this method varies from country to country: 20% for Norway, 29% for Portugal, 18% for Sweden and only 1% for Germany. Direct contacts (tele-recruiting) were made for the Swiss FDB. For Swiss French 66% agreed to call; 63% of these actually called resulting in an overall success rate of 42%. The corresponding success rate for Swiss German is reported also to be below 50%.

Snowball. The snowball scheme works quite well, in particular if the recruiter can see an immediate return. For Danish 81% of the speakers were recruited via snowball recruitment. For Swedish the corresponding figure is 75%. Some recruiters were very helpful providing more than 10 speakers. On average a recruiter provided 8.8 potential callers (for Danish) of which on average 4.8 speakers completed the call. For Swedish it was observed that if extra incentives were added to control e.g. the calling environment (by offering extra lottery tickets if the speaker called from a certain environment), the average response and completion rate was reduced. In that case the success rate of snowball recruitment was reduced from 48% to 41%. In some cases recruiters were asked explicitly to recruit speakers of e.g. a certain age-category or region. In this way an attempt was made to control the demographics during the snowball recruitment process itself.

Other methods. In local newspaper calls for participation a success rate of 48% was obtained for Danish; 1061 signed in for participation out of a target population of 500,000. In Germany the corresponding figure was 35%, where 285 signed in for participation out of a target population of 600,000.

WWW-based recruitment resulted in a success rate of 69% for Swedish, once the 526 potential speakers had signed in for participation.

The success rate for Welsh was 49% (of 3300) measured across a number of different recruitment schemes as listed in Table 2.

4.2 MDB/SDB Recruitment

For the MDB databases the task of recruiting speakers is more difficult as there are additional constraints to be fulfilled. It is therefore often difficult to find the required number of speakers, as these are caused more inconvenience (e.g. by having to make the recordings at a public place), and because the environment of their calls has to meet the specified distribution. Thus, it was observed for the Dutch MDB that only 20% of the callers already signed in (excluding friends and colleagues) actually completed the calls. Similarly, in Sweden it was experienced that snowball recruitment was less attractive (only a success rate of 22%).

Unfortunately, at the present stage of the project, insufficient information is available on the SDB-databases.

5. CONCLUSION

Apart from using a market research company it may be concluded that no single recruitment scheme is probably sufficient for large-scale data collections to be conducted within a given period of time. Instead, one should try many schemes in parallel, study the impact of each strategy and be willing to modify them during the course of the recruitment action.

It is also learned that it is utmost important to respond immediately to prompt sheet requests. If prompt sheets are sent out more than a week after being requested, the rate of return is clearly lower than when they are sent out on the next day.

There is no doubt that one of the most successful approaches to recruitment experienced in the present project is snowball recruitment. At some sites a success rate of 55% was obtained using this scheme.

In the long run, recruiting speakers via the Internet seems to be quite attractive: wide circulation in news groups and mailing lists, and on-line distribution of prompting material. However, there is a risk that the users are not representative for the total population. Also, as long as English is the dominant language in news groups and mailing lists, recruiting speakers for languages other than English languages is difficult.

6. ACKNOWLEDGEMENTS

Part of the SpeechDat project is funded by the Commission of the European Communities, Telematics Applications Programme, Language Engineering, Contract LE2-4001.

Colleagues within the SpeechDat project are gratefully acknowledged for their contributions to the present paper. The individual databases within the SpeechDat project are separately described in design documents. In some cases data has been extracted from these design documents without explicitly referencing the (long list of) particular design documents.

7. REFERENCES

1. Draxler, C., Heuvel, H. van den & Trops, H. (1998) SpeechDat Experiences in Creating Large Multilingual Speech Databases for Teleservices. In *Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain*.
2. Heuvel, H. van den (1997). Validation Criteria. *SpeechDat Technical Report SD1.3.3*.
3. Höge, H., H.S. Trops, R. Winski, H. van den Heuvel, R. Haeb-Umbach & K. Choukri (1997). European Speech Databases for Telephone Applications. In *ICASSP'97, Munich*.
4. Kordi, K. (1996). Definition of Corpus, Scripts, and Standards for Speaker Verification. *SpeechDat Technical Report SD1.1.3*.
5. Moreno, A. et al. (1998). Recruiting Speakers and Documentation on Speaker Typology. *SpeechDat Technical Report SD 2.2.1*.
6. Senia, F. et al. (1997). Environmental and Speaker Specific Coverage for Fixed Networks. *SpeechDat Technical Report SD1.2.1*.
7. Velden, J.G. van, D. Langmann, M. Pawlewski (1996). Specification of Speech Data Collection over Mobile Telephone Networks. *SpeechDat Technical Report SD1.1.2/1.2.2*.
8. Winski, R. (1997). Definition of corpus, scripts and standards for Fixed Networks. *SpeechDat Technical Report SD1.1*.