

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/76459>

Please be advised that this information was generated on 2021-06-22 and may be subject to change.

Speaker Recruitment for Speech Databases

Eric Sanders, Henk van den Heuvel
SPEX

P.O. Box 9103, 6500 HD Nijmegen, the Netherlands
eric@spex.nl

Abstract

In this paper, the aspects of speaker recruitment, that we encountered during our speech database collections, are discussed. A number of parameters that influence speaker recruitment are pointed out and strategies to reach possible participants are given. The listings are not exhaustive, as we have focussed on our own experiences.

1. Introduction

As more and more automatic speech recognition systems are being developed for a growing number of languages, the need for speech databases is growing. The production of a speech database is full of challenges. In general, one of the most difficult problems is speaker recruitment. In the SpeechDat projects [1], for example, speaker recruitment has been called "the most difficult task for all partners, which took longer than expected for most partners" [2].

This paper gives an overview of what to expect and what to consider with respect to speaker recruitment. For a large part, the issues addressed will be based on experience, since the authors were and are involved in the production of a large number of speech databases, mostly recorded over the fixed telephone network, the mobile telephone network and in cars. Each time, speaker recruitment appeared to be the bottle-neck of the project. Various methods have been used to recruit speakers and a lot was learned.

The paper is structured as follows: in the first part attention is paid to the requirements of a database production and its influence on speaker recruitment. In the second part, various strategies to recruit speakers are described with their pros and cons. Finally some thoughts are spent on rewards for speakers.

2. Parameters of database collection

Speech databases differ from each other on a number of levels. This has its influence on the strategy that is needed to recruit the speakers. Important parameters of a speech database collection are:

1. Recording conditions
2. Special tasks for speakers
3. Country/culture where recordings take place
4. Number and distribution of speakers
5. Time of recording

Each of these parameters and their effects on speaker recruitment are discussed in the following subsections.

2.1. Recording conditions

The purpose for which a speech databases is collected can vary and this has a clear impact on the recording methods, platforms and environments, which in turn has its effect on speaker recruitment. Some examples of recording methods are a 'Wizard of Oz' system [3], spontaneous dialogues, and reading from text. In Wizard of Oz recordings, people who are aware of the recording strategy are not eligible for participation, because their knowledge might influence their way of behaving (and thus speaking) towards the system. This means that colleagues and close relatives (who know about the work you are doing) cannot participate and strangers have to be recruited. For recordings of read speech, alphabetic people are ruled out. In that case, recruitment could be done in such a way that only people who can read are likely to react, for example by using written advertisements.

The kind of recording device and the environment are closely related. Combinations that are often used are a high quality microphone in a studio environment, a fixed network telephone in a living room or office environment, a cell phone in public places and a headset microphone on a desktop

computer. For collections that take place in a studio, for example, it is necessary that speakers come to a place set up by the producer of the database. In this case speakers should be found in the neighbourhood, for example by advertising in local newspapers.

In other collections, e.g. through cell phones, speakers are expected to have the equipment themselves. We did some projects of recordings over the mobile network in a time that relatively few people owned a cell phone. We looked for owners of a mobile telephone by advertising in a magazine for cell phone users and using a directory listing of cell phone users.

2.2. Special tasks for speakers

For some databases a special task is demanded from the speakers. For example, in SpeechDat, in the databases that were built for speaker verification purposes, people were asked to call 20 times, with three days in between calls [4]. Furthermore, in some of the databases recorded over the mobile network, the participants were requested to call four times from different environments [5]. This lays an extra burden upon the speakers and a higher reward should be provided to keep people motivated. There is a serious risk that participants do not finish their series of calls, and they should be reminded that they still have some calls to make.

In the Spoken Dutch Corpus project [6] many speech styles are recorded among which spontaneous dialogues. To collect these, people are asked to record discussions at home on a MiniDisc. It appears that not everybody is able to use the device, which makes recruitment more difficult.

For database collections in a car, it might be necessary that the speaker is driving during the recordings. In this case, obviously, only speakers with a driving licence and perhaps a car should be recruited.

2.3. Country and culture where recordings take place

In some countries speaker recruitment is in general easier than in others. In highly individual, prosperous societies, people are less willing to be of assistance in speech collections. People are more likely to participate if they are less well to do (if rewards are given), or if social pressure is higher. In Mediterranean countries social and family structures are tighter than in northern European countries, which makes recruitment via word of mouth easier. In a project of car recordings we are

doing, in which we ask people to record family and friends, recruitment appears to be much harder in the Netherlands than in Italy, partly because of this reason.

One of the experience in the SpeechDat projects was that in some countries, in which the society is more hierarchically structured, for example in former socialist countries, people are less likely to refuse participation if their employer tells them to.

In countries where more than one language is spoken, optimal strategies for recruiting native speakers of the target language should be found. In countries with a high percentage of illiteracy, finding people who are able to read prompt sheets will be difficult.

2.4. Number and distribution of speakers

The number of speakers that are needed in the database and the effort the recruitment will take are not linearly related. The more participants are needed, the more difficult it will be. When only a few speakers are needed, direct colleagues and family members, who are easiest to recruit, will be enough. When a bit more people are needed, other colleagues and acquaintances might be willing to join in, but when many people are needed, strangers have to be recruited. The people easiest to recruit are the first that will be found. The last few needed are always the hardest to get.

In most databases an equal distribution of gender, age and accent of the speakers is strived at. In some databases, balance of socio-situational settings is needed as well. Having to take these distributions into account makes speaker recruitment more difficult. In our experience, there is always a group of people that is hard to cover in the database. In our experience, people over 50 years of age and people with an accent from specific parts of the country are hard to find. In recruiting speakers for spontaneous dialogues in the Spoken Dutch Corpus, people with a low education level appeared to be more reluctant to participate than people with higher education. They were suspicious about the purpose of the recordings and feared their privacy.

2.5. Time of recording

During holidays, it is difficult to find people, because many are away. On the other hand, people have more time during holidays, because they do not have to work. Wintertime, when it is cold in some countries, is not suitable when one wants to collect mobile recordings made in the streets, as we

experienced ourselves in Dutch SpeechDat recordings [7].

3. Strategies

The most important part of speaker recruitment is advertising the project and convincing people to take part in the recordings. Possible strategies to reach the public are:

1. Friends, relatives, colleagues and acquaintances
2. Advertisements
3. Press release
4. Direct mail/telephone
5. Email lists
6. Posters, flyers, contact on the street
7. World wide web
8. Snowball/recruitment of recruiters

These strategies are described here, with their advantages and disadvantages (see table 1) and personal experiences.

An alternative way to deal with speaker recruitment is to hire a marketing agency. These are specialised in finding people for interviews, but not really for the task described here. Such an agency will take all the work from your hands, but is very expensive. It is not really a strategy itself, since a marketing agency will most probably make use of (some of) the strategies described here.

3.1. Friends et al

Friends, relatives, colleagues and acquaintances are more likely to participate than others, because there is a social pressure on this group that is absent with strangers. Also students from a course given by (someone close to) the producer of a database can be kindly asked or pressed to participate.

The drawback is that there won't be very many from these groups. Furthermore, there is a danger of an imbalance in the speaker distribution: it is not very likely that this group is balanced in age, sex, speaker accent and socio-situational setting.

In all our data collections, these people are represented, sometimes in a quite substantial part, but most of the times only in a relatively small part.

3.2. Advertisements

It can be very effective to place an advertisement in a newspaper, a magazine, or on the radio. Attention should be paid in which medium the advertisement is placed and moreover what kind of readers or listeners can be expected. If an ad is placed in a regional newspaper, a bias of respondents with an accent of that particular region should be expected.

On the other hand, if the recordings take place on a fixed spot, local newspapers are more useful than national ones. If an equal distribution of men and women is necessary, an ad should not be placed in a typical man's magazine (like the *Playboy*). But when there are e.g. already enough male speakers and only female speakers are wanted, an ad in a typical woman's magazine can be very useful. With an advertisement in a big newspaper, many people can be reached. Unfortunately, placing an advertisement in media with a large audience is very expensive.

We placed an advertisement a few times with moderate success. It gave us some participants, but never enough.

3.3. Press release

A method to try to reach the same public as with an advertisement, but at a much lower cost is to have a press release in the hope to get an article or interview. People might be even more interested in comparison to an advertisement, but one depends completely on journalists, who will pay attention to the press release. One has also little influence on what exactly is published. Our experience is that in a relatively new science field like speech processing, a press release will get quite some attention.

We did a press release in the SpeechDat mobile collection. That led to two interviews on local radio, a front page article in a regional newspaper, an article in the university news magazine and some smaller articles in newspapers all over the country. A very small article (on page seven) in a national newspaper without any details or telephone number gave us many participants.

3.4. Direct mail/telephone calls

Sending people a letter, is much more personal than advertisements. Moreover, it gives the opportunity to explain what is expected from participants. This is even more so for telephone calls, which has the additional advantage that one can directly try to persuade people to participate. However, both methods are expensive, especially the telephone calls, since time (thus money) is spent for each person who is contacted. Also, these methods are not appreciated by many people.

In one of the first large telephone data collections in Europe, the first in the Netherlands, Polyphone [8], direct mail was used by picking addresses from the telephone book randomly. The response rate was 15 percent, which was less than expected.

3.5. Email lists

The big advantage of sending a request to participate by email is that in this way many people can be reached very easily at very low cost or effort. On the other hand, this way of recruitment is not very sympathetic and people might be offended by it. It is also not guaranteed that a usable email list is present and one might need to compile one.

We tried this method in our current collection. Secretaries who have lists with email addresses of students are reluctant to send an email to these lists, so we compiled one ourselves, by looking for email addresses on websites. This was quite time consuming and the response rate was very low.

3.6. Posters/flyers/contact on the street

A relatively easy recruitment method is to hang posters on several places and spread flyers around. An extension of this is to hand out flyers and to try to persuade people directly. With posters and flyers, many people can be reached at a relatively low cost. When one also talks to the people it is very time consuming, like the direct telephone calls. However, persuasion power is larger with direct contact. A strong disadvantage of this method is that recruitment can be done only on a local level. Hanging posts and flyering should be repeated often and can only be done nearby (unless one travels around or hires persons on other locations).

We used this methods in several projects. We set up a page on the WWW and spread posters and flyers with a pointer to the web site and with appealing texts. The flyers were mostly spread out in computer rooms with internet access and in canteens. The website was not visited as often as expected and hanging posters and flyering had to be done over and over again, since posters and flyers tend to disappear very quickly.

3.7. World wide web

Since a couple of years, it is also possible to use the world wide web for recruitment. People can register on a web page that is advertised on the web. It is easy and cheap, but because a web page is a needle in a haystack, the pages on which is advertised must be very well visited for this method to be successful. We have not done commercial advertisements (banners), but we tried to get a link to our page on sites with 'things for free', which are very popular in the Netherlands. This proved to be a successful recruitment method.

3.8. Snowball/recruiting recruiters

One of the more effective methods proved to be spreading the message by word of mouth, especially when it is rewarded. In that case, people who already participated in the recordings (and others) get a reward for every new participant they register. Of course, the new participant has to complete the task before rewards are given. This is also an easy way of recruiting speakers, but it makes the recordings more expensive.

We use this method in most of our recordings in the last stages of a project, when we are looking for specific categories (mostly certain regional accents).

A related method, which we use in our current (in car) recording project, is to look for people who recruit and also record other people. The recording procedure should be easy enough that it can be done by people without recording experience. If someone does not understand the procedure, many recordings might go wrong. The advantage of this method is that fewer people have to be recruited, but recruitment will be harder, since the task that is asked from participants is much harder.

Method	Advantage	Disadvantage
friends et al	social pressure	low coverage, bad distribution
adverts	high coverage	expensive
press release	high coverage	dependent on journalists
direct mail/teleph	personal	expensive, unsympathetic
email lists	high coverage	unsympathetic
posters/flyers	high coverage	laborious
WWW	easy	low coverage
snowball	less effort	extra costs

Table 1 Overview of different recruiting strategies

4. Registration

Unless people can participate at once, e.g. in a telephone recording in which prompt sheets are sent by direct mail, they have to register somehow. Some ways to do that are:

1. WWW fill-in form
2. Paper fill-in form
3. Telephone call

It is important to make registration as easy as possible. It should also be taken care of that registration is possible all the time, e.g. by installing an answering machine if registration goes

via telephone calls. We also experienced that people should not be pressed too hard; if a decision is asked immediately people might refuse, so they should be given some time to decide.

4.1. WWW fill-in form

This is the most important method we use nowadays. A web site with a fill-in form is easily set up. Automatic checking whether all important data is filled in and automatic processing is guaranteed. It is also easy for the people who fill in the form and in many countries most people have internet access. Usually it is available all the time.

4.2. Paper fill-in form

This is essentially the same as the web form, but a pencil and some effort are needed to send the registration in. It is a bit more work and costly for both participant and recruiter. It is still more widely accessible than the web form, though. It is desirable to have a postage paid return envelope available.

4.3. Telephone call

A telephone call for registration is a lower threshold than the paper fill in form, but it is much more time-consuming for the producer to register them and telephone calls might come in at inconvenient times. When there is no-one at the telephone, a answering machine should be available.

5. Rewards

People can volunteer for recordings for various reasons: because they are interested or because of social pressure, for example of close relatives to the producers of a database. Most of the time, these are not sufficient and rewards are needed to persuade people to participate. Some things to bear in mind with respect to rewards are the following:

1. Type
2. Size
3. Appropriateness

5.1. Type

The reward can be a gift, varying from money and value checks to telephone cards. Alternatively or additionally, one can hold a lottery in which participants can win a big prize. Our experience is that (not surprisingly) people prefer money, but rewarding money can be a legal problem, since a sort of employer employee relationship is

established, which is not free of tax duties. Once we did a lottery of which the (only) prize was a weekend to Paris for two, which was received quite positively.

5.2. Size

A very important thing to take care of is that the rewards are big enough. When the rewards are too small people will not easily participate and a lot of effort will have to be put in recruiting people who will participate for this small reward. In the end, this could turn out to be more costly than having started with a bigger reward. Higher rewards can be given to speakers from categories (like accent, age) that are hard to recruit.

5.3. Appropriateness

Rewards should be appropriate and as neutral as possible. Obvious errors to avoid are giving a telephone card to people who participate in a mobile speech database recording, with participants with their own cell phone or giving book vouchers to people in a region with a high illiteracy rate.

6. Conclusions

In this paper we tried to give an overview of possible recruitment strategies for participants of a speech database. Without a doubt, there are more strategies and variations than the ones described here. It is very hard to evaluate the different strategies, because there are many differences in the circumstances that influence the success of a strategy. We tried to point out what to take into account when one chooses a certain strategy, based on our experience, mostly in the Netherlands.

In the recordings we did, we tried all the strategies described here. In most of the recordings we use more than one strategy and most partners in the SpeechDat projects also used various strategies.

In our experience, using some sort of snowball strategy works most of the times, but whenever possible, media coverage should be used to recruit speakers. For registration, a WWW fill-in form is preferred, if internet access is widely available. Finally, much frustration (and money) can be spared by having attractive rewards.

References

1. The SpeechDat Family,
<http://www.speechdat.org>

2. Lindberg, B., Comeyne, R., Draxler, C., Senia, F. (1999) "Speaker Recruitment Methods and Speaker Coverage – Experiences from a Large Multilingual Speech Database Collection" *Proceedings of ICSLP '99, Sydney, Australia*
3. Fraser, N.M. and Gilbert, G.N. (1991) "Simulating Speech Systems" *Computer Speech and Language, vol 5, pp. 81-89*
4. Kordi, K. (1996) "Definition of corpus, scripts, and standards for Speaker Verification" *SpeechDat Technical Report SD1.1.3*
5. Velden, J.G. van, Langman, D., Pawlewski, M. (1996) "Specification of speech data collection over mobile telephone networks" *SpeechDat Technical Report SD1.1.2/1.2.2*
6. The Dutch Spoken Corpus,
<http://lands.let.kun.nl/cgn/ehome.htm>
7. Sanders, E., Heuvel, H. van den, Choukri, K. (1999) "Building Speech Databases for Cellular Networks" *Proceedings of Eurospeech '99, Budapest, Hungary*
8. Damhuis, M., Boogaart, T., Veld, C. in 't, Versteijlen, M., Schelvis, W., Bos, L., Boves, L. (1994), "Creation and Analysis of the Dutch Polyphone Corpus" *Proceedings of ICSLP '94, Yokohama, Japan*
9. Gibbon, D., Moore, R., Winski, R. (eds.) (1997) "Handbook of Standards and Resources for Spoken Language Systems", *Berlin, New York: Mouton de Gruyter*