OrienTel: Recording Telephone Speech of Turkish Speakers in Germany

Chr. Draxler

Bavarian Archive for Speech Signals (BAS)
c/o IPSK, Ludwig-Maximilians-Universität München
draxler@phonetik.uni-muenchen.de

Abstract

OrienTel is a project to create telephone speech databases for both the local and the business languages of the Mediterranean and the Arab Emirates. In Germany, 300 Turkish speakers speaking German were to be recorded. The database is an extension of the SpeechDat databases. This paper outlines the recording setup, the recruitment strategy and the annotation procedure. Recruiting the speakers was a particular challenge because none of the recruitment strategies used in previous SpeechDat projects in Germany did work and a new approach had to be found.

1. Introduction

OrienTel [8] is an EU-funded project (IST-2000-28373) for the collection of telephone speech for the main local and business languages spoken in the Mediterranean and the Arab Emirates. The project coordinator is Philips Speech Technology (now Scansoft). A total of 21 speech databases will be collected within the project (Table 1). Each country contributes 1750 to 2000 speakers - in the case of Turkey, 1700 speakers are recorded in Turkey, 300 in Germany.

The official project partner for Turkey and Germany is SIEMENS AG (Munich, Germany). SIEMENS has charged ELDA with the recordings. ELDA in turn has charged the Middle East Technical University in Ankara with the recordings in Turkey, and the University of Munich with the recordings in Germany.

2. Database Contents and Demographics

The database contents of OrienTel is an extension and adaptation of the original SpeechDat database contents [6], [7]. Extensions were needed to cover languages with a non-Latin script and different standard and colloquial languages, e.g. Arabic and Hebrew, and to meet the specific requirements of the project partners [3]. The core vocabulary is given in Table 2.

The main demographic criteria are gender, region, age and calling environment. Both genders should be distributed evenly. Three age classes were defined for the speakers: 16-30, 30-45, and 45 and older. The proportion of each age class was set to a minimum of 30%, 20%, and 10% respectively. For the regional distribution, four to six regions should be specified based on accent or urban vs. rural areas. 30% of all recordings should originate from a fixed network phone, 70% from a mobile phone. Of the mobile network calls, a minimum of 20% should come from home or office environments, 20% from a public place, e.g. café, street, train station, etc., and 15% from a moving vehicle. For all these criteria, a tolerance was defined.

2.1. German Recordings

The German OrienTel speech database must consist of at least 300 calls. All speakers must be of Turkish origin, 90% of them must be first generation immigrants, i.e. must have grown up in Turkey and come to Germany at least four years ago. These criteria are described in detail in [1].

<table>
<thead>
<tr>
<th>Country</th>
<th>Languages collected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>Modern</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>Standard</td>
</tr>
<tr>
<td>Egypt</td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td></td>
</tr>
<tr>
<td>Morocco</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>Turkish</td>
</tr>
<tr>
<td>Cyprus</td>
<td>Cypriote Greek</td>
</tr>
</tbody>
</table>

Table 1: OrienTel databases

<table>
<thead>
<tr>
<th>Code</th>
<th>#</th>
<th>Utterance description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, B</td>
<td>2</td>
<td>isolated digit items</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>digit/number strings</td>
</tr>
<tr>
<td>N</td>
<td>2</td>
<td>natural numbers</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>currency amount</td>
</tr>
<tr>
<td>Q</td>
<td>2</td>
<td>yes/no questions</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>dates</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>times</td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>application keywords/keyphrases</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>word spotting phrase</td>
</tr>
<tr>
<td>O</td>
<td>5</td>
<td>directory assistance names</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>spellings:</td>
</tr>
<tr>
<td>W</td>
<td>4</td>
<td>phonetically rich words</td>
</tr>
<tr>
<td>S</td>
<td>9</td>
<td>phonetically rich sentences</td>
</tr>
<tr>
<td>X</td>
<td>2</td>
<td>Spontaneous items (for control only)</td>
</tr>
</tbody>
</table>

Table 2: OrienTel database contents
For German, some extra items were added (Table 3). The L, M, Q, and W items were added to provide a better coverage of items because of the small size of the database.

The extra O items are the same as the ones collected in Turkey to allow comparisons between Turkish spoken in Turkey and Turkish with German influence.

The X items are used to gather information on the speaker and the call: gender, age, first name, environment from which the call is made, type of handset used, federal state of residence, city where he or she spent childhood, and the time he or she has been living in Germany.

2.2. Speaker Prompts

Speakers are sent a prompt sheet with instructions and the text to read. In Germany, this prompt sheet consists of a double-sided A4 sheet with the prompt sheet in tabular form. Instructions were printed in plain face, text spoken by the telephone server was printed in italics, and the text to be spoken by the speaker in bold face. The beep was represented by a black bullet and a hint on how to speak the item was given in the first column, e.g. read the item.

3. Data Collection

3.1. Prompt Sheet Generation

The prompt sheets for the German data collection were taken from the SpeechDat-II German data collection and adapted to the requirements of the OrienTel project. The original German SpeechDat sets of phonetically rich words and sentences were taken from newspaper texts. Such texts are not intended to be read aloud, and hence even many native speakers of German had problems reading these sentences in SpeechDat.

To reduce these problems, the sets of phonetically rich words and sentences were redefined and checked for readability. Furthermore, all sentences were checked for potential offense towards Turkish speakers – sentences containing references to political parties, religious communities, etc. were removed. Even then, not all sentences were easy to read.

150 different German prompt sheets were generated. The prompt sheets were validated by SPEX prior to the recordings [4]. SPEX discovered some deviations from the specifications which were then corrected for the actual recordings.

3.2. Minidatabase Collection

In OrienTel, all database producers are required to have a small 10 speaker database validated before the main recordings begin.

The German minidatabase was delivered to SPEX in December 2002. Because the project was behind schedule already, it was decided to go on with the recordings although the minidatabase had not yet been validated.

Early February 2003 the minidatabase was validated and only minor deviations were found, mainly in the documentation.

3.3. Speaker Recruitment

In previous SpeechDat data collection projects in Germany, the most successful speaker recruitment strategies were advertisements and articles in company and telecom subscriber journals and the regional pages in local newspapers, and a hierarchical snowball system in which speakers recruited other speakers [5].

Originally it was planned to simply use the same recruitment strategy for OrienTel. However, it quickly turned out that these strategies were not effective – in fact, some of them were counter productive. The following observations were made:
• People often did not understand the motivation for the recording and thus did not participate.
• Speakers often did not want a reward for their phone call – especially not money.
• People almost never called when they were given a prompt sheet for later calls.
• It seemed that non-Turkish experimenters were not well accepted.
• There were only few downloads of prompt sheets from the web site.

It became clear that new recruitment strategies were needed.

3.3.1. Interviews with Turkish students

In a first step, Turkish students were interviewed for an explanation of the observations, and they were asked to recruit a few speakers and report on their experiences.

The student recruiters confirmed the observations listed above. However, they were only slightly more successful than non-native experimenters in explaining the project to potential speakers and thus their rate of success was not much higher.

3.3.2. Article in SiemensWelt journal

Because the recordings were made for SIEMENS, the internal monthly journal SiemensWelt was asked to write an article on the project. SiemensWelt devoted an entire page to the project. As an eye-catcher, the headline was written in Turkish, and it mentioned special incentives for organized groups of speakers. Furthermore, the article contained photos of the speech research group within SIEMENS and screen shots showing how different the same word appears when spoken by different speakers. At the end of the article, Internet links, phone and fax numbers were provided.

The article appeared shortly before Christmas 2002. A total of about 50 people responded by fax or phone to request prompt sheets, and a few sheets were downloaded from the web site. Requests for prompt sheets continued well into February 2003.

As a follow-up, 850 e-mails were sent to Turkish employees within SIEMENS in February. Approximately 25 people requested prompt sheets by e-mail, and further 20 people downloaded prompt sheets directly from the German OrienTel web site (www.speechdat.org/orientel).

3.3.3. Distribution of sheets via multiplicators

In January 2003, three volunteers offered to distribute prompt sheets. They had access to large communities of Turkish people due to their position in these communities. One was a journalist, the other the head of a cultural society, the third a former radio speaker and owner of a successful catering service and restaurant.

These multiplicators each received prompt sheets with coded prompt sheet numbers. Any call with one of these prompt sheet numbers could thus be traced to the recruiter, eliminating the need for returning individual data sheets.

A total of 86 calls was obtained in this manner. However, because the multiplicators live in Munich, the geographic distribution of speakers was not at all balanced.

3.3.4. Distribution via higher education institutions

In January 2003, more than 300 German institutions of higher education were asked whether they would distribute to their Turkish students pre-packed and stamped letters with invitations to participate. For privacy reasons, the letters were to be addressed by the institution.

More than 60% of all institutions responded positively, and a total of more than 3000 prompt sheets was distributed from January to March 2003.

Initially, institutions were asked to select students of Turkish nationality older than 25 years of age who were born in Turkey. This search criterion was difficult to use because the place of birth was not always registered, and it did not say anything about the time spent in Germany.

A better search criterion, proposed by the institutions, was to select students according to their final school degree: “Bildungsausländer” vs. “Bildungsinnenländer”, i.e. a foreign school degree vs. a German Abitur. This refined search criterium cut down the number of addressees considerably, but was much more focused.

More than 300 calls were recruited in this manner, and the regional distribution was rather good, except for Eastern Germany where the number of Turkish students is extremely low.

3.3.5. Oversampling and snow-ball recruitment

Although the recruitment via educational institutions was quite successful, it was considered important to get speakers to recruit further speakers. We thus decided to record every speaker regardless of whether his or her call was needed to cover the demographic criteria. The rationale of this decision was to try and reach new speaker groups via those recruiters.

As an incentive we offered 100 € for every ten speakers recruited and recorded who met the demographic criteria. The threshold of ten speakers was set to obtain sufficiently many speakers quickly, and to be able to offer a relatively large reward.

It was important to keep the recruitment process simple and straightforward. We thus sent the recruiter a letter explaining the procedure, a speaker list and ten prompt sheets. The recruiters were asked to put the prompt sheet number and the first name of the recruited speaker on the speaker list and return the list together with a bank account number to our institute once ten people had called. We then transferred the reward to the recruiter. Recruiters could always check the status of their recruitment on the project web site.

The letters to the recruiters were customized to steer the recruitment of speakers. For example, to record more speakers from public places, recruiters were informed that a call from a public place would count as two calls.

In general, this recruitment scheme worked rather well. It was particularly successful if the recruiter was a member of a large family or a local community, e.g. a student organization. However, many recruiters found it very difficult to successfully recruit further speakers for the reasons listed above.

3.4. Status of the Recordings

The German OrienTel project was scheduled to end in February 2003. However, at the end of March 2003 recordings
were still going on. Although far more than 300 calls have been recorded some demographic classes have not yet been covered sufficiently: there are too few recordings of speakers older than 45, of speakers calling from a car, and of speakers from Eastern Germany. Recruiting is now limited to speakers from these classes.

4. Annotation

The recordings were annotated according the OrienTel annotation guidelines, i.e. an orthographic annotation with additional marker symbols for hesitation phenomena, articulatory noises, e.g. laughing or coughing, and intermittent or stationary non-speech noise, e.g. door slamming or noisy connections. All annotations are stored in SAM label files [3].

The annotation was performed using the WWWTranscribe system [2]. WWWTranscribe retrieves the next utterance to transcribe from the transcription database and presents it acoustically via a standard web browser. To speed up the annotation process, WWWTranscribe features editing buttons for common annotation tasks such as converting from digit strings to their explicit orthographic form, e.g. “1 2 3” to “eins zwei drei” (Fig. 2).

As of March 25, 2003, a total of 11,527 utterances have been annotated; this corresponds to more than 190 recording sessions. 11,026 annotations are marked as OK, 171 as garbage, and 303 as being out of sync. This last category describes utterances that do not appear in the correct position within a recording, e.g. because they were uttered to early or too late; most of them can be moved to their proper position simply by renaming the signal and label files.

1093 annotations contain markers for mispronunciations or repair phenomena, e.g. repetitions of words or parts of the utterance. The utterances affected most are sentences, spellings, and some spontaneous utterances. Sentences are long and sometimes difficult to read and hence prone to repairs. The pronunciation of letters in Turkish spelling is different from German, and speakers often use the Turkish letter names which are then consequently marked as mispronunciations by German standards.

260 annotations contain markers for filled pauses or hesitations. The most affected items are the spontaneous question for the time of day, money amounts, telephone numbers and the spelling of an artificial letter sequence. An explanation is that these items impose cognitive load on the speakers because they either have to discover the type of the item from its format, or, as in the case of the spelling sequence, have to produce the names of very rare letters, some of which occur only in German but not in Turkish.

5. Summary

Recruiting speakers was the most difficult part of the data collection. The recruitment strategies employed in previous projects did not work well for the non-native speakers in OrienTel in Germany. An explanation is cultural difference in combination with an uneasy feeling by Turkish speakers due to their position as immigrants in a foreign country.

An example for cultural difference: for Turkish speakers giving one's voice is seen as doing the project a favor: if the speaker decides to participate, then he or she does not want an immediate and personal reward, especially not money. In fact, money is considered as something that reduces the value of the favor. If the money goes to a community, a family, etc. then it is much more likely to be accepted. In previous German data collections, an immediate and personal reward was a necessity.

The uneasy feeling was caused by the formal nature of the recording process: a data sheet to be filled in with personal data, a two-page prompt sheet with strange German sentences, an impersonal telephone computer adhering to a strict recording protocol, all this caused people to be reluctant to participate. One reason may be that many first generation immigrants have rather low level of education and thus feel intimidated by a task requiring reading skills in a foreign language. This explanation is supported by the fact that young speakers with a high education level, e.g. from university, were much more willing to participate. However, even they were not able to recruit older family members - most of the recruited speakers were of the same age as the recruiter or younger.

6. References

[2] Draxler, Chr. WWWTranscribe - a modular transcription system based on the world wide web, Proc. of Eurospeech 95, Rhodes
[8] www.orientel.org