

# RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus

Ingunn Amdal<sup>(1)</sup>, Ole Morten Strand<sup>(1,2)</sup>, Jørn Almberg<sup>(1)</sup>, and Torbjørn Svendsen<sup>(1)</sup>

<sup>(1)</sup> Department of Electronics and Telecommunications  
Norwegian University of Science and Technology  
NO-7049 Trondheim, Norway

<sup>(2)</sup> Norwegian Defence Research Establishment  
NO-2027 Kjeller, Norway

E-mail: {ingunn.amdal,torbjorn}@iet.ntnu.no, jorn.almberg@hf.ntnu.no, ole-morten.strand@ffi.no

## Abstract

This paper describes the Norwegian broadcast news speech corpus RUNDKAST. The corpus contains recordings of approximately 77 hours of broadcast news shows from the Norwegian broadcasting company NRK. The corpus covers both read and spontaneous speech as well as spontaneous dialogues and multipart discussions, including frequent occurrences of non-speech material (e.g. music, jingles). The recordings have large variations in speaking styles, dialect use and recording/transmission quality. RUNDKAST has been annotated for research in speech technology. The entire corpus has been manually segmented and transcribed using hierarchical levels. A subset of one hour of read and spontaneous speech from 10 different speakers has been manually annotated using broad phonetic labels. We provide a description of the database content, the annotation tools and strategies, and the conventions used for the different levels of annotation. A corpus of this kind has up to this point not been available for Norwegian, but is considered a necessary part of the infrastructure for language technology research in Norway. The RUNDKAST corpus is planned to be included in a future national Norwegian language resource bank.

## 1 Introduction

Large amounts of speech data are a prerequisite for speech technology. This paper presents RUNDKAST<sup>1</sup>, a database of broadcast news that will reduce the current shortage of speech data for research activities on the Norwegian language. It supplements existing language data and will be a natural part of a future national language resource bank (Svendsen et al, 2002).

The database will be of particular benefit to speech technology in two application domains, namely user interfaces and audio indexing. User interfaces based on speech alone, or speech in combination with other modalities, are often considered for applications that require hands-free interaction, involve mobile devices, or run over the telephone network. However, the success of such applications depends on speech technology that is adequate for practical use by most people in most everyday situations. To achieve this performance it is crucial to have access to relevant data in sufficient amounts. RUNDKAST does to some extent contain such data.

The goal in audio indexing is to enable searching and browsing of audio data (Makhoul et al, 2000). The main target for this technology has been broadcast news, exemplified by the yearly NIST evaluations<sup>2</sup>. In addition, it is now increasingly considered for other types of audio data such as recordings of meetings and court trials. Audio indexing depends on a number of basic speech technologies such as speech recognition (including speaker adaptation), topic classification, speech/non-speech detection, and speaker diarization. RUNDKAST provides a starting point for research on audio indexing also for Norwegian.

A corpus like RUNDKAST will serve as an entry ticket to international research cooperation involving speech technology for broadcast news applications. This is an arena where Norway so far has lacked the necessary data to join. One example is the COST278 special interest group on Broadcast News for many European languages (Žibert et al, 2005).

The interest in alternative paradigms for automatic speech recognition has increased recently (Lee, 2004; Moore, 2007). One of the alternative approaches is based on utilization of phonetic knowledge in combination with a statistical framework. Much of the initial, basic research in this area has focused on experimentation on phonetically annotated databases, such as the acoustic-phonetically annotated TIMIT corpus (Garfalo et al, 1993). In order to provide an experimental database for Norwegian, useful for e.g. investigating the validity of these approaches across languages, we decided to provide broad phonetic annotation of a subset of RUNDKAST. The SIRKUS project<sup>3</sup> at NTNU, which is working in this area, plans to use the RUNDKAST data.

The outline of the paper is as follows: In section 2 we present a short overview of the RUNDKAST contents. More details from the specification and development are given in section 3. Discussion, conclusion and further work are given in sections 4 and 5, respectively.

## 2 An overview of RUNDKAST

The first task of the project developing the RUNDKAST corpus was to establish agreements with information providers. We were fortunate to reach an agreement with Norway's major broadcasting company NRK (the Norwegian Broadcasting Corporation), which has a geographical coverage of approximately 100%. NRK is a public broadcaster funded by a license fee and has no

<sup>1</sup> <http://www.iet.ntnu.no/projects/rundkast/>

<sup>2</sup> <http://www.nist.gov/speech/tests/bnr/>

<sup>3</sup> <http://www.iet.ntnu.no/projects/sirkus/>

commercials in radio broadcasts. A broadcast company like NRK would greatly benefit from an available audio indexing system to browse and search their vast archive of audio data.

The entire corpus is manually transcribed and segmented using hierarchical levels. An independent level contains background noise labeling. A subset of one hour of read and spontaneous speech from 10 different speakers was manually annotated using broad phonetic labels.

The corpus was annotated during a period of 18 months in 2006 and 2007. The high level annotation was done by technology students in telecommunications whereas phonetics students did the broad phonetic annotation.

The RUNDKAST corpus also includes documentation of formats and annotation specifications. In preparing for distribution more than half of the recordings and transcriptions were internally validated by a supervisor. We also plan an external validation before distribution.

## 2.1 Selected material

RUNDKAST contains about 77 hours of broadcast news. The content of the corpus covers both read and spontaneous speech, as well as spontaneous dialogues and multipart discussions. There is large variation between speakers, speaking styles, and topics. Speaker turns may moreover be rapid and several speakers may talk simultaneously. The quality of the recordings include studio, telephone (land line, mobile and satellite telephony), and occasionally also radio link. The corpus contains frequent occurrences of background noise, jingles, music, and audio illustrations.

The RUNDKAST corpus is limited to eight different news programs from the NRK radio channels P1 and P2, see Table 1. The emphasis on P2 is due to its focus on news. The programs that were chosen each has a specific mix of five broad categories: *short news*, *long news*, *debates*, *interviews*, and *commentaries*.

- *Short news* is typically 10-40 seconds long and read from a manuscript by the program host. Sometimes short news also includes short sound clips from another speaker (e.g. a speech or an interview).
- *Long news* is typically 1-20 minutes long and has mixed content. Typical content is manuscript read speech from the studio host or a reporter, spontaneous on-site reporting, interviews on-site or by telephone, and excerpts from speeches or press conferences.
- A *debate* is typically 5-20 minutes long and contains spontaneous speech from the program host and two or more debaters.
- An *interview* is typically 1-10 minutes long and contains spontaneous speech from the program host and an interviewee. Debaters and interviewees may attend by telephone.
- A *commentary* is typically 5-10 minutes long and contains manuscript read speech from an in-house commentator or a foreign correspondent.

Debates, interviews and commentaries are all typically introduced with manuscript read speech from the program host. Notice that some content may best be considered a

mixture of two or more of these five categories. Three of the programs; "Her og nå", "Søndagsavisa", and "Ukeslutt", have markedly softer profiles than the other programs. Also, "Verden på lørdag" focuses on international affairs, while "Politisk kvarter" focuses on politics.

For each program between 7 and 22 episodes were chosen giving a total of 115 episodes from the period 1995-2006. Most of these are from the years 2003, 2005, and 2006. Each episode is from 15 to 60 minutes long, giving a total of about 77 hours, see Table 1. The episodes were selected by NRK and reflect typical content of the selected news programs during this period.

Program names	Duration per prog. (minutes)	Number of episodes	Total duration (hours)
Dagsnytt 1230 (P2)	15	14	3.5
Dagsnytt 1730 (P2)	30	22	11
Dagsnytt atten (P2)	55	15	13.75
Her og nå (P1)	55	17	15.5
Politisk kvarter (P2)	15	15	3.75
Søndagsavisa (P2)	55	7	6.5
Ukeslutt (P1 and P2)	60	9	9
Verden på lørdag (P2)	55	16	14.25
<b>SUM:</b>		<b>115</b>	<b>77.25</b>

Table 1 Overview of selected material

## 3 RUNDKAST Specification and Corpus Development

### 3.1 Tools, data formats, and structure of annotation

We have defined standards for annotation and transcription based on international conventions and tools. For the high level annotation, including orthographic transcription, we used *Transcriber* (Barras et al, 2001)<sup>4</sup>. This tool is specifically designed and widely used for annotation of broadcast news recordings and spontaneous speech, e.g. in the Norwegian NoTa corpus (Johannessen et al, 2007). The broad phonetic annotation was performed using *Praat* (Boersma & Weenink, 2006)<sup>5</sup>, a tool suitable for phonetic annotation. It is widely used for phonetic annotations, e.g. in the Dutch CGN corpus (Schuurman et al, 2003).

For the high level annotation we used the four levels of annotation supported by *Transcriber*. Three of these are hierarchical: an *episode* is divided into *sections*, sections are divided into *speaker turns*, and speaker turns are divided into *segments*. A graphical overview is given in Figure 1. In addition, there is another level, *background*, for lasting acoustic background conditions, e.g. the "music" label shown in Figure 2.

<sup>4</sup> Version 1.5.1 from <http://trans.sourceforge.net/>

<sup>5</sup> Version 4.5.01 from <http://www.praat.org/>

annotation level: ← one episode file →

1 report nontrans filler report ...

2 speaker 1 speaker 2 no speaker speaker 1 ...

3 [i] blah blah ... [lp] more blah ... [b]-noisy blah[-b] ... ...

Transcriber 1.5.1

File Edit Signal Segmentation Options Help

tredjeplass etter den åttende fartsprøven .

[p] [i] det var nyhetene . [p]

report - kultur

(no speaker)

Prinsesse Martha Louise

det eventyret jeg skal lese for deg nå , [p]

det er et av de eventyrene jeg er mest glad i .

[n] det var en gang en konge

music

sport

larit Kolberg

(no ... Prinsesse Martha Louise

regjerende ... [i] mens [e] ... [p] [i]. det ... det er et ... [p] det.

tte løpet , ... fartsprøven . ... [p] nå , [p] ... i . konge .

2:40 2:45 2:50 2:55 3:00

### 3.1.1 Section level of annotation

- *Nontrans* is used for longer audio sequences not to be transcribed, such as jingles, music, and sound illustrations.
- *Filler* is used for headlines, announcements, small talk, and other types of transcribable spoken sequences that are not regarded as proper news items.
- *Report* is used for regular news content (see section 2.1), and make up the bulk of the transcriptions. Furthermore, reports are annotated according to topic to facilitate research on topic classification (domestic, foreign, economy, sports, culture, or traffic).

### 3.1.2 Speaker turn level of annotation

Mother tongue	Written standard	Dialect regions
Native	Bokmål	Østland (South-East Norway)
Nonnative	Nynorsk	Sørland (South Norway)
		Vestland (West Norway)
		Trøndelag (Mid Norway)
		Nord-Norge (North Norway)

Second, each speaker turn is labeled according to speaking mode, channel, and fidelity. The possible classifications are given in Table 3. Note that the fidelity judgments have different meanings according to the channel label, as specified in Table 4. The speaker turn information is essential for research on audio classification and speaker diarization. It also gives means to classify the degree of difficulty the data will give in a speech technology research experiment.

Table 3 Categories for speaker turnsTable 4 Fidelity specification

Segments are typically 2-5 seconds long and transcribed orthographically (see section 3.2), including labels for

acoustic events, see Table 5. The segment boundaries are placed at pauses or speaker breathings when possible.

### 3.1.4 Background level of annotation

At the independent level for lasting background conditions, audio sequences were marked as containing zero or more of the following four categories of background noise:

- Music.
- Speech (from background speakers).
- Ssh (stable broadband noise).
- Other (all other noise types).

While *Transcriber* allows arbitrarily placed background boundaries, the convention in RUNDKAST is that such boundaries must coincide with segment boundaries.

## 3.2 Segment level orthographic transcription

### 3.2.1 Written standards of Norwegian

Orthographic transcription of spoken language is a challenge, especially for Norwegian. Using dialect also in official circumstances is more and more accepted in Norway. NRK has defined two pronunciation standards based on the written standards of Norwegian: Bokmål and Nynorsk. Reading of manuscript based news should comply with one of these standards, but this principle has been softened recently. The majority of RUNDKAST is not manuscript read and therefore not compliant to the NRK standard pronunciations. The pronunciation is then more or less flavored by the speaker's dialect.

In previous large orthographically annotated Norwegian speech databases, the speech was either read or from a restricted part of the South-East region closely associated with Bokmål (Johannesen et al, 2007). In the Norwegian part of SpeechDat a mix of Bokmål and Nynorsk was allowed (Johansen et al, 1997).

A Norwegian's choice of writing standard will probably not influence her or his spoken language, and there is no unambiguous mapping from dialect to written standard. Two speakers of the same dialect using fairly similar pronunciation may chose different writing standard. Still, Bokmål is used by the vast majority in the South-East, Mid and North Norway while Nynorsk is used more frequently in West Norway.

The challenge is therefore not the two written standards (many countries have more than one), but to define how to classify spoken Norwegian by a given speaker to the appropriate standard. Both standards include many alternative spellings and inflection variants, some corresponding to dialect pronunciations. The standards are moreover continually changing. As an example, an alternative spelling of the number 7 was added during the RUNDKAST specification ("syv" in addition to "sju"). We have used the spelling as defined in the dictionaries "Bokmålsordboka" and "Nynorskordboka"<sup>6</sup> that is closest to the actual pronunciation used.

### 3.2.2 Orthographic transcription conventions

The aim of the conventions for the orthographic transcription in RUNDKAST is to minimize uncertainty about pronunciations and facilitate consistency. Each speaker's language variant was classified as being closer to the written standard of either Bokmål or Nynorsk. The orthographic transcription is therefore in compliance with one of these standards with a few exceptions (restrictions). For manuscript read speech we have used the writing standard of the speaker. For dialect speech we have chosen the written standard where as many words as possible are inside the standard. The final conventions for the orthographic transcription were based on experiences from (LDC, 2004) and (COST278, 2003) as well as the annotation guidelines provided with *Transcriber*:

- Words are transcribed with the written forms that are closest to actual pronunciations. A limited number of interjections are allowed.
- Text codes are used to mark mispronunciations, truncations, and unknown words<sup>7</sup>.
- Numbers and symbols are written out as words (e.g. three point six).
- Space is used between spelled letters, also when acronyms have spelled pronunciation.
- Capital letters are used in proper names, spellings, and acronyms (e.g. London, B B C, and NATO), but not at the start of sentences.
- Abbreviations are not used.
- Compound words not in the dictionary are usually written with a hyphen. This is mandatory when one of the roots is a proper name (e.g. Telenor-ansatt).
- Punctuation marks are restricted to comma, period, and question mark.

	Type	Label
<b>Speech not transcribed</b>	Unintelligible	[?]
	Overlapping	[o]
	Not Norwegian	[lang=country]
<b>Speaker noise</b>	Inhalation	[i]
	Exhalation	[u]
	Vocal hesitation	[e]
	Nasal hesitation	[m]
	Other	[s]
<b>Pauses</b>	Short pause (> 200 ms)	[p]
	Long pause (> 1 sec)	[lp]
<b>Transient background noise</b>	Speech	[t]
	Other	[b]

Table 5 Acoustic event categories

*Transcriber*'s mechanism for annotation of events (category, description, extent) is used for acoustic events as shown in Table 5. These categories are all instantaneous, except for the transient background noise which may be instantaneous but typically is applied to one or more affected words.

<sup>6</sup> <http://www.dokpro.uio.no/ordboksoek.html> (in Norwegian)

<sup>7</sup> Typically dialect pronunciation, or words a Bokmål-defined speaker borrows from Nynorsk or vice versa.

### 3.3 Broad phonetic annotation of a subset

A subset consisting of one hour of speech from 10 speakers (5 female and 5 male) was manually annotated in broad phonetic segments to provide material for research on utilization of phonetic knowledge in speech technology. We selected 5 minutes of planned speech and 1 minute of spontaneous speech per speaker. This material consists solely of segments with studio bandwidth, high fidelity and a very low level of background noise.

For such a limited number of speakers we did not try to achieve any dialect coverage. On the contrary, we selected only Bokmål-defined speakers in order to limit the variability. In spite of the Bokmål limitation the speakers are rather heterogeneous. Most of the 10 speakers come from the "traditional Bokmål-region" (South-East Norway), but the North Norway, Mid Norway and South Norway regions are also represented. The subset therefore consists of speakers using the Bokmål pronunciation standard, but with some accent variation from speaker to speaker.

The annotation was to be mainly phonemic, but acoustic boundaries were marked following the guidelines for TIMIT (Garfalo et al, 1993). In defining the guidelines we also relied on experience gained from annotation of Norwegian databases for speech synthesis (Amdal & Svendsen, 2006).

We used SAMPA format symbols (Wells, 1997) with a core set based on the Norwegian SAMPA phoneme inventory<sup>8</sup>. We supplemented the inventory with some allophones and acoustically motivated symbols to mark major acoustic boundaries in accordance with the guidelines for TIMIT. It is therefore more correct to use the term "broad phonetic" than phonemic. In TIMIT the term "acoustic-phonetic" is used for the same type of labels.

Only symbols for allophones we believe can be differentiated fairly reliably were included, such as the (dialectally based) distinction between apical and dorsal /r/. A small number of English phonemes that are regularly used in Norwegian pronunciation of English loan words were also included in the inventory, see Table 6. The closure and release of plosives are separated into two segments and the release is categorized regarding aspiration. We added syllabic versions of all sonorants and a voiced version of /h/.

Type	Label(s)
Glottal stop	[ʔ]
Additional allophones of /r/	[X] and [R]
Palatal allophones	[c], [J], [j], and [L]
From British English SAMPA	[eI], [@U], [aU], [T], [D], [z], [Z], [w], and [r\]

Table 6 Specification of amendment to Norwegian SAMPA phoneme inventory

#### 3.3.1 Broad-phonetic annotation specification and procedure

The main principle for the annotation was that consistency in annotation is of utmost importance! This resulted in the following supervisory guidelines for the annotation:

- A transcription as close as possible to the citation form is preferred.
- Norwegian standard SAMPA is preferred over English symbols.
- The waveform is the most important information source for setting segment boundaries.

Examples from the further restrictions in the detailed guidelines:

- Vowel length should be labeled phonemically.
- For syllabic consonants a transcription using /@/ (*schwa*) is preferred if there is any evidence of a preceding /@/.

The annotation scripts were based on *Praat*. We used the following annotation procedure:

1. Conversion of orthographic transcription to a format suitable for automatic transcription.
2. Automatic segmentation with a phonotypical transcription using a speech recognizer.
3. Manual correction of both segments and labels by four phonetics students.
4. Format check.
5. Control of all annotation by one supervisor.

The four annotators were given an introductory course including a test annotation session. The supervising annotator corrected and commented these test annotations to ensure common annotation practice. All the annotated material was finally controlled by the supervising annotator, who also reported possible deviating solutions to the respective annotators during the annotation period. The annotators reported difficulties in detailed logs to the supervisor.

We chose to let one supervising controller do all the verification work. Alternatives include several annotators working independently, or annotators correcting each other's annotations. The main reason for our solution was that we believed that this strategy would give the most consistent annotation result. In addition, a single supervisor simplifies the administration of the project.

#### 3.3.2 Format of the broad-phonetic annotation

The annotation consists of four tiers: "phone", "phone comment", "word", and "utterance". In an intermediate format we included a fifth tier "auto" to show the original automatic annotation to the supervising annotator, see Figure 3. The labels in the "auto" tier were never changed by the annotators, but boundaries could be moved. SAMPA symbols were used for the linguistic segments in the phone tier labels. We also included extralinguistic symbols such as breathing and hesitation types, using symbols corresponding to those used in the orthographic transcription.

<sup>8</sup> <http://www.phon.ucl.ac.uk/home/sampa/norweg.htm>

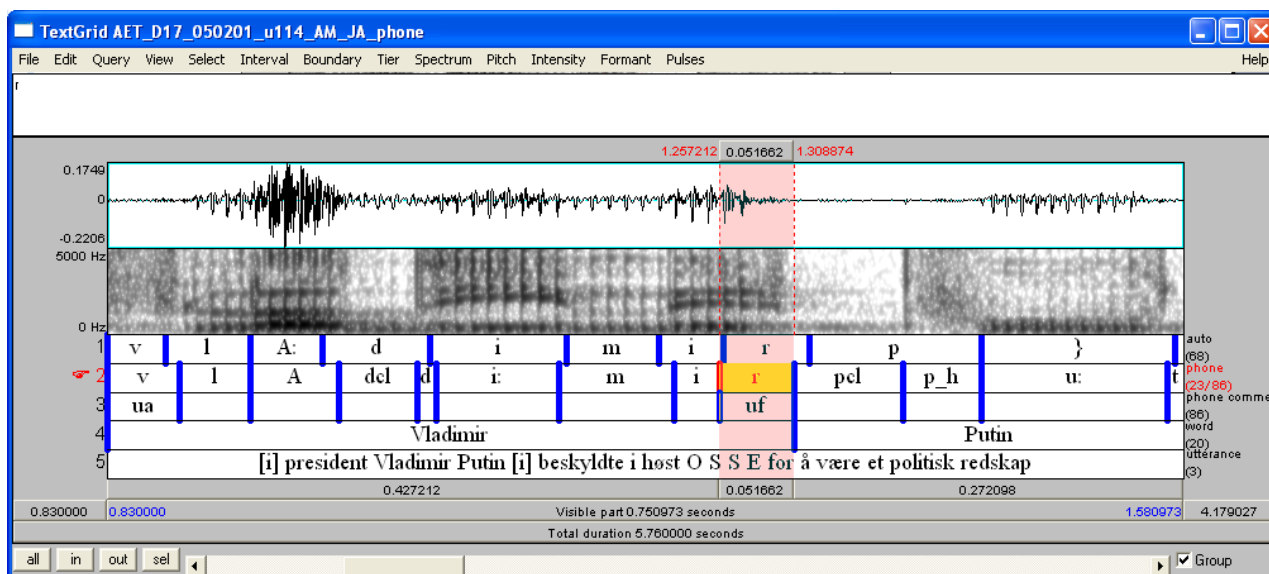


Figure 3 Example of broad phonetic annotation using *Praat*.  
(The auto tier contains the automatic transcription suggestion and is not included in the final file.)

The phone comment tier is synchronized with the phone tier and consists of a set of standard comments for common variation, e.g. devoicing of a voiced phone, epenthetic sounds and pauses, or creaky voice. Furthermore, a majority of the standard comments also include information defining which part of the segment the comments describe.

An example: In Figure 3 the /r/ in "Vladimir" has the comment "uf" that means that it is unvoiced finally. Among other frequent comments are alternative phone symbol suggestions when the annotator has been in doubt. Furthermore, Figure 3 shows that the bursts of voiced plosives (as for the "d" in Vladimir) are also annotated. For the unvoiced plosives the bursts were separated into two different categories: unaspirated and aspirated. The "p\_h" in "Putin" is an example of the latter type.

## 4 Discussion

### 4.1 Availability

It is the intention of both NTNU and NRK that the database shall be made available for speech technology research, and that it shall be included in a planned collection of language resources for Norwegian, a Norwegian language bank. There are however some issues regarding intellectual property rights, particularly concerning musical segments, that need to be resolved before the database can be considered for distribution to the research community.

### 4.2 High level annotation

The transcribers working on the high level annotation were asked to report problems in logs and ask for help or advice when necessary. Their work was also regularly evaluated by a supervisor. These efforts gave valuable insight into problematic aspects of the high-level annotation. One example is the treatment of background

noise, particularly whether it was too weak to mark, or too sporadic to be regarded as lasting. Transcription of mispronunciations and speaker noises, categorization of fidelity, and finding the written form of words closest to actual pronunciations were all associated with similar trade-off difficulties. Such problems are increased by drift in subjective norms between transcription sessions.

For simplicity we chose to limit the number of speaker noise events to five, see Table 5, including an open event *other*. RUNDKAST users should be aware that the latter therefore covers a very wide range of speaker noise phenomena.

Another issue is the use of punctuation marks. It is usually quite manageable to place punctuation marks in manuscript read speech, but this is often much more difficult in spontaneous speech. Sometimes punctuation marks are less useful, for instance in the case of repeated utterance restarts, or when one speaker in the middle of an utterance loses his or her turn to another interrupting speaker. Nevertheless, we chose to use punctuation marks, and asked the transcribers to use their best judgment in problematic cases.

Lastly there were two issues regarding speaker IDs. Finding the correct orthography for speaker names was more problematic than expected. The transcribers were asked to use Internet sources and tools when needed, but for many speakers this was still a difficult and time-consuming task. Another difficulty was the classification of a speaker's language variant. Some speakers have pronunciations that fall between the categories in Table 2. Moreover, some speakers tried to follow one of the two NRK pronunciation standards for manuscript read speech, but reverted to their original dialect for spontaneous speech. Ensuring correct or at least consistent naming and language variant classification for speakers across transcription files was therefore a prioritized subtask in the internal validation.

The high-level annotation required about 70 hours of work per hour of speech, including various support to transcribers and internal validation, but excluding administration.

### 4.3 Broad-phonetic annotation

Broad-phonetic annotation is a time consuming task, the entire annotation of 1 hour required approximately 1000 working hours including control and support, but excluding administration. We also encountered several phenomena that were difficult to label consistently. The phone comment tier provides structured information on the doubt cases. We believe this will be an advantage e.g. for research on fine phonetic detail.

The annotators found that the categorization of aspirated versus unaspirated bursts of unvoiced plosives was rather problematic. Future users of RUNDKAST should therefore consider treating aspirated and unaspirated bursts as a single category.

## 5 Conclusions and further work

RUNDKAST is the first Norwegian broadcast news speech corpus annotated for research in speech technology. The intention is to provide resources for research on various topics in speech technology, in particular audio indexing. The corpus development has utilized existing tools and guidelines where available. We have defined conventions for orthographic and broad phonetic annotation of Norwegian spontaneous speech containing heterogeneous speaking styles and dialect usage. The documentation of our efforts may be used in the production of future databases and will hopefully lower the threshold for similar data collections.

The availability of user interfaces in the domestic language is vital to both productivity and usability and for supporting the Norwegian language in an increasingly English-dominated international society. RUNDKAST is planned to be included for non-commercial use in a future Norwegian language bank where it will complement other corpora also intended to be included<sup>9</sup>. The most important further work is to make the Norwegian language bank really happen, but that is beyond the power of the authors.

## 6 Acknowledgements

The authors wish to thank the Faculty of Information Technology, Mathematics and Electrical Engineering at NTNU for funding the project, NRK for providing data, and last but not least the annotators for their hard work on a tedious task.

## 7 References

- Amdal, I., Svendsen, T. (2006). FonDat1: A Speech Synthesis Corpus for Norwegian. In *Proceedings of LREC-2006*. Genova, Italy
- Barras, C., Geoffrois, E., Wu, Z., Liberman, M. (2001).

Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication* 33(1-2) pp 5–22

- Boersma, P., Weenink, D. (2006). *Praat: doing phonetics by computer*. Accessible from <http://www.praat.org/>
- COST278 (BN special interest group) (2003): *Report on the Lisbon Broadcast News Workshop*. Accessible from [https://speech.elis.ugent.be/cost/database\\_report.pdf](https://speech.elis.ugent.be/cost/database_report.pdf)
- Garfalo, J., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L. (1993). *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Printed Documentation NISTIR 4930: LDC
- Johannessen, J.B., Hagen, K., Priestley, J., Nygaard, L. (2007). An Advanced Speech Corpus for Norwegian. In *Proceedings of NODALIDA-2007*. Tartu, Estonia.
- Johansen, F.T., Amdal, I., Kvale, K. (1997). The Norwegian part of SpeechDat: A European speech database for creation of voice driven teleservices. In *Proceedings of NORSIG-1997*, Tromsø, Norway
- LDC (Linguistic Data Consortium) (2004): *Guidelines for RT-04 Transcription*. Accessible from <http://projects.ldc.upenn.edu/Transcription/>
- Lee, C.-H., (2004). From knowledge-ignorant to knowledge-rich modeling: A new speech research paradigm for next generation automatic speech recognition. In *Proceedings of Interspeech-2004*, Jeju, South Korea
- Makhoul, J., Kubala, F., Leek, T., Liu, D., Nguyen, L., Schwartz, R., Srivastava, A. (2000). Speech and Language Technologies for Audio Indexing and Retrieval. *Proceedings of the IEEE* 88(8) pp 1138–1353
- Moore, R.K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication* 49(5) pp 418–435
- Schuurman, I., Schoupe, M., Hoekstra, H., van der Wouden, T. (2003). CGN, an Annotated Corpus of Spoken Dutch. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*. Budapest, Hungary.
- Svendsen, T., Nordgård, T., Andreassen, L.H., Berg, J.T., Kvale, K., Espeli, T., Johansson, S., Breivik, T. (2002). Consolidating and increasing the availability of Norwegian human language technology resources. Report. *The Norwegian Ministry of Culture and Church Affairs*. Accessible from [http://www.sprakrad.no/upload/1308/sbank\\_en\\_prosjektrapport.pdf](http://www.sprakrad.no/upload/1308/sbank_en_prosjektrapport.pdf)
- Wells, J.C. (1997). SAMPA computer readable phonetic alphabet. In Gibbon, D., Moore, R., Winski, R. (Eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B
- Žibert, J., Mihelič, F., Martens, J.-P., Meinedo, H., Neto, J., Docio, L., Garcia-Mateo, C., David, P., Zdansky, J., Pleva, M., Cizmar, A., Žgank, A., Kačič, Z., Teleki, C., Vicsi, K. (2005). The COST278 Broadcast News Segmentation and Speaker Clustering Evaluation - Overview, Methodology, Systems, Results. In *Proceedings of Interspeech-2005*. Lisboa, Portugal.

<sup>9</sup> Updates on the Norwegian language bank can be found at <http://www.sprakrad.no/Spraakstyrking/IKT/Spraakteknologi/Spraakbank/> (in Norwegian)