

# ***Acoustic database for Norwegian***

## **1. BASIC INFORMATION**

### *1.1 Corpus composition*

#### **Databases for speech recognition/ dictation**

SpeechDat database (mobile phone and landline data for Norwegian, Swedish and Danish)  
Recordings for speech recognition: ADB\_OD\_Nor.NOR database (recordings in office environment based on phonetically balanced manuscripts produced from sentences from the NST Norwegian corpus ) Recordings for dictation: ADB\_D\_IBM-N database produced from recordings in office environment based on phonetically balanced manuscripts produced from text material from the newspaper *Aftenposten*. Recordings for telephone services: ADB\_T\_Nor.NOR landline and mobile telephone recordings Database of non verbal sounds marking hesitation

#### **Databases for speech synthesis**

RealSpeak  
IBM speech synthesis  
IBM Phrase Splicing

### *1.2 Representation of the corpora (flat files, database, markup)*

Database

### *1.3 Character encoding*

Not relevant

## **2. ADMINISTRATIVE INFORMATION**

### *2.1 Contact person (name, e-mail)*

Arne Martinus Lindstad, [ArneMartinus.Lindstad@nb.no](mailto:ArneMartinus.Lindstad@nb.no)

### *2.2 Copyright statement and information on IPR*

Open Source

## **3. TECHNICAL INFORMATION**

### *3.1 Data structure of an entry*

Soundfiles in PCM/wav-format

Spl-loggfiler in txt

### *3.2 Corpora size (num. of tokens)*

873915 recordings

## **4. CONTENT INFORMATION**

### *4.1 Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)*

Monolingual

### *4.2 The natural language(s) of the corpus*

Norwegian

### *4.3 Domain(s)/register(s) of the corpus*

General speech

Phone conversations

### *4.4 Annotations in the corpus (if an annotated corpus)*

Not relevant

### *4.5 Intended application of the corpus*

Speech recognition, dictation, speech synthesis

### *4.6 Reliability of the annotations (automatically/manually assigned) – if any*

Automatically transcribed, validated by groups led by professionally trained linguists.

## **5. RELEVANT REFERENCES AND OTHER INFORMATION**

<http://www.nb.no/spraakbanken/tilgjengelege-ressursar/taledatabasar>