Samrómur Children: An Icelandic Speech Corpus

Carlos Mena, David Erik Mollberg, Michal Borský, Jón Guðnason

Language and Voice Lab, Reykjavík University

{carlosm, de14 ,michalb, jg}@ru.is

Corpus Characteristics

The main aspects of Samrómur Children are:

- Samrómur Children is an Icelandic speech corpus intended for the field of Automatic Speech Recognition.
- It contains 131 hours of read speech from Icelandic children aged between 4 to 17 years.
- Audio Format: Linear PCM, 16khz@16bit, 1 channel.
- In terms of time-length, the average is around 3.4 seconds approximately.

Corpus Portions

The Table shows the corpus portions broken down into gender of the speakers.

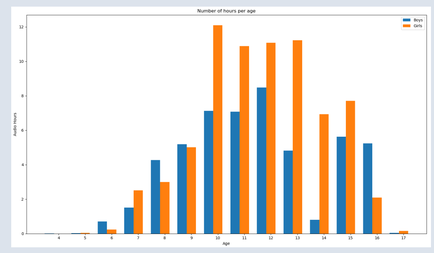
Gender	Female	Male	Unknown
Duration	$73h38m$	$52h26m$	$05h02m$
Utterances	$78,993$	$53,927$	$4,677$
Speakers	$1,667$	$1,412$	96

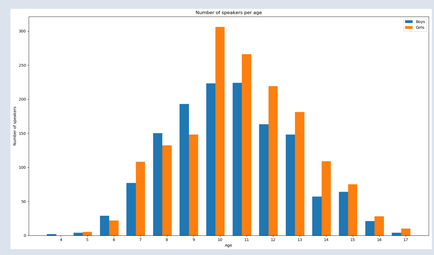
• However, there is a few number of speakers with unknown gender information.

• The speakers with unknown gender information are used in the development portion of the corpus.

Audio Per Speaker

The Figures show the hours of audio per range of age (top) and the number of speakers per range of age (botton) of the whole corpus.





Train Portion

The training portion of the corpus has a total of $134,394$ utterances from $2,517$ speakers. The total duration of this portion is 127 hours and 25 minutes.

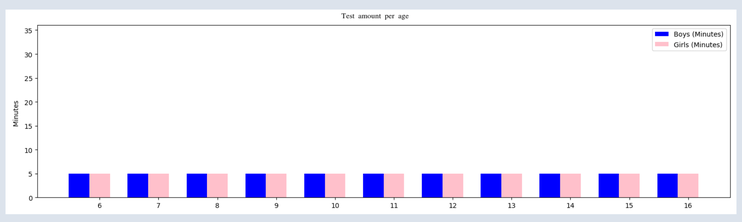
The Table shows the statistics of the train portion broken down into gender of the speakers.

Gender	Female	Male	Unknown
Duration	$72h43m$	$51h30m$	$03h11m$
Utterances	$78,148$	$53,058$	$3,188$
Speakers	$1,357$	$1,097$	63

Test Portion

The test portion of the corpus has a total of 845 utterances from 625 speakers. The total duration of this portion is 1 hour and 50 minutes. The shows the statistics of the test portion broken down into gender of the speakers.

Gender	Female	Male	Unknown
Duration	$00h55m$	$00h55m$	$0h0m$
Utterances	845	869	0
Speakers	310	315	0



Development Portion

The development portion of the corpus has a total of $1,489$ utterances from 34 speakers. The total duration of this portion is 1 hour and 50 minutes.

The Table shows the statistics of the development portion broken down into gender of the speakers.

Gender	Female	Male	Unknown
Duration	$00h00m$	$00h00m$	$01h50m$
Utterances	0	0	$1,489$
Speakers	0	0	34

Kaldi Setup

We followed the Kaldi recipe designed for the TED-LIUM corpus:

- As a first stage, we generated an HMM triphone model with LDA/MLLT training adaptations.
- Next is to augment the training data using speed perturbation in two different ratios with respect to the original speed (0.9 and 1.1).
- We Calculated iVectors for the whole corpus (including the augmented data).
- We implemented a TDNN-LSTM network.

Language Model

The language model was created using the Icelandic Gigaword corpus as well as the training prompts.

- The Gigaword corpus contains text from: newspapers, speeches, books, etc.
- The resulting text has a length of more than 44 million lines of text (5.3 GB approximately).
- It was created a 3-gram LM for decoding.
- And a 4-gram LM for re-scoring.
- We used the SRILM toolkit to produce the LMs.

Pronouncing Dictionary

The pronouncing dictionary was created with the words extracted from the text for the language model.

- To produce the pronunciations we used Sequitur-G2P which is a trainable grapheme-to-phoneme tool.
- The resulting pronouncing dictionary contains more than 960,000 entries.
- The phoneme set of Icelandic counts with 58 phonetic symbols in IPA.

%WER Children vs Adults

WER results obtained with Children Corpus

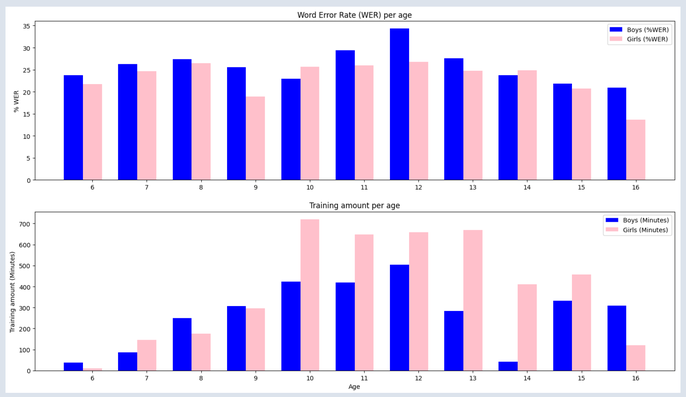
Experiment	%WER Dev	%WER Test
Kaldi HMMs	32.76	43.71
Kaldi LSTM	21.11	24.47

WER obtained with adult's speech of Samrómur

Portion	Duration	Best %WER
Dev	$15h16m$	11.48
Test	$15h51m$	12.98

Training Amount and WER

Notice how the %WER is spread almost evenly through all the ranges of age. This is an interesting behavior found with our experiments.



%WER Per Age

Age (years)	%WER Fem	%WER Male
6	21.7	23.7
7	24.7	26.3
8	26.5	27.4
9	18.9	25.6
10	25.7	22.9
11	26.0	29.4
12	26.8	34.3
13	24.8	27.6
14	24.9	23.7
15	20.7	21.8
16	13.7	20.9