



# Best possible speech recognizer on your own data

#15



Tomáš Sýkora  
Ing. Igor Szőke Ph.D.

## How to beat Google speech recognition:

First pass transcription

Automatic timestamps

Timestamp correction +  
partial text normalization

Automatic normalization

Automatic cleaning

Speed perturbation

Room impulse responses

18 hours of Parrot data

500 hours Supreme Court

GMM phoneme alignmet

Speaker adaptation

CNN-TDNNF

Recurrent rescoring

Dataset	Size	Weight	Perplexity
Court decisions	15.5M	0.2	340
Legal transcriptions	2.5M	0.55	176
Supreme Court	1.5M	-0.1	337
Supreme Court from before 2000	2.5M	-0.1	340
Parrot transcriptions	17k	0.45	139
Interpolated LM	-	-	123

AM	Training info: LM	Dataset	WER [%] test (10h)
google-asr	unknown	unknown	35.1
cnn-tdnnf	4-gram	18h parrot + 200h aug. + 990h s.c.	24.42
cnn-tdnnf	4-gram + lat. rescoring	18h parrot + 200h aug. + 990h s.c.	<b>22.32</b>