Delft University of Technology, Bachelor Seminar of Computer Science and Engineering

Training and testing the TDNN-OPGRU acoustic model on English read and spontaneous speech

Georgi Genkov
Supervisor: dr. Siyuan Feng
Responsible Professor: dr. Odette Scharenborg

TU Delft

June 27, 2021

Abstract

Automatic phoneme recognition (APR) is the process of recognizing phonemes (spoken sounds) in a recording of speech. It can be used for any application requiring fast and accurate transcription, i.e. a courthouse. This research creates such a model using the TDNN-OPGRU architecture and trains it on two datasets of recorded English speech - "TIMIT" for prewritten sentences being read out (prepared/read speech) and "Buckeye" for recorded interviews (spontaneous speech). The results of the model are analyzed and compared to similar research. The main conclusion is that the results obtained do not exceed previous research and in some cases are considerably worse. The reasoning for that is also included.

1 Introduction

Automatic speech recognition (ASR) is the process of transcribing a speech signal into a sequence (string) of words. It is a well-developed field of machine learning but it has one severe limitation: it can only recognize predefined words in its lexicon. This makes it inconvenient for situations requiring a lot of technical jargon or colloquialisms. In order to be able to deal with words outside of the lexicon, automatic phoneme recognizers (APR) are often employed. APRs convert a recording of speech into a sequence of phonemes - the smallest distinct unit of speech, often representing a single sound [1]. Since there is a small and finite amount of phonemes in each language, a neural network for recognizing them can be constructed, allowing for arbitrary words to be recognized. There exists research on the performance of APRs [2], but they typically focus on evaluating different neural network implementations on the same dataset. For the Research Project "What is the Best Automatic Phoneme Recognition System?" me and my peers - Jordy van der Tang, Irene Klom and Mihail Chiroşca - will evaluate multiple deep learning architectures on multiple languages and speech styles. In particular the acoustic models TDNN-OPGRU and TDNN-BLSTM will be trained and evaluated on both English and Mandarin spoken in two ways - reading out prepared sentences and conversing freely. After these researches it will be easy to compare these two acoustic models based on language and manner of speech.

1.1 Research Questions

The overall research question for this project is "What is the Best Automatic Phoneme Recognition System?". In particular, the aim of this research is to "Train and test the TDNN-OPGRU acoustic model on English read and spontaneous speech". After performing evaluation on the trained model, error analysis and comparisons with the results of peer students' models will be made to determine the advantages and disadvantages of TDNN-OPGRU and TDNN-BLSTM. The acoustic model configurations will be based on the work of R. Levenbach in his Master thesis "Phon Times: Improving Dutch phoneme recognition" [3].

From this research the following sub-questions will be answered:

- How does the trained TDNN-OPGRU perform for prepared English speech from the speech corpus TIMIT?
- How does the trained TDNN-OPGRU perform for spontaneous English speech from the speech corpus Buckeye?
- How do the results for English prepared speech produced by TDNN-OPGRU compare to the other research for TIMIT?
- How do the results for English spontaneous speech produced by TDNN-OPGRU compare to the other research for Buckeye?

2 Problem Description

As previously mentioned the research topic presented is to train the TDNN-OPGRU acoustic model for two speech corpora: "TIMIT Acoustic-Phonetic Continuous Speech Corpus" and "Buckeye Speech Corpus". The research topic will be further broken down in the following sub-topics:

- Process the English prepared and spontaneous speech datasets in a format, understood by the speech recognition software Kaldi.
- Design a TDNN-OPGRU acoustic model, train it on English prepared speech and perform model testing to receive initial results.
- Adjust the parameters of the acoustic model to obtain better results and perform training and testing with the improved models for both the English prepared and spontaneous speech datasets.
- Analyse the results of the trained models, perform error analysis and compare with the findings of peers.

These sub-topics have to be completed sequentially as each one requires knowledge or data from the previous one.

The TIMIT and Buckeye speech corpora contain recordings of English speech (prepared and spontaneous respectively), transcriptions and phonetic labelling of the conversations. This data will need to be processed to a specific format, in order to be understood by the software used. Due to the size of the corpora, this process will be automated using Bash shell scripts. The TDNN-OPGRU is a hybrid model combining time-delayed neural networks [3] and the Output-gate Projected Gated Recurrent Unit [4] architecture. The acoustic model will be implemented using the Kaldi speech recognition toolkit [5] once again using Bash shell scripts.

3 Experimental Setup

This section describes the experimental setup used for this research. First, the data preparation performed on both datasets are described in detail (Section 3.1). Then the parameters used for the acoustic model and the reasoning behind them is provided (Section 3.2). Finally the method used for evaluation of the results is given (Section 3.3).

3.1 Data Preparation

The TIMIT corpus contains recording of 630 American speakers, divided in eight major dialect and a transcription for every recording on both word on phoneme level [6]. Each speaker reads out 10 sentences for a total of 6300 utterances. The phoneme transcriptions provided by TIMIT contains 60 different phonemes, however these are reduced to 48 during the data preparation step. This is done because the corpus contain rare phonemes that sound very similar to another phoneme and thus they can be reduced to their more common counterpart, resulting in a simpler model.

The Buckeye corpus contains conversational speech interviews of 40 speakers from Columbus, Ohio [7]. The transcription contains 59 different phonemes which were not reduced to a lower number due to time constraints.

The setup provided by our supervisor Siyuan Feng already contains shell scripts for processing the Timit corpus to a format understood by Kaldi. This setup uses the full training data of 462 speakers (326 male and 136 female divided by 8 dialects) and the smaller "core" test set of 24 speakers - 16 male and 8 female. In collaboration with Irene Klom similar scripts were created for automatic pre-processing of the Buckeye corpus. To achieve this, a small subset of the Buckeye data is used, consisting of 5 speakers each speaking for about an hour. Four of the speakers (2 male, 2 female) are to be used as training data for TDNN-OPGRU and the other female speaker as testing data for evaluation of the model.

3.2 Acoustic model

In this experiment a DNN-HMM (Deep Neural Network - Hidden Markov Model) hybrid model is used. This means that a statistical model is combined with the Neural Network which helps with modelling variability like accents and dialects [3][8]. This involves a few steps, namely MFCC feature extraction, and the training of one monophone and three triphone models (Delta+delta-delta, LDA + MLLT and LDA + MLLT + SAT). These steps were already implemented in the default Kaldi recipe for TIMIT and were adapted from there[1]. Their purpose is outside the scope of this research, as it is more focused on the TDNN-OPGRU, but it can be read about here - https://www.eleanorchodroff.com/tutorial/kaldi/training-overview.html

These models will be the base of the experiments, as their parameters will be tweaked and the resulting outcomes analyzed. The following configuration was used as the initial setup of the TDNN-OPGRU:

[1] https://github.com/kaldi-asr/kaldi/blob/master/egs/timit/s5/run.sh

- 7 TDNN layers
- 3 OPGRU layers
- Layer layout: 3 TDNN/1 OPGRU/2 TDNN/1 OPGRU/2 TDNN/ 1 OPGRU/ Output layer
- Layer dimension (TDNN/OPGRU layers): 1024 / 512
- Training epochs: 6
- Dropout schedule: 0,0@0.20,0.3@0.50,0
- L2 Regularisation: 0.00005
- Mini-batch size: 128
- Frames per chunk: 150
- Chunk left and right context: 40;0
- Initial and final learning rate: 0.001 and 0.0001

These parameters were taken from Robert Levenbach's Phon Times paper [3] provided by our supervisors.

The shell scripts describing the configuration of the TDNN-OPGRU acoustic model were also created collaboratively with fellow peer Jordy van der Tang, whose research topic also involves using this model. This configuration describes the different neural network layers and other related parameters and will be adjusted further in the research to obtain a better model. Using these initial scripts the model was trained and tested a few times on the TIMIT dataset until satisfactory baseline results were obtained. Parameters that were experimented with were the number and positioning of layers, their dimensions, the number of training epochs, the minibatch size and the learning rates. This is the configuration used for these baseline results with the changes in bold:

- 7 TDNN layers
- 3 OPGRU layers
- Layer layout: 3 TDNN/1 OPGRU/2 TDNN/1 OPGRU/2 TDNN/ 1 OPGRU/ Output layer
- Layer dimension (TDNN/OPGRU layers): **256 / 128**
- Training epochs: **6**
- Dropout schedule: 0,0@0.20,0.3@0.50,0
- L2 Regularisation: 0.00005
- Mini-batch size: 128
- Frames per chunk: 150
- Chunk left and right context: 40;0
- Initial and final learning rate: **0.05 and 0.005**

Configuration 1: Baseline

In Section 4.2 a modified version of Configuration 1 is presented, which will be named Configuration 2.

- 7 TDNN layers
- 3 OPGRU layers
- Layer layout: 3 TDNN/1 OPGRU/2 TDNN/1 OPGRU/2 TDNN/ 1 OPGRU/ Output layer
- Layer dimension (TDNN/OPGRU layers): **256 / 128**
- Training epochs: **10**
- Dropout schedule: 0,0@0.20,0.3@0.50,0
- L2 Regularisation: 0.00005
- Mini-batch size: 128
- Frames per chunk: 150
- Chunk left and right context: 40;0
- Initial and final learning rate: **0.005 and 0.0005**

Configuration 2

3.3 Evaluation

To present the answers to the research questions the report will contain numerical and graphical analysis of the results on mainly two criteria: The Phoneme Error Rate (PER) and confusion matrices.

The phoneme error rate is calculated by taking the predicted sequence of phonemes and counting the number of differences compared to the known truth sequence. A difference is defined as either a deletion (a phoneme from the ground-truth sequence is not present in the predicted sequence), substitution (a phoneme from the ground-truth sequence has been changed to another phoneme in the predicted sequence) or insertions (a phoneme from the predicted sequence that is not present in the ground-truth). The actual phoneme error rate is calculated as follows:

$$PER = \frac{N_{substitutions} + N_{insertions} + N_{deletions}}{N_{allphonemes}}$$

A confusion matrix is a matrix used to visualize phoneme errors. Every row in the table is a ground-truth phoneme and there is one column for every phoneme and two additional columns for inserted and deleted phonemes. The diagonal of the matrix match a phoneme with itself, which gives the number of correctly classified phonemes. These numbers can also be represented as a percentage of successful classifications, or normalized as a value between 0 and 1.

In the simplified example (Figure 1) there are only 6 phonemes. As you can see, the phoneme /d/ is correctly recognized the most with 92.7% accuracy. Higher number on the diagonal (in bold) indicate better recognition. Numbers not on the diagonal are errors. The most common mistake is recognizing /b/ as /d/ with 4.2%.

	aa	ae	ah	b	d	ey
aa	**88.5**	2.2	4.1	0.2	0.2	0.3
ae	2.1	**86.8**	2.6	0.2	0.1	0.2
ah	2.2	2.6	**89.1**	0.1	0.4	0.5
b	0.1	0.2	0.1	**90.4**	4.2	0.1
d	0.3	0.1	0.3	4.0	**92.7**	0.1
ey	0.6	0.3	0.5	0.2	0.1	**90.2**

Figure 1: Example Confusion Matrix.

4 Experimental Results

This section is split in 3 main parts. First the initial (baseline) results, which were obtained without many changes to the configuration of the model (Section 4.1). Then the results and improvements for both the TIMIT and Buckeye corpora are presented in detail (Section 4.2) and the final results of the research are analyzed (Section 4.3).

4.1 Initial Results

After applying the data processing to the TIMIT dataset and training a base OPGRU model, the following results were obtained. The results for monophone and triphone HMM models are also included for comparison.

Stage	PER
Monophone	36.66%
Triphone1	30.48%
Triphone2	28.00%
Triphone3	25.56%
TDNN-OPGRU (initial run)	**32.57%**

Figure 2: Phoneme Error Rates of Monophone and Triphone models and the initial TDDN-OPGRU configuration

These results are quite poor as the TDNN-OPGRU performs worse than most of the basic HMM models preceding it. After determining that something was wrong with the configuration of the model, the initial assumption was that it was too complex for the small subset of that data which was used. Following a recommendation by the supervisor to reduce the number of layers and their dimension, a number of other configurations were tested but they all achieved similarly bad results. After more experimentation, it was deduced that increasing the learning rate by as much as 10 times drastically helped reduce the PER. This configuration was named "the baseline" and it's properties can be found in Configuration 1 described here.

TDNN-OPGRU (initial run)	32.57%
TDNN-OPGRU (Configuration 1)	**31.55%**

Figure 2: Phoneme Error Rates of Configuration 1 for TIMIT

4.2 Results and improvements for both corpora

After training on the Buckeye corpus with Configuration 1, the following results were achieved:

Stage	PER
Monophone	57.28%
Triphone1	55.35%
Triphone2	55.50%
Triphone3	47.77%
TDNN-OPGRU (Configuration 1)	**52.21%**

Figure 3: Phoneme Error Rates of Configuration 1 for Buckeye

The results for Buckeye are consistently worse than TIMIT in both the HMM and TDNN runs. This is somewhat to be expected as TIMIT contains every speaker reading out the same prepared transcript so the model might recognize these patterns. There are three other notable things in these results compared to the ones in Section 4.1:

- The Triphone 2 model performs slightly worse than the Triphone 1. This is somewhat unexpected as Triphone 2 build upon the results of Triphone 1 and can probably be attributed to the particular subset of speakers used.
- The Triphone 3 model improves 7.7% on the results of Triphone 2, compared to a 2.4% improvement with TIMIT.
- The TDNN-OPGRU model is better than Triphone 1 and 2, which means it performs significantly better for Buckeye than for TIMIT.

A noticeable improvement of the PER for TIMIT came from increasing the number of training epochs while decreasing the learning rate. This causes the TDNN to find more precise weights at the expense of more computational time. Decreasing the learning rate too much can also lead to the training process getting stuck at a high error rate [9]. This is how the results changed for both TIMIT and Buckeye using this technique:

# Epochs	Learning rates	TIMIT PER	Buckeye PER
6	0.01 and 0.001	31.48%	53.71%
10	0.005 and 0.0005	30.82%	51.79%
12	0.001 and 0.0001	30.82%	52.14%

Figure 4: Phoneme Error Rates of the iterative changes to the model

Here you can see that the modification with 10 epochs and 0.005/0.0005 learning rate produces the best result for both TIMIT and Buckeye. This configuration will be called Configuration 2 and will be used in the final analysis. Further more, the third modification doesn't improve the results for both corpora, probably because of the previously mentioned drawback of the technique.

4.3 Final results and analysis

After extensive testing, the model which achieved the best results for Buckeye and TIMIT was Configuration 2. Note that rerunning the same configuration multiple times might produce a slightly different results because of the inherent initial randomness of DNN layers. The PERs that Kaldi presents are 51.79% for Buckeye and 30.82% for TIMIT. However the scoring algorithm in Kaldi also counts the insertions and deletions of silences as errors but they are not relevant to the task of phoneme recognition. The remainder of this section will use the following corrected PERs to account for this:

	TIMIT PER	Buckeye PER
Kaldi PER	30.82%	51.10%
Corrected PER	**25.98%**	**49.31%**

Figure 5: Correcting the PER

This analysis will also refer to the confusion matrices that were generated from Kaldi's output. Due to the size of the matrices they can be found in Appendix A and B and only portions of them will be presented here.

4.3.1 Buckeye analysis

This is the breakdown for the number of errors and correctly identified phonemes for the Buckeye model:

- 11426 substitutions; 61.9% of all errors; 30.5% of all phonemes
- 764 insertions; 4.1% of all errors; 2.04% of all phonemes
- 6266 deletions; 33.95% of all errors; 16.74% of all phonemes
- 18968 correctly identified; 50.68% of all phonemes

When comparing this to the findings for TDNN-OPGRU in "Phon Times", there is a significantly higher error rate than for the Dutch spontaneous speech corpus [3]. This can be attributed to the limited number of training data used: About 6 hours for the Buckeye subset vs. about 140 hours for the Dutch corpus. This limitation is discussed more in Section 6. The ratios of the types of error also differs significantly - for the Dutch corpus 45.4% of all errors are substitutions, 23.7% - insertions and 30.8% - deletions. This is a rather even distribution of the errors when compared to the results obtained for Buckeye.

Compared to the findings for Buckeye using the TDDN-BLSTM acoustic model done by fellow peer Irene Klom [10], the PERs achieved are quite similar - 51.5% for OPGRU vs 54.03% for BLSTM. The PER is slightly higher for TDDN-BLSTM which is consistent with both the results from "Phon Times" and the results for the Chinese (Mandarin) corpora made by peers Jordy van der Tang and Mihail Chiroşca [11] [12].

When looking at individual phonemes, it quickly becomes apparent that the most frequent phonemes are also the ones that have been both correctly identified and mistaken the most. So below these results are presented normalized to the number of occurrences of each phoneme:

Phoneme	s	iy	ay	z	w
% Correctly identified	80.78	71.68	71.64	71.23	67.48

Figure 6: Most correctly identified phonemes Buckeye (normalized %)

Phoneme	aan	aen	ahn	ehn	eng	eyn	ihn	iyn	uhn	em	nx	own
% Correctly identified	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.24	8.51	9.25
Occurrences (test set)	6	11	21	18	4	3	34	3	1	165	411	54

Figure 7: Most incorrectly identified phonemes Buckeye (normalized %)

8 phonemes are never correctly identified but they are also very rarely used (all are under 35 occurrences). 3 phonemes were only present in the train set and never in the test set. This could have been solved by reducing the number of phonemes like for the TIMIT setup but due to the time limitations of the project this was not done.

There is very limited past research on APR or ASR architectures using Buckeye and most of them use Word Error Rate (WER) as a metric instead of PER. A Probabilistic model used for text-to-speech achieved the PER of 23.4% - significantly better than the model used here [13]. However, for this research a very small subset of the Buckeye corpus was used, so it is very hard to make comparisons.

4.3.2 TIMIT analysis

This is the breakdown for the number of errors and correctly identified phonemes for the TIMIT model:

- 1228 substitutions; 67.95% of all errors; 17.65% of all phonemes
- 169 insertions; 9.35% of all errors; 2.43% of all phonemes
- 410 deletions; 22.68% of all errors; 5.89% of all phonemes
- 5147 correctly identified; 74.01% of all phonemes

Similarly to the findings for Buckeye, when compared to Robert Levenbach's findings with Dutch prepared speech and TDNN-OPGRU, our PER for TIMIT is higher. [3] This can again be explained by the fact that the TIMIT corpus contains 462 speakers and each speaker only says 10 sentences (utterances) [6]. The distribution in errors is also very different as with the Dutch corpus 45.46% of the errors are substitutions, 20.34% are insertions and 34.18% - deletions. For the TIMIT model there are significantly more substitutions than insertions and deletions.

When compared to the experiments performed by Irene Klom with TIMIT and TDNN-BLSTM, her model once again performed slightly worse (at 31.78% PER) than TDNN-OPGRU. This is again consistent with Levenbach's findings in "Phon Times" and with the experiments performed by Jordy van der Tang and Mihail Chiroşca [11] [12].

When looking at individual phoneme's contributions to the error rate we can see that the phonemes themselves are quite different to the results from Buckeye and that for TIMIT the results are hardly dependent on how frequently a phoneme appears. Overall this shows that TDNN-OPGRU does not have a tendency to score some phonemes way better than others - although it is worth noting that in both corpora consonants are frequent in the top scoring section and vowels are common in the worst performing section.

Phoneme	jh	s	aw	cl	k
% Correctly identified	88.09	87.96	86.66	86.35	85.88
Occurrences (test set)	42	324	30	623	170

Figure 8: Most correctly identified phonemes TIMIT (normalized %)

Phoneme	epi	uh	ax	en	ih	ix
% Correctly identified	36.50	37.93	46.59	48.27	55.82	59.53
Occurrences (test set)	63	29	191	29	206	388

Figure 9: Most incorrectly identified phonemes TIMIT (normalized %)

Comparing the experimental results for TIMIT with past research on the corpus, it is clear that many other architectures outperform TDNN-OPGRU. However, for this research TIMIT was only tested on the smaller "core" test set, instead of the full test set which may alter the results. Here are some results from previous research obtained from a previous Bachelor literature review [2]:

Source	**Architecture**	**PER**
Ravanelli et al. (2018)	Li-GRU	14.9
Chorowski et al. (2015)	ARSG	16.8
Chorowski et al. (2014)	RNN	18.57
Schwarz et al. (2006)	TRAP	24.84
This research	*TDNN-OPGRU*	*25.98*
Graves & Schmidhuber(2005)	DBLSTM	30.2

5 Responsible Research

Since this research involves analyzing recorded speech from many people it is important to discuss the ethical implications that arise from that (Section 5.1). The reproducibility of the research will also be described in Section 5.2.

5.1 Ethical considerations

Both datasets contain sound files consisting of recordings of people speaking. The creators of Buckeye and TIMIT have anonymized the speaker information behind generic names, such as "s01" (for "Speaker 1") and "FAEM0" (for "Female, initials A.E.M") respectively. This removes any possibility of private information being handled or released inappropriately during this research. Further more, all of these recordings were hosted only on the TU Delft's HPC platform where the experiments were performed.

Both datasets also include general information about the age and gender distribution of the participants as well as some data about their location which can be useful for any applications that involve the accent/dialect of the participant. Understanding these distributions can help uncover some shortcomings of this research. For example, the Buckeye corpus only contains Caucasian speakers who are long-time residents of Columbus, Ohio, USA [7]. This severely limits the range of accents that the acoustic model can use to learn and it might perform worse on people from a different race or with a very different accent. The TIMIT corpus contains speakers from various regions of the United States but they are primarily employees of Texas Instruments [6] which limits their educational and professional background.

Since only a subset of the corpora is used it is important that this subset is representative of the whole data. Initially for Buckeye, the first 5 speakers were picked for the subset but that caused a great imbalance in the gender representation (4 female to 1 male) so one female speaker was exchanged for a male one to achieve a distribution similar to the one in the original dataset.

5.2 Reproducibility

The setup used for this research is based on the default Kaldi setup for TIMIT. Two important changes are made:

- In "run.sh" everything after "tri3 : LDA + MLLT + SAT Training & Decoding" is not used and is instead replaced with a Kaldi 'chain' model setup for TDNN-OPGRU.
- The scoring metric is changed to Word Error Rate (in reality Phoneme Error Rate as every 'word' is configured to be a phoneme) instead of the sclite scoring used for TIMIT.

The final source code and models for this research is available at a private TU Delft Git repository due to the licenses of the corpora [2]. The configuration used for the TDNN-OPGRU at any stage in the research is also given in Section 3.2 and the steps taken to process the data in Section 3.1. This should allow anyone familiar with Kaldi to reproduce the results obtained from this research. If necessary, files from the configuration can be provided after contacting by email.

6 Conclusions and Future Work

The purpose of this research was to train and test the TDNN-OPGRU acoustic model on English read and spontaneous speech - using the corpus "TIMIT" for prepared speech and the corpus "Buckeye" for spontaneous speech. To do so, the data from these corpora had to be processed to a specific format, the acoustic model had to be configured and the optimal settings discovered experimentally. Then the results were analyzed and compared with previous research and the findings of other students from the peer group.

The results obtained were as follows:

- TIMIT prepared speech + TDNN-OPGRU: **25.98%** corrected Phoneme Error Rate
- Buckeye spontaneous speech + TDNN-OPGRU: **49.31%** corrected Phoneme Error Rate

These results are poor compared to previous research but very similar to the results achieved by members of the peer group. This is due to several limitations which will now be discussed.

First and foremost, this researched used very limited data for simplicity and due to the short time span of the project. For the TIMIT corpus, only the "core" testing set was used which contains of the full test data provided. For the Buckeye corpus, only 6 out of 40 speakers were used to keep the overall recording length similar to that of TIMIT. Using the full data in these corpora may improve the results significantly.

Further more, during the data preparation part for Buckeye a decision was made to not reduce the number of phonemes to a number similar to TIMIT. This meant that there were more phonemes in the Buckeye corpus (59 vs 48) and many of these are only used a handful of times in the transcriptions. This made the models more complex and less accurate. Reducing some of the lesser used phonemes to a similar sounding phoneme (like is done in TIMIT) can improve the overall results.

Lastly, again due to the 10 week limitation of the project, only so many experiments with different configurations could be ran. Because of the licences on the corpora, these experiments could only be ran on TU Delft's HPC platform which gives students a 2 hour time limit and was frequently overloaded with jobs or down entirely. Given more time, a more optimal configuration of TDNN-OPGRU can be achieved.

References

[1] Britannica, T. Editors of Encyclopaedia. Phoneme. Encyclopedia Britannica. https://www.britannica.com/topic/phoneme

[2] https://gitlab.ewi.tudelft.nl/cse3000/2020-2021/rp-group-33/rp-group-33-ggenkov/

[2] van Geffen, H., Smit, M., Warners, A., Warners, F., Yarally, T. (2019). "A review of deep neural network-based phoneme recognition systems". BSc group project, Delft University of Technology

[3] Levenbach, R. (2021). "Phon Times: Improving Dutch phoneme recognition". MSc thesis, Delft University of Technology.

[4] Cheng, Gaofeng & ZHANG, Pengyuan & XU, Ji. (2019). Automatic Speech Recognition System with Output-Gate Projected Gated Recurrent Unit. IEICE Transactions on Information and Systems.

[5] About the Kaldi project.
`https://kaldi-asr.org/doc/about.html`

[6] Garofolo, J. S., Lamel, L. F, Fisher, W. M., Fuscus, J. G., Pallett, D. S, Dahlgren, N. L. (1993) DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus

[7] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

[8] Padmanabhan, Jayashree & Premkumar, Melvin. (2015). "Machine Learning in Automatic Speech Recognition: A Survey". IETE Technical Review.

[9] Goodfellow, I., Bengio, Y., Courville, A. (2016). "Deep Learning". MIT Press

[10] Klom, I. (2021) "Assessing the performance of the TDNN-BLSTM architecture for phoneme recognition of English speech". Research project, Delft University of Technology

[11] van Der Tang, J. (2021) "Evaluation of phoneme recognition through TDNN-OPGRU on Mandarin speech". Research project, Delft University of Technology

[12] Chiroşca, M. (2021) "Evaluating the performance of TDNN-BLSTM on Mandarin read and spontaneous speech". Research project, Delft University of Technology

[13] Qader, R., Lecorve, G., Lolive, D., Sebillot, P. (2015) "Probabilistic Speaker Pronunciation Adaptation for Spontaneous Speech Synthesis Using Linguistic Features". International Conference on Statistical Language and Speech Processing

A Buckeye Confusion Matrix

	aa	aan	ae	aen	ah	ahn	ao	aon	aw	awn	ay	ayn	b	ch	d	dh	dx	eh	ehn	el	em	en	eng	er	ey	eyn	f	g	hh	ih	ihn	iy	iyn	jh	k	l	m	n	ng	nx	ow	own	oy	p	r	s	sh	t	th	tq	uh	uhn	uw	v	w	y	z	zh	SIL	INS	DEL
aa	38.22		1.24		13.33	0.10	3.10		1.14		6.82		0.21		0.21	0.21	0.52	1.03		0.10		0.21		0.31	0.10		0.10	0.10	0.21	0.21		0.10			0.31	1.45	0.21	0.52			0.93			0.10	0.93	0.52	0.10	0.21	0.10	0.10	0.10		0.10		0.83	0.21	0.10			2.79	22.83
aan	16.67				16.67		16.67																																																						50.00
ae	2.69	0.11	52.42	0.11	5.80		0.21		1.83		1.83		0.11		0.43	0.64	0.97	7.52	0.11	0.21		0.43		0.32	0.43			0.11	0.43	2.58		0.64			0.32	0.54		0.54	0.11	0.21	0.32				0.64	0.32		0.21	0.97	0.97				0.21	0.11	0.54	0.43			1.61	12.03
aen			9.09															18.18												9.09		9.09						18.18																						18.18	18.18
ah	1.85		2.00		45.99		0.84		0.33		1.71		0.15		0.36	0.62	0.54	2.80		0.40	0.07	0.40		0.84	0.18		0.40	0.22	0.47	7.56		0.18		0.07	0.18	1.27	0.25	1.05	0.15	0.04	1.96		0.15	0.15	0.47	0.36		0.36	0.07	0.33	1.42		0.80	0.33	0.40	0.25	0.15			2.91	19.00
ahn					19.05													4.76												9.52						4.76		9.52			4.76								4.76				4.76		4.76						33.33
ao	8.57		0.26		6.75		41.56		0.26		1.82		0.52		0.26		0.26	0.78		0.52		0.78		0.26	0.26		0.52		0.52	0.78		0.26			0.78	2.34		1.04		0.26	3.64				0.52	0.52		0.52			0.26			0.52	2.60					2.86	19.22
aon																																																													
aw	5.59		9.50		3.91		0.56		43.02		2.79				0.56		1.12	2.23										0.56	0.56	0.56		1.12				2.79	0.56	1.12	0.56		5.59				2.79	0.56				0.56				0.56	1.12					0.56	11.17
awn																																																													
ay	2.74	0.19	1.04		3.78	0.09	0.66		0.09		71.64		0.09		0.28	0.09	0.47	0.85		0.09		0.09		0.19	0.47		0.38	0.09	0.09	0.57		0.66			0.09	0.66		0.85	0.09	0.09	0.38			0.09	1.51	0.09		0.09	0.19	0.09			0.19		0.19	0.09	0.09			0.95	9.55
ayn																																																													
b	0.54		0.27		2.31						0.41		38.13		4.75	8.01	0.81	0.41				0.27		0.68			1.49	2.04	0.54	1.36		0.14			0.14	1.09	1.49	2.58	0.14	0.14	0.14			1.76	1.22	0.41		0.14	0.54	0.41	0.41		0.54	7.87	1.36	0.14				1.22	16.15
ch	0.45				1.35		0.45				0.45			49.10	1.80			1.80				0.45					0.45	0.45	0.45			0.45		4.05	1.80										0.45	7.66	5.86	12.61			0.45						0.90	0.45		0.90	7.21
d	0.19		0.47		2.15		0.09		0.47		0.47		0.56	0.19	38.82	5.24	4.12	0.75		0.37		0.28		0.37	0.28		0.28	1.96	0.09	3.09		2.25		0.19	0.47	0.47	0.47	3.55	0.37	0.19	0.28	0.19		0.19	0.47	0.47		4.21	0.09	0.65	0.47		0.47	0.47	0.19	0.75	0.28			3.09	19.55
dh	0.47		0.36		1.42	0.12	0.36				0.47		0.95		2.96	43.08	1.78	1.07		0.12				0.24	0.36		1.78	0.24	0.59	1.66		0.12			0.36	2.01	0.24	2.37			0.12			0.59	0.83	1.07		0.59	2.72	0.12	0.12		0.47	1.54	0.24	0.24	0.71			2.72	24.85
dx	0.48	0.16	0.32		2.24		0.16		0.16		1.12				4.01	2.24	49.04	0.48		0.16				0.16	0.16		0.48	0.80	0.48	1.60		0.64				0.80	0.48	1.44		0.80	0.32		0.16	0.16	0.96	0.16		0.16		0.32	0.16		0.32	0.32	1.12	0.48	0.48			3.53	22.92
eh	0.54		6.49		9.64		0.13				1.34		0.27		0.27	0.87	0.67	38.09		0.07	0.13	0.33		0.33	1.07		0.07	0.07	0.27	10.11		0.74	0.07	0.07	0.20	0.27	0.13	1.47	0.07	0.20	0.67			0.07	0.33			0.27	0.27	0.33	0.40		0.20	0.07	0.07	0.13	0.13			2.88	20.21
ehn			5.56										5.56			5.56		16.67																														5.56							5.56						55.56
el					6.22				0.52				1.04		1.55	0.52		1.04		25.39					0.52					2.59		1.04			1.04	20.73	0.52	0.52			4.66				1.04	0.52		0.52			1.55		1.55		2.07					2.59	22.28
em	0.61				4.24										0.61	1.21	0.61		5.45	0.61	4.24	6.67		2.42	0.61				0.61	1.21		1.21				1.82	5.45	16.36								0.61		0.61	1.21	0.61	0.61		3.03	0.61			0.61				38.18
en	1.25		0.42		2.50						0.42				3.33		0.83	2.08		0.42	0.42	23.75	0.83	0.42	0.42		0.42			6.25		4.17			0.42		1.67	22.50	2.08				0.42		0.42			0.42		0.42	0.42		1.67		0.42	0.42	0.42			2.92	17.50
eng																																25.00				25.00																									50.00
er	0.42				3.09		0.84				0.14		0.28		0.42			1.13		0.28				60.90			0.56		0.28	1.55		0.14		0.28	0.28	0.42		0.84			0.84			0.14	14.06	0.28		0.42	0.14	0.56			0.28	0.14	0.14	0.14				0.28	10.69
ey			1.40		1.40						1.40				0.20	0.60	0.20	3.41			0.20	0.20		0.20	58.72		0.20	0.20	0.40	8.22		5.61			0.40	0.20	0.20	0.40	0.20						0.80			0.40	0.20	0.60	0.20		0.20		0.40	0.20				0.60	12.42
eyn																																						33.33	33.33																						33.33
f	0.36		0.36		0.36						0.18				0.18			0.36				0.18		0.54	0.36		65.53		0.90	1.08					1.08	1.08	0.18	0.54			0.18			2.15	0.36	1.08	0.54	2.15	2.15	0.36	0.72		0.18	2.15	0.18	0.36	0.72			1.26	12.21
g	0.86		0.22	0.22	1.51		0.22				0.65		0.65		6.48	1.94	1.51	0.22				0.22		1.51	0.65		0.86	44.28	0.43	1.30		0.65		0.22	7.99	1.08		1.08	0.65		0.22	0.22		0.22	0.65	0.22		0.65	0.22	0.22	0.65			0.22	0.22	0.65	0.22			1.51	18.57
hh	0.47		0.63		1.74		0.16		0.16		0.47		0.16		0.32	0.63	0.79	1.27	0.16			0.16		0.47	0.32		0.16	0.47	50.32	0.16		0.47			0.47	1.42	0.16	3.96	0.16	0.32	0.16			0.16	0.32	0.79		0.47		0.32	0.16		0.32		0.47	4.11	0.47			4.91	21.36
ih	0.13		0.68		6.85		0.23				0.78	0.03	0.13	0.07	0.72	0.78	0.55	4.01		0.07		0.42		0.88	1.24		0.23	0.13	0.36	50.80	0.07	4.27		0.07	0.13	0.59	0.16	1.11		0.03	0.39		0.10	0.07	0.52	0.42	0.13	0.49	0.13	0.23	0.88		1.30	0.29	0.16	0.36	0.36	0.10		2.02	16.53
ihn					14.71						8.82					2.94		5.88						2.94						17.65						2.94		11.76							2.94						2.94		2.94		2.94						20.59
iy	0.21		0.07		0.43		0.07		0.07		0.64		0.21	0.07	0.43	0.14	0.07	0.21				0.29		0.50	3.57		0.07	0.07	0.14	4.49		71.68		0.14	0.36	0.36		0.86	0.21		0.14				0.43			0.29	0.07				1.57	0.14	0.07	1.71	0.14	0.14		0.78	9.13
iyn																														33.33		33.33									33.33																				
jh					1.16									2.33	11.63									0.58	0.58			1.16	1.74	0.58				52.33	0.58			0.58								0.58	1.16	7.56									2.33	2.33		0.58	12.21
k	0.37		0.09		0.64		0.09				0.27		0.09	0.09	1.56	0.37	0.27	0.37		0.18		0.09		0.37	0.27		1.37	2.93	0.64	1.10		0.37			64.19	0.37		1.01	0.09	0.09	0.27		0.09	1.28	0.82	0.37	0.09	3.57	0.37	0.27	0.18			0.18	0.18		0.46			0.64	13.92
l	0.55				1.38		1.01		0.37		0.09		0.18		0.28	0.74	0.37	0.18		1.75		0.09		0.28	0.09		0.46	0.55	0.37	0.18		0.37			0.09	60.26	0.46	1.10	0.09	0.09	4.05		0.18		1.10	0.64		0.18			0.28		0.37	0.28	5.34	0.09	0.09			3.04	12.97
m	0.96		0.44		2.79		0.51		0.51		0.22		0.44		0.73	1.18	1.03	0.66	0.15	0.07	1.54	1.40		0.59			0.07		0.37	1.18		0.51			0.15	1.98	42.76	15.06	0.29	0.44	0.59			0.15	0.66	0.15		0.51	0.22	0.44	0.07		0.66	1.25	1.76	0.15	0.15			0.29	16.90
n	0.18		0.36	0.04	1.78		0.27		0.22		0.67		0.31	0.04	1.87	1.24	3.77	1.11		0.22	0.18	1.73	0.13	0.22	0.49	0.04		0.36	0.22	2.22		1.47			0.22	1.55	1.42	51.07	1.60	0.49	0.80	0.40			0.58	0.09		0.93	0.04	0.49	0.13		0.36	0.22	0.22	1.02	0.31			2.62	16.30
ng	0.28		0.56		0.56						1.67		0.28	0.28		0.84	0.28	0.84		0.28		0.28	0.28	0.28	2.79			0.56	0.28	0.84		1.39			0.28	1.11	0.56	16.99	45.40	0.56		0.28	0.28		0.28			0.28	0.28	0.28				0.28	0.56	1.11	0.28			0.84	17.83
nx	1.46		0.49		4.62		1.46				1.22		0.24		1.46	1.22	13.14	4.14				1.22		0.24	0.73		0.49	0.49	0.24	1.46		2.92	0.24			1.46	0.24	15.57	0.97	8.52	0.49				1.46	0.49				0.49			0.49	0.49	0.49	0.49	0.24			1.46	29.20
ow	1.48		1.38		12.73	0.11	3.39		0.53		0.32		0.21		0.85		1.06	1.48	0.11	0.53	0.11	0.11		0.95	0.32			0.32	0.21	1.38	0.11	0.53		0.11	0.21	6.89	0.95	1.48	0.42	0.11	39.66	0.21	0.21	0.11	0.64	0.21		0.21		0.42	0.74		1.38	0.32	1.38		0.21			1.91	14.00
own			1.85		3.70				1.85		3.70				7.41													1.85	3.70			5.56				5.56					12.96	9.26			1.85			1.85	1.85												37.04
oy							2.78								2.78			2.78							16.67					2.78		2.78											33.33										2.78		16.67						16.67
p	0.40		1.59		3.39				0.60		0.40		1.79		0.40	0.60		0.60		0.20				0.80			4.38	0.20	2.19	1.99		0.60			4.38	0.40		0.20			0.40			46.41	0.40			3.98	0.40	1.20	0.20		1.39	2.19	0.60	0.80	0.20			0.60	16.14
r	0.09		0.17		1.28				0.09		0.26		0.17		0.43	0.17	0.26	0.34		0.26		0.17		10.12	0.09		0.34	0.09	0.26	0.94		0.34		0.09	0.43	0.43	0.68	0.60			0.34		0.09	0.17	62.07	0.68	0.09	0.26	0.09	0.17	0.17		0.09	0.43	1.79		0.09			2.04	13.44
s	0.06		0.06		0.84		0.06				0.24				0.36	0.24	0.06	0.18		0.06		0.06		0.30	0.18		0.48			0.66		0.42			0.06	0.12	0.06	0.18			0.06		0.06	0.12	0.06	80.79	0.12	0.18	0.36	0.06	0.12		0.36	0.12	0.12		5.67	0.06		1.31	5.79
sh	0.34				0.34		0.34				0.34			3.02			0.34	0.67				0.34		0.67		0.34			0.67	1.01		0.67		0.67	0.67	1.01		0.67						0.34	1.01	14.09	59.06	1.01								0.34	1.34			0.34	10.40
t	0.22		1.01		1.80		0.29		0.07		0.72		0.29	0.65	5.83	1.51	1.51	0.94		0.07		0.22		0.36	0.29		0.65	0.36	0.36	1.80	0.07	0.29		0.43	1.58	0.29	0.07	0.86	0.14	0.14			0.07	0.58	0.29	2.16	0.07	42.19	1.51	1.87	0.36		0.50	0.07	0.14	0.36	0.65			2.74	23.61
th	0.52				0.79						0.52		0.26		1.31	6.04	0.26	0.52		0.26		0.52		1.05	0.26		7.35		1.57	1.57					1.31	0.52		0.26	0.26	0.52				1.05	0.52	2.62		2.89	40.42	0.26	0.26			2.10	0.26	0.26	1.84			1.57	20.21
tq	0.83	0.12	2.73	0.12	6.18	0.12	0.48		0.48		2.49		0.48	0.12	2.02	0.59	2.26	4.51		0.59				0.71	1.19		0.24	0.36	0.24	1.43		0.59			0.12	1.54	0.36	1.54	0.12		1.90	0.24		0.24	1.78	0.95	0.12	2.14	0.12	23.16	0.36		0.24	0.71	0.24	0.12	0.12			2.97	32.07
uh			0.27		12.00		0.27		0.27		0.27		0.53	0.53	0.53	0.53	0.53	1.33		0.27		0.80		0.27			0.53			12.80		1.87		0.27	0.27	2.40	0.27	1.33		0.27	2.67		0.27	0.27	0.53	0.80		0.80			19.73		2.93	0.80	1.07	0.80				1.60	29.33
uhn																																																						100.00							
uw					3.14				0.18				0.55		1.11	0.37	0.18	0.74		0.18	0.18	0.37		1.48	0.74			0.74	0.18	5.73		4.07		0.18	0.18	1.66		1.11	0.37		0.92			0.37	1.48	0.37	0.18	0.92		0.37	1.11		54.16		0.37	0.37	0.18			2.59	13.12
v	0.65		1.08		3.89		0.43		1.08		0.22		1.51		2.81	5.18	1.08	0.86		0.65		0.22		0.65			1.51	0.43		2.81				0.22	0.22	0.86	1.94	1.73		0.43	0.65			0.43	1.08	0.22		0.86	0.22	0.22			1.30	33.91	0.86	0.43	0.22			2.16	27.00
w	0.10		0.10		1.13		0.21		0.21		0.10		0.41		0.10	0.10	0.31	0.31		0.21		0.10		0.31	0.10		0.41	0.21	0.10	0.62		0.21			0.31	3.91	1.75	0.72	0.21		0.93			0.10	1.23	0.21		0.21		0.31	0.21		1.03	0.31	67.49	0.21	0.10			1.95	13.48
y	0.11				0.23		0.23				0.45		0.11		0.56	0.68	0.90	0.45				0.23		0.11	0.79		0.11	0.23	1.58	1.24		2.71		0.34	0.11	0.34		0.90	0.11	0.11	0.11				0.68			0.11	0.23	0.11			0.34	0.45	0.23	62.12	0.11			4.17	18.71
z		0.12			1.00						0.12				0.12	0.50	0.25	0.25		0.12		0.25			0.25				0.12	1.25		0.25			0.12	0.50		0.62			0.50			0.25		10.71		0.50	0.12		0.12		0.25	0.62			71.23			1.62	8.22
zh					3.51												1.75							1.75	1.75							1.75		3.51		1.75					5.26						5.26	1.75					3.51			3.51	14.04	29.82		3.51	17.54

B TIMIT Confusion Matrix

	aa	ae	ah	ao	aw	ax	ay	b	ch	cl	d	dh	dx	eh	el	en	epi	er	ey	f	g	hh	ih	ix	iy	jh	k	l	m	n	ng	ow	oy	p	r	s	sh	t	th	uh	uw	v	vcl	w	y	z	zh	SIL	INS	DEL
aa	67.67	3.01	4.51	6.77	3.01		3.76			0.75				1.50						0.75															0.75		0.75												1.50	5.26
ae	0.85	61.02				0.85	0.85							12.71	0.85			0.85	2.54					0.85			0.85																						11.02	6.78
ah	6.52	0.72	63.77	1.45	0.72	5.80	0.72							7.25				0.72					1.45	2.17						0.72		2.17								1.45									2.17	2.17
ao	7.55		1.89	65.09	1.89	0.94												1.89														3.77										0.94							8.49	7.55
aw	3.33	6.67			86.67																											3.33																		
ax	0.52	0.52	9.42	0.52		46.60								1.05	1.05		0.52	2.62		0.52			1.05	17.28				0.52				0.52				0.52				0.52			1.57	0.52					2.62	11.52
ay	7.78	1.11	1.11				80.00												2.22			1.11						1.11																					1.11	4.44
b								80.45		2.26	0.75						0.75			1.50										0.75				5.26					1.50				0.75	1.50					0.75	3.76
ch									82.93																	2.44									2.44	2.44		7.32											2.44	
cl			0.16							86.36			0.16	0.16			0.16		0.16	0.16				0.32	0.16						0.16				0.48			0.16			0.16	0.16	5.30						1.93	3.85
d				0.83				0.83			69.17	1.67	1.67				0.83				0.83			0.83			0.83											9.17							0.83				5.83	6.67
dh			0.76					0.76			3.05	61.83	2.29												0.76		0.76	1.53	0.76	3.05				0.76				1.53	6.11			2.29	2.29			1.53			2.29	7.63
dx											2.22		83.33														1.11			3.33					1.11								1.11							7.78
eh	0.52	5.21	5.73			2.08	1.04							62.50				1.04	1.56				9.38	2.08						0.52						0.52			0.52										1.56	5.73
el						1.69		1.69			1.69				61.02													16.95				10.17												1.69					3.39	1.69
en						6.90										48.28								3.45			3.45		3.45	27.59																				6.90
epi										1.59							36.51		1.59												1.59					1.59							4.76						22.22	30.16
er			0.43	0.43		0.85				0.43				0.43				79.91					0.43	3.85	0.43										9.83						0.43		0.85							1.71
ey		0.88						0.88						3.51					81.58	0.88			4.39		4.39					0.88																				2.63
f										0.75		0.75		0.75			1.50			83.46		0.75												1.50					3.01			4.51			0.75				1.50	0.75
g								3.03			1.52										83.33						1.52											3.03												7.58
hh		1.25								1.25												72.50					5.00			1.25				2.50									1.25		2.50				3.75	8.75
ih		0.49	0.97			1.46								4.37				0.49	1.46			0.49	55.83	18.93	4.37						0.49	0.49	0.49		0.49	0.49					1.46		0.49		0.49				1.46	4.85
ix						10.82					0.26			1.55		0.26		2.58	0.26				6.70	59.54	2.84			0.52		1.29		0.26			0.52	0.26				0.52	0.26	0.26	0.26						2.84	8.25
iy		0.39						0.39											1.56				1.95	4.28	80.16	0.39	0.39								0.39			0.39			0.39				0.78				5.45	3.11
jh									2.38		2.38										2.38					88.10									2.38											2.38				
k										1.18	1.18						0.59					0.59					85.88							0.59				1.76											2.94	5.29
l	0.85	0.42		0.42	1.69			0.42				0.85		0.42	1.27								0.42				0.42	74.15	0.85	0.42		0.85	0.42	0.85	0.42	0.42		0.85	0.85			0.42		0.85					0.85	10.59
m						0.97	0.49	0.49						0.49		0.49	0.49							0.49				0.97	75.73	8.25											0.97	0.49	0.97	0.49					0.97	7.28
n			0.28	1.11		0.28		0.28		0.56	0.28	0.56	1.39	0.28		1.11							0.56	0.28	0.56				2.22	77.22	3.06							0.28					0.28						1.11	8.33
ng																	1.92		3.85				1.92	1.92					1.92	19.23	61.54							1.92				1.92								3.85
ow			6.74	4.49	2.25	6.74				1.12				1.12	2.25			1.12										1.12				61.80									1.12	1.12								8.99
oy															6.25				6.25						6.25			6.25			6.25		68.75																	
p								5.84				0.73						0.73		2.19								0.73						78.83				2.92	1.46				0.73						0.73	5.11
r							0.71		0.36				0.36	0.36	0.36			3.20		0.36						0.36		0.36		0.36				0.36	77.58			1.07	0.36	0.36		0.36							3.91	9.25
s	0.31					0.31		0.31	1.23	0.93					0.31			0.31		0.31		0.31		0.62	0.31		0.31									87.96		0.93					0.31			1.85			0.93	2.47
sh									5.97	1.49																			1.49							5.97	83.58	1.49												
t									0.52	1.04	5.18	2.07				0.52				0.52				0.52			0.52									0.52		81.35								0.52			3.11	3.63
th						5.13				2.56	2.56	2.56								5.13														2.56		2.56		2.56	61.54				2.56						2.56	7.69
uh			6.90			13.79									3.45			3.45					10.34	10.34								3.45								37.93										10.34
uw						1.35				1.35				1.35				4.05	1.35				2.70	4.05	4.05										1.35					1.35	68.92		1.35							6.76
v					1.04	1.04	2.08	2.08			1.04	2.08	1.04							2.08	1.04								2.08			1.04			1.04							71.88	1.04						3.12	6.25
vcl									0.26	8.40	0.26	0.26	0.79	0.26	0.26		0.52			0.26			0.26	0.26							0.26											0.52	77.43						2.62	7.35
w				0.67													0.67		0.67									4.03				0.67										0.67		82.55					3.36	6.71
y													1.82									1.82			7.27					1.82	1.82				1.82		1.82	1.82							60.00				9.09	10.91
z										0.55		0.55											1.09	0.55	0.55				0.55							12.02			0.55							80.87	0.55		1.09	1.09
zh																									10.00												10.00								10.00		60.00			10.00