Speaker Recognition using Neural Networks and MFCC Features

Hamza Zamani1, Taishi Kato1, David Rosenwasser1

1University of California, Los Angeles

hwzamani@g.ucla.edu, taishikato10@g.ucla.edu, dbrosenwasser@gmail.com

Abstract

This paper implements mel-frequency cepstral coefficients (MFCC) as the feature vector to a neural network in order to develop a speaker recognition model. MFCCs are a commonly used speaker identification feature vector as they are modeled after the human auditory system. <need more here, less awkward transition> A data set consisting of 565 audio samples ranging from 3 to 5 seconds were using to form the model. Two sets of training data were trained against three different cases. The first set of training data consisted of various speakers reading the same sentence (read set) while the second set consisted of participants speaking in casual speech (phone set). The model will use the training data to form a model which will make predictions on the identity of the test data. The test data contains a read set, a phone set, and a mismatched set. By obtaining a high identification rate in all cases, the model would provide evidence to being a considerably robust speaker recognition system. The model obtains a <insert accuracy rates>

**Index Terms**: speech recognition, mel-frequency cepstral coefficients, neural networks, feature extraction

# Introduction

This section should talk generally about the MFCC, and speech processing stuff. Should include the challenges in speaker identification and our problem specifically. Should include some of the literature survey. Should talk generally about neural networks but not get into specifics about how the model is formed.

<steam of cosciousnes>: This section provides an overview of the speaker identification tasks and methodologies that a user can take in order to solve the problem. The challenge in speaker recognition resides in the nature of human speech. The short duration of the samples of speech used to form the model and the stylistic mismatches of the read vs. read speech. The identity information of a speaker is an entity of how the speech is spoken in contrast to what is being said. Speech therefore possesses a large degree of variability between speakers. Humans have an uncanny ability to recognize the subtleties in human speech to provide an extremely accurate method of speaker identification. Machines rely on mathematical models based off of studies of human biology to provide an estimate of speaker recognition. Systems to detect speakers must develop a model of the speaker in the training stage. The features are extracted, and buh bam. The mel-frequency cepstral coefficients used in this project aim to model human auditory perception. [speaker recognition by machines and humans]. First describe what the LPC is, and why formants are important for recognizing vowels/speech. Taking the LPC of a signal and then doing the cepstrum of that gives you the LPCC. Talk about what the cepstrum is and why its important. Originally thought of as deconvolution for echo cancellation purposes, the cepstrum is used to provide robust uniqueness quantifiers for a sampled speech signal. The cepstrum can be used to reliably distinguish between the source, and the transfer function. The transfer function is provides the unique features that may be used for speaker identification like vocal tract length. The cepstrum is initially calculated by taking the logarithm of the spectrum, and then performing the inverse fourier transform to enter the cepstral domain. May want to give some equations here or save it for later.

Mel-frequency cepstral coefficients function in the mel domain. The mel domain is a warped frequency domain meant to model how a human hears. It’s the DFT, then warped frequency domain, then cepstrum, then inverse. The complex exponential can be expressed in terms of cosines and sines, but we only take the cosiens, DCT. The deltas/derivatives capture the dynamics of each frame to each frame and their formants, which is useful for considering the uniqueness of the speech. Challenges: short duration. Stylistic mismatch. Auditory filters in the ear are narrow as the frequency is low. This relates to the critical band filters in the Rabiner book. P. 144 has the mel-frequency reation. P 464 has the MFCC eqn. Mel-scale imitates that. The mel domain allows for improved resolution at lower frequencies which comes into play during the mel-filter bank of the MFCC.

MFCCs implicitly capture speaker-specific information [a new set of features…] Cite voicebox.

# Methodology

This section describes the methodology forming the speaker identification system. This section should show what we actually do. The general block diagram is as follows:



## Preprocessing

Preprocessing should cover… sampling frequency? What other preprocessing steps do we take. If not many, this can be the feature extraction section.

### Headings

Section headings are centered in boldface with the first word capitalized and the rest of the heading in lower case. Sub- headings appear like major headings, except they start at the left margin in the column. Sub-sub-headings appear like sub-headings, except they are in italics and not boldface. See the examples in this file. No more than 3 levels of headings should be used.

## Feature Extraction

We extract the features as such: Speak to this figure



## Neural Network Model

This section will talk specifics about the NN.

All figures must be centered on the column (or page, if the figure spans both columns). Figure captions should follow each figure and have the format given in Figure 1.

Figures should be preferably line drawings. If they contain gray levels or colors, they should be checked to print well on a high-quality non-color laser printer.

Graphics (i. e., illustrations, figures) must not use stipple fill patterns because they will not reproduce properly in Adobe PDF. Please use only SOLID FILL COLORS.

Figures which span 2 columns (i. e., occupy full page width) must be placed at the top or bottom of the page.

## Tables

An example of a table is shown in Table 1. The caption text must be above the table.

Table 1: *This is an example of a table.*

|  |  |
| --- | --- |
| Ratio | Decibels |
| 1/1 | 0 |
| 2/1 | ≈ 6 |
| 3.16 | 10 |
| 1/10 | 20 |
| 10/1 | -20 |
| 100/1 | 40 |
| 1000/1 | 60 |

Figure 1: *Schematic diagram of speech production.*

## Equations

Equations should be placed on separate lines and numbered. Examples of equations are given below. Particularly,

 (1)

where  is a special warping function

 (2)

A residue theorem states that

 (3)

Applying (3) to (1), it is quite straightforward to see that

 (4)

Finally we have proven the secret theorem of all speech sciences. No more math is needed to show how useful the result is!

## Information for Word users only

For ease of formatting, please use the styles listed in Table 2. The styles are defined in this template file and are shown in the order in which they would be used when writing a paper.

When the heading styles in Table 2 are used, section numbers are no longer required to be typed in because they will be automatically numbered by Word. Similarly, reference items will be automatically numbered by Word when the “Reference” style is used.

Table 2: *Main predefined styles in Word.*

|  |  |
| --- | --- |
| Style Name | Entities in a Paper |
| Title | Title |
| Author | Author name |
| Affiliation | Author affiliation |
| Email | Email address |
| AbstractHeading | Abstract section heading |
| Body Text | First paragraph in abstract |
| Body Text Next | Following paragraphs in abstract |
| Index | Index terms |
| 1. Heading 1 | 1st level section heading |
| 1.1 Heading 2 | 2nd level section heading |
| 1.1.1 Heading 3 | 3rd level section heading |
| Body Text | First paragraph in section |
| Body Text Next | Following paragraphs in section |
| Figure Caption | Figure caption |
| Table Caption | Table caption |
| Equation | Equations |
| • List Bullet | Bulleted lists |
| [1] Reference | References |

If your Word document contains equations, you must not save your Word document from “.docx” to “.doc” because when doing so, Word will convert all equations to images of unacceptably low resolution.

## Hyperlinks

For technical reasons, the proceedings editor will strip all active links from the papers during processing. Hyperlinks can be included in your paper, if written in full, e. g. “http://www.foo.com/index.html”. The link text must be all black. Please make sure that they present no problems in printing to paper.

## Multimedia Files

The INTERSPEECH organizing committee offers the possibility to submit multimedia files. These files are meant for audio-visual illustrations that cannot be conveyed in text, tables and graphs. Just like you would when including graphics, make sure that you have sufficient author rights to the multimedia materials that you submit for publication. The proceedings media will NOT contain readers or players, so be sure to use widely accepted file formats, such as MPEG, Windows WAVE PCM (.wav) or Windows Media Video (.wmv) using standard codecs.

Your multimedia files must be submitted in a single ZIP file for each separate paper. Within the ZIP file you can use folders and filenames to help organize the multimedia files. In the ZIP file you should include a TEXT or HTML index file which describes the purpose and significance of each multimedia file. From within the manuscript, refer to a multimedia illustration by its filename. Use short file names without blanks for clarity.

The ZIP file you submit will be included as-is in the proceedings media and will be linked to your paper in the navigation interface of the proceedings. The publisher and the conference committee will not check or change the contents of your ZIP file.

Users of the proceedings who wish to access your multimedia files will click the link to the ZIP file which will then be opened by the operating system of their computer. Access to the contents of the ZIP file will be governed entirely by the operating system of the user's computer.

## Page Numbering

Page numbers will be added later to the document electro­nically. *Don't make any footers or headers*!

## References

The reference format is the standard IEEE one. References should be numbered in order of appearance, for example [1], [2], [3, pp. 417–422], and [4].

## Abstract

The total length of the abstract is limited to 200 words. The abstract included in your paper and the one you enter during web-based submission must be identical. Avoid non-ASCII characters or symbols as they may not display correctly in the abstract book.

## Author affiliation

Please list country names as part of the affiliation for each country.

## Number of authors in the author list

The maximum number of authors in the author list is twenty. If the number of contributing authors is more than twenty, they should be listed in a footnote or in acknowledgement section, as appropriate.

## Submitted files

Authors are requested to submit PDF files of their manu­scripts. You can use commercially available tools or for in­stance http://www.pdfforge.org/products/pdfcreator. The PDF file should comply with the following requirements: (a) there must be no PASSWORD protection on the PDF file at all; (b) all fonts must be embedded; and (c) the file must be text searchable (do CTRL-F and try to find a common word such as ‘the’). The proceedings editors will contact authors of non-complying files to obtain a replace­ment. In order not to endanger the preparation of the pro­ceedings, papers for which a replacement is not provided timely will be withdrawn.

# Discussion

This is the discussion. This is the discussion. This is the discus­sion. Is there any discussion?

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras consequat mollis odio, nec venenatis enim auctor sed. In- teger tincidunt fringilla lectus eget condimentum. In eget sapien id eros dapibus interdum vel ac quam. Aenean vitae rutrum erat. Aenean et risus pharetra, lacinia augue ut, fermentum ante. In- teger dui arcu, interdum at ornare a, faucibus quis est. Mauris quis quam felis. Etiam pulvinar massa et turpis lacinia, eu po- suere mi iaculis. Fusce at velit quis leo dignissim porttitor.

Fusce ut nunc eu sapien venenatis finibus a vel ligula. Pel- lentesque habitant morbi tristique senectus et netus et male- suada fames ac turpis egestas. Ut quam eros, volutpat at gravida consectetur, rutrum ut leo. Aenean cursus euismod feugiat. Cras hendrerit, ligula eu feugiat malesuada, neque turpis auc- tor lacus, sit amet accumsan neque orci a quam. Mauris suscipit ultrices mattis. Nulla at interdum metus, id pharetra diam. Cur- abitur at vestibulum sem, sed elementum massa. Donec iaculis et arcu ut rutrum. Fusce gravida, mauris porta volutpat eleifend, enim mauris eleifend orci, eu ultrices leo purus vitae metus. In pretium dolor ut magna dictum, at imperdiet lectus porta.

Quisque mollis lectus id risus pretium mattis. Morbi scelerisque posuere est, id efficitur urna luctus non. Praesent quam lacus, facilisis id ante eu, vehicula maximus ex. Nullam mollis in arcu vitae efficitur. Aliquam molestie eleifend ante, in pretium velit ultrices ac. Etiam laoreet nec sem non pulvinar.

Integer ligula felis, interdum non lacus id, malesuada imperdiet turpis.

Aenean sit amet volutpat nisi. Aliquam eu erat quis tortor ultrices laoreet. Vivamus fermentum semper metus, non fau- cibus libero euismod vitae. Sed efficitur porta congue. Aenean in faucibus nisi. Donec suscipit augue vitae orci consequat, sit amet aliquet felis varius. Duis efficitur lacinia dolor sit amet lobortis. Curabitur erat sapien, molestie nec nisi eu, dignissim accumsan ipsum. Fusce id nibh nec risus dictum posuere in ac magna. Donec malesuada massa sed erat lacinia cursus. Sus- pendisse ornare augue nec volutpat consequat.

Vestibulum et vulputate nisi, a malesuada mi. Nam pel- lentesque arcu sapien, at placerat odio imperdiet ut. Curabitur nec venenatis tellus, vel aliquet nisi. Curabitur vel ligula sit amet metus auctor pretium. Nullam nulla mi, blandit a mattis id, vulputate sit amet enim. Proin mollis fringilla dictum. Proin lacinia orci purus.

Curabitur porttitor bibendum dolor, nec consectetur sapien pulvinar id. Donec eleifend, est vel dignissim pretium, tortor augue euismod nunc, id fermentum erat felis ac neque. Morbi id lectus ultricies, rutrum justo eu, sollicitudin risus. Suspendisse lobortis efficitur nisi sit amet pellentesque. Ut eget augue at mi aliquet mattis. Proin et feugiat erat, sit amet sodales eros. In- teger sed elit quis est mattis ullamcorper. Pellentesque lectus nisi, vulputate a imperdiet tincidunt, auctor nec orci. Pellen- tesque sagittis nisl orci, vitae placerat massa lacinia nec. Sed egestas magna sed augue sollicitudin luctus. Praesent interdum bibendum tortor, eu porta purus. Aliquam convallis velit id mi fermentum, sed ornare eros cursus.

Quisque congue leo a fringilla pharetra. Phasellus sed tem- por est, sed auctor purus. Morbi vel lacus ullamcorper, auctor mauris id, pulvinar lorem. Suspendisse potenti. Nam porta, purus non eleifend bibendum, erat metus pellentesque elit, non luctus nibh nunc ornare nisl. Sed rutrum lacinia nisi ac sus- cipit. Curabitur non blandit augue. Integer viverra, ipsum vel molestie euismod, sem quam tempus massa, eget efficitur ante turpis non metus. Quisque efficitur posuere velit in iaculis. Cras imperdiet varius urna vitae vestibulum. Donec accumsan eget nisi sed pellentesque. Vestibulum id quam ut urna volutpat ul- lamcorper gravida sit amet libero. Aliquam bibendum, ligula vitae porta malesuada, arcu diam congue erat, a pharetra diam sem vulputate tortor. Etiam luctus iaculis leo cursus tristique.

Mauris mattis sem dolor, sit amet ullamcorper arcu tin- cidunt ac. Vestibulum at blandit tortor. Quisque bibendum congue leo, vitae eleifend massa. Vestibulum vitae odio elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut sagittis quam vel felis ornare, in gravida felis tempor. Donec molestie dui quis leo venenatis blandit. Nunc sit amet finibus metus. Cras ut lectus ex.

Suspendisse commodo libero vel leo tincidunt, a tempus mauris porta. Integer varius eros ac sapien lacinia vehicula. Donec porttitor, lacus faucibus rhoncus venenatis, neque quam imperdiet nunc, id consectetur metus purus quis sapien. Phasel- lus interdum nulla vel euismod posuere. Vestibulum finibus magna vel finibus mollis. Curabitur mollis turpis tortor, hen- drerit vulputate justo egestas quis. Nam dignissim luctus leo non elementum. Phasellus a metus at leo malesuada bibendum. Mauris quis eleifend magna, nec vehicula ex. Donec venenatis urna fermentum commodo vehicula. Ut mattis scelerisque ali- quam. Vivamus pulvinar erat metus, id tempus mi vulputate quis.

Fusce lobortis a urna eget blandit. Vivamus in eleifend neque, at sollicitudin lectus. Quisque faucibus egestas lorem, in commodo diam maximus eu. Morbi finibus ante ac felis port-

titor euismod. Donec lobortis aliquam ipsum sit amet luctus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam rutrum neque sapien, eget luctus turpis iaculis pulvinar. Duis quis pulvinar nunc, nec bibendum ligula. Phasellus suscipit sagittis lacus molestie laoreet. Pellen- tesque lacus diam, tincidunt a aliquam vitae, aliquet non justo.

Etiam lectus lacus, commodo eget consectetur eget, auc- tor vitae leo. Praesent vitae erat in diam blandit semper vitae et eros. Maecenas auctor pharetra nibh eget egestas. Donec accumsan ut risus eget rhoncus. Nam placerat, erat sit amet gravida mollis, purus arcu accumsan diam, tempus pharetra risus mi ac sapien. Ut et tortor porta, pulvinar elit vitae, tem- por mi. Nam interdum, nisl non pharetra molestie, turpis neque commodo ligula, sit amet pretium nisl nibh quis ante. Quisque et ex eget velit lobortis suscipit. Integer aliquam finibus mo- lestie. Sed pellentesque neque eu turpis aliquet, mattis ornare enim finibus. In hac habitasse platea dictumst.

Integer congue quis justo a posuere. Quisque porta, ante et dignissim suscipit, arcu mauris ultrices libero, nec sollicitudin purus lacus a enim. Aliquam feugiat eget lacus molestie so- dales. Duis blandit placerat nunc, et venenatis turpis dictum vel. Nulla facilisi. Nam ullamcorper, tellus eu posuere mattis, arcu lacus dictum nulla, vel mattis nisi sem posuere tellus. Etiam quis eros condimentum lectus lobortis eleifend. In ex lacus, so- dales scelerisque egestas ac, aliquam nec purus. Nunc sit amet magna non libero ullamcorper dictum. Phasellus porta faucibus tempus. Praesent blandit tortor sed tellus ornare consectetur. Sed sed nisi id neque porta varius eu eu velit. Curabitur var- ius convallis justo id facilisis. Mauris auctor velit nec aliquam cursus.

Integer suscipit scelerisque leo sed faucibus. Ut commodo nulla luctus diam posuere egestas. Integer ut augue ac velit ul- lamcorper tempus. Pellentesque in mi rhoncus, sodales sem quis, commodo sem. Aenean dapibus euismod diam id rhon- cus. Nullam vehicula placerat eros consectetur luctus. Aliquam auctor ipsum vitae egestas imperdiet. Ut nulla lacus, imperdiet quis urna vel, ornare imperdiet tortor. Mauris nec diam ac nunc laoreet volutpat at id turpis. Nulla eu neque a risus feugiat iac- ulis ac vel risus. Ut tempus elementum lorem eget porta. Nul- lam et ullamcorper urna. Cum sociis natoque penatibus et mag- nis dis parturient montes, nascetur ridiculus mus. Phasellus eget dui vitae nulla hendrerit ultrices quis rutrum leo.

Proin consectetur lacus sit amet eleifend varius. Etiam eu blandit risus. Curabitur pellentesque urna sed dolor congue mattis. Vestibulum ut velit posuere, feugiat leo a, dignissim massa. Proin eu nulla risus. Fusce luctus bibendum est, sit amet venenatis nisi finibus non. Donec ultricies ornare nunc at lobortis. Nam a auctor metus. Vestibulum pretium condimen- tum turpis ac mattis. Curabitur semper sagittis rhoncus. Duis molestie facilisis mattis. Sed pharetra lorem id tortor efficitur, sed maximus leo posuere. Quisque suscipit molestie convallis. Duis imperdiet placerat congue. Morbi placerat, velit ut tempor porta, ex nisi imperdiet purus, non feugiat ex velit nec nulla.

Phasellus mattis at erat eget lobortis. Vivamus sodales odio non erat luctus faucibus. Curabitur aliquam luctus nulla quis consectetur. Fusce vulputate finibus vulputate. Cras at condimentum massa. Duis vestibulum ipsum ac tortor lobortis fringilla. Cras non neque at nunc pellentesque mollis. Sed quis erat mauris. Ut dapibus sem lectus, quis imperdiet diam biben- dum et. Maecenas quis venenatis ante. Nunc blandit a risus sed scelerisque. Praesent cursus est sit amet nisi tempus, quis plac- erat libero rutrum. Phasellus lacinia nisi quis consequat mollis. Phasellus sagittis aliquam lacus.

Sed dolor quam, posuere nec nunc eget, feugiat lobortis

ligula. Fusce lacinia fermentum dolor, luctus dapibus ex ve- nenatis feugiat. In hac habitasse platea dictumst. Nullam vitae ligula dignissim, interdum turpis quis, tincidunt metus. Nul- lam in nisl vitae mauris egestas porta. Nam fringilla aliquet sapien, non dapibus nunc sollicitudin id. Nunc hendrerit felis et vehicula consequat. Praesent varius libero id volutpat iaculis. Aliquam vel dui imperdiet, pharetra augue sed, iaculis nulla. Quisque mollis orci nec odio eleifend, eget laoreet nunc feu- giat. Cras varius tortor a fringilla gravida.

Sed posuere erat eu dolor consequat euismod. Donec im- perdiet, tellus nec convallis commodo, lorem sem lobortis pu- rus, a lacinia massa ipsum vitae nisl. Quisque mollis orci nec odio eleifend, eget laoreet nunc feu- giat. Cras varius tortor a fringilla gravida. Vivamus auctor tellus in urna iaculis luctus. In dui nibh, posuere a erat a, lobortis finibus nulla. Sed vel suscipit nisi. Nunc eget nibh risus.

# Conclusions

Authors must proof read their PDF file prior to submission to ensure it is correct. Authors should not rely on proof­reading the Word file. Please proofread the PDF file before it is submitted.

# Acknowledgements

The ISCA Board would like to thank the organizing committees of the past INTERSPEECH conferences for their help and for kindly providing the template files.

Note to authors: Authors should not use logos in acknowledgement section; rather authors should acknowledge corporations by naming them only.

# References

1. S. B. Davis and P. Mermelstein, “Comparison of parametric representation for monosyllabic word recognition in con­ti­nuous­ly spoken sentences,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.
2. L. R. Rabiner, “A tutorial on hidden Markov models and selec­ted applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
3. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statis­tical Learning – Data Mining, Inference, and Prediction*. New York: Springer, 2009.
4. F. Lastname1, F. Lastname2, and F. Lastname3, “Title of your INTERSPEECH 2019 publication,” in *INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association, September 15-19, Graz, Austria, Proceedings*, 2019, pp. 100–104