Automatic Language Identification (LiD) Through Machine Learning

Alexander W. Hosford

Candidate No. 55095

Supervisor: Dr. Nicholas Collins

Music Informatics BSc

School of Informatics

University of Sussex

May 2011

Abstract

The following report outlines the creation of a system for the automatic identification of a given language from a selection of possible world languages. It is driven by a number of feature extraction techniques and an artificial neural net, and is written in the SuperCollider language utilising the WEKA machine learning tools. The highest level of accuracy obtained was a rate of 89.01% accuracy in discrimination between twelve languages. The system works on acoustical features alone and does not utilise any statistical models.

Contents

Acknowledgements

The author is indebted to the assistance of many individuals without whom the completion of his studies would not have been possible. Nicholas Collins, whose course was so varied and thought-provoking, possesses an insight into innumerable areas of expertise and has never failed to provide invaluable information; I also thank him for the authoring and use of the SuperCollider Music Information Retrieval Library. Kerith Harris, an ear above ears without whom I would sincerely not have completed my degree. Melanie Griffiths has provided much-needed emotional support in trying times, for your patience and understanding you are thanked. Finally his parents, who have never ceased to believe and tolerated so much noise.

Chapter One

Introduction

Language Identification (LiD) is concerned with the identification of a spoken language, uttered by an anonymous speaker using a given speech signal (Adda-Decker 2008, Muthusamy 1994, Navrátil 2006).

Human beings are most capable discriminators of spoken language; when presented with even short excerpts of speech a person is able to estimate which language they may be auditioning. This ability appears to develop during the earlier stages of infancy, as although newborn babies are capable of the perception and production of an incredibly wide range of sounds, prosodic contours specific to their mother tongue are some of the first linguistic skills to be acquired (Levitt 1991, Hallé 1991, cited in Adda-Decker p.8). Throughout the first months of a child's life and towards the end of the first year, an infant also shows native language abilities in terms of consonants and vowels used, and syllabic structure (McNeill 1970, Boysson-Bardies 1991).

Although a given individual may be able to take cues from language-specific phonemes or words of which they may have previous knowledge, this ability also extends to those languages with which the listener is unfamiliar and one may be relatively successful in judging which language is being presented to them.

LiD amongst humans is a desirable skill with many practical applications and the automation of this task is attractive in a world of increasing communication and multicultural exchange. There are currently 6,909 unique languages in the world (Ethnologue, 2010) however only 6% of these languages are spoken by 94% of the world's population of 6.4 billion people. Furthermore, only 5-10% of languages possess a corresponding writing system (Adda-Decker, p.6). Given these discrepancies it would be advantageous if artificial systems possessing the capacity to store models of all known languages (and discriminate between them) carry out the task of LiD automatically.

Employment of LiD methods is beneficial to numerous areas, notably that as a front-end extension to existing Automatic Speech Recognition systems. Global call-centres would benefit, so that callers may be directed to speakers of their native language, a task currently carried out by human beings. International environments such as airports would also find use; the ability to provide more meaningful customer service and overall LiD methods can be envisaged as the initial stage of future universal translation models. The development of LiD systems also allows for greater analysis of spoken natural language and may give an insight into differences between dialects and aid in linguistic research.

I detail the exploration of a number of techniques to successfully identify a presented natural language from a number of possible choices exclusively utilising the acoustical features contained therein. Features are extracted from a selection of speech samples and machine learning algorithms trained to discriminate between them.

The context within applicable fields in which the project was engaged is covered in Chapter Two. In Chapter Three I give attention to the professional considerations of which I had to be aware whilst undertaking the project. Chapter Four gives an overview of the system architecture and the necessary steps taken in order to reproduce the results seen, with a more detailed inspection of the individual system components taking place in Chapter Five. Recorded results of the system are shown in Chapter Six and I draw conclusions from the project, including what has been and what still could be achieved, in Chapter Seven.

Chapter Two

Area Review

This research project covers three main fields, those of linguistics, information retrieval and machine learning. I present an evaluation of these domains so that my chosen methods for this investigation are better understood.

2.1 Languages and their Relationships

The LiD task described herein aims toward the successful discrimination between twelve languages belonging to a number of language families. Three Germanic languages were selected (English, German and Dutch) alongside three members of the Romantic family (French, Italian and Spanish), three of the Slavic languages (Czech, Polish and Russian) and finally three languages of Eastern Asian origin, Mandarin, Korean and Japanese.

With the exception of Mandarin, Korean and Japanese, all of the above languages are members of the Indo-European language tree shown in Figure 1 (Ramat 1998). It is this family that possesses the greatest number of speakers globally and within, relationships between languages have been intensely studied in an attempt to trace their origins and development.

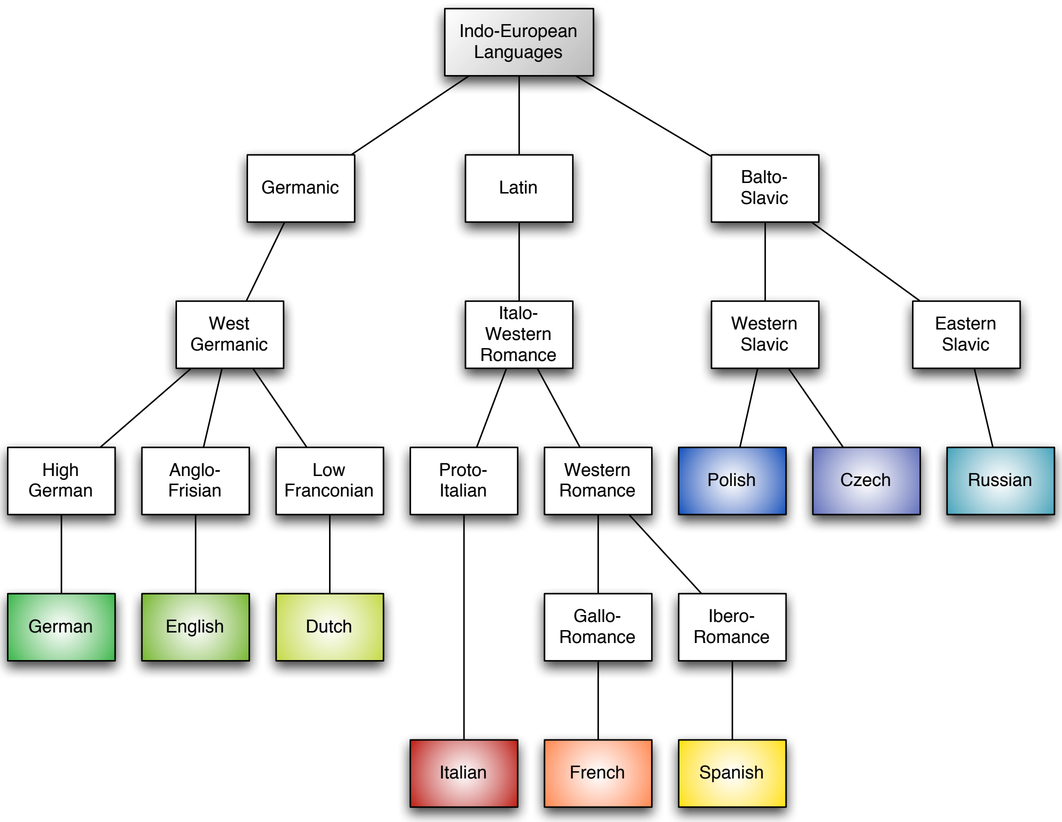


Figure 1 - The Indo-European language tree (Ramat 1998)

The three Asian languages within this project possess different origins, some of which are highly disputed. Mandarin Chinese comes under a branch of Sino-Tibetan languages (Figure 2); with Korean traditionally considered an isolate language and Japanese inhabiting its own isolate language family, Japonic. Recently, it has been suggested that the latter two of these three languages are an extension of the Altaic group designated 'Macro-Altaic', shown in Figure 3. Greenberg (Greenberg, 2000) presents lexical evidence that Indo-European and Altaic languages share a common root in the form of 'Eurasiatic' languages, shown in Figure 4.

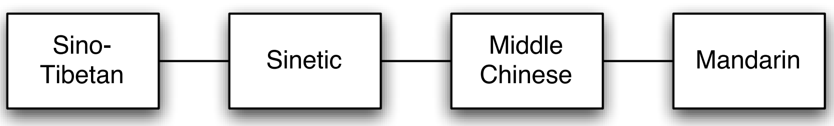


Figure 2 - The heritage of Mandarin Chinese (Coblin 2000)

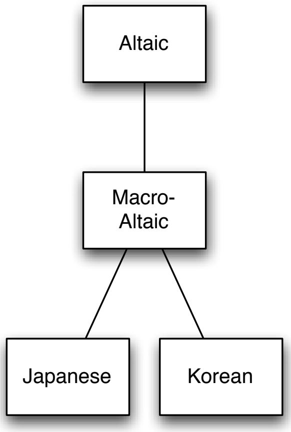


Figure 3 - Proposed membership of the Macro-Altaic language group for Japanese & Korean

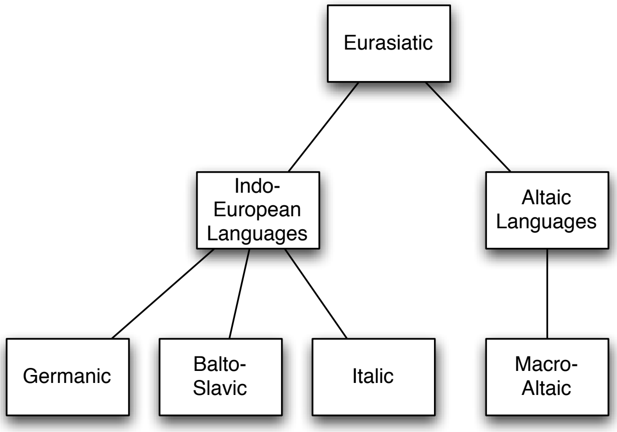


Figure 4 - Proposed 'Eurasiatic' language tree (Greenberg, 2000)

The relationships between Korean and Japanese are a subject of debate. Evidence has been presented that Japanese is a relative of 'Goguryeo', an ancient language that was spoken in the geographical area of North Korean until the 7th Century. Modern Korean is generally assumed to be derived from Silla, the language of the south-eastern state of the three kingdoms of ancient Korea that co-existed with Goguryeo. An alternative, however not exclusive hypothesis, is that Japanese is directly linked to Korean through lexical similarities (Beckwith, 2004). These relationships are shown in Figure 5.

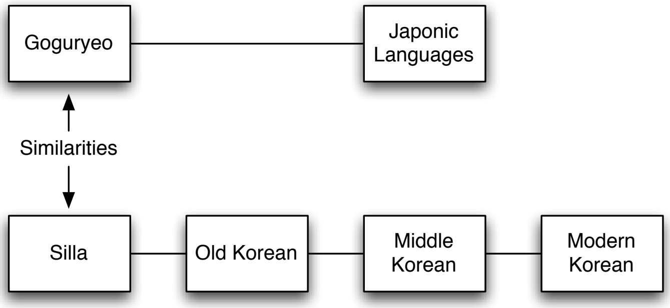


Figure 5 - Proposed relationships between Korean & Japanese (Beckwith, 2004)

The complexity of the task is expected to increase as the number of languages to be compared increases also. I would expect to see a greater confusion between those languages that are closely related such as Dutch and German, than languages between which a greater distance exists, for example English and Japanese. Of interest will be results of the system when attempting to discriminate between Korean and Japanese, given the current academic ambiguity over their relationship, similarities and origins.

2.2 Differences between Languages

Within spoken natural languages, many forms of information exist that make one discernable form the other.

The most important of these cues include (Matějka 2004);

- Phonemes – A limited set of recurring, distinctive speech sounds. A phoneme is the smallest unit that can be used to differentiate between speech signals. One phoneme may be more frequent in one language than another.
- Prosody – The characteristic rhythm, stress, and intonation of a given speech signal. This includes length phoneme length and pitch (f_0) contour.

In stress-based languages prosody is suprasegmental but its importance within phonemes is heightened when considering tonal languages such as Mandarin or Japanese.

- Phonotactics – The rules that govern the allowed sequence of phonemes in speech signals. For instance, a sequence of phonemes that is valid in one language may be illegal in another.
- Syntax – This deals with rules along the same line as phonotactics, however it relates to words and their admissible sequencing.

As well as being the most capable recognisers of speech in regards to Automatic Speech Recognition, human beings are currently also the most capable identifiers of language. The human brain is most competent at pattern recognition and the LiD problem can be seen as an extension of this task. Given that humans can make reasonable estimates on the spoken language presented within a few seconds of having heard it, it is reasonable to assume that such identification capabilities are based on phonological information rather than larger linguistic constructs such as word content or phrasing (Muthusamy 1993). An individual may have an incomplete model of a given language in memory and how that language sounds.

For example, the nasal vowels of French are in contrast to the diphthongs of English that do not occur in the former; Some languages display greater tonality than others, such as Mandarin and Japanese, and the rhythm of others may be an important cue for languages such as Italian or Spanish. Humans may also make use of language-specific acoustic identifiers, such as the palatised consonants of Slavic languages and the ‘clicks’ apparent in many African dialects

(Adda-Decker p7). It is likely that within these models we possess not only one feature per language, but also a range of cues that upon audition contribute to a level of certainty over which language is currently being heard.

Navrátil performed perceptual tests using five foreign languages to demonstrate the importance of various cues in LiD (Navrátil 2001). Three sets of stimulus were used to assess the identification capabilities of a range of human listeners;

1. Original stimuli, unaltered speech signals
2. Extracted Syllables randomly sequenced and
3. Filtered stimuli that preserve only the f_0 contour of the speech signal.

Test	English	German	French	Mandarin	Japanese	Average
Original (3s)	100.00	98.7	98.7	88.7	81.7	93.6
Shuffled Syllables (6s)	98.7	79.7	79.1	57.7	54.6	73.9
f_0 Contour (6s)	34.3	34.3	69.4	65.9	45.3	49.4

Figure 6 - Language classification accuracy (%) using different speech stimuli (from Navrátil 2001)

The results of the experiment are displayed in Figure 6 and show that for those languages in which the listener has a high level of background knowledge, identification rates are high. When the syllables are concatenated in a random fashion the suprasegmental prosodic information is lost. However, this has a minimal impact on the mother tongue or familiar languages and suggests that for well-known languages an amount of pre-informed knowledge is utilised in the identification.

When only an f_0 contour remains, performance drops across the board; for example the rate of successful identification for English is behind that of French and Mandarin. These results show that successful LiD relies not only on one set of cues but takes from many – some of which may be fairly redundant when it comes to making a decision (Adda-Decker, p11).

Successful LiD is also dependent on the distance that exists between two compared languages. Discrimination between English and Japanese should theoretically be easier to attain than discrimination between two closer related languages such as Italian and Spanish or Dutch and Flemish. An effective automatic LiD system should make use of many of the above cues in order to provide successful discrimination.

2.3 Previous Approaches to Automatic Language Identification

The earliest research into the area of language discrimination is that of Leonard & Doddington between 1974 and 1980, funded by the United States Air Force and carried out under the supervision of Texas Instruments. Such early work was highly confidential due to the implications for national security and communications monitoring. It was not until 1977 that the first methodological work was carried out into the possibilities of the use of computers for the task of discrimination (House & Neuburg 1977). The processing power available to researchers during this time was not sufficient for the proposed problem and as such this research outlined theoretical discrimination of languages using Markov chains, working on statistical constraints observed during the sequence of phonemes within languages.

Phonetic data was generated manually from transcriptions and results showed that discrimination should be possible once adequate computing power was available; in fact their system showed perfect discrimination. However, it also assumed that separation between phonemes had been performed immaculately and that phonemes had been classified without error. When applied to real speech data (Li & Edwards 1980) these concepts were shown to be relatively effective at discrimination using broad phonemic categories, gaining 80% accuracy on five languages, however comparison with other efforts would not be meaningful as the study focused on male speakers only and the languages compared were never disclosed (Muthusamy 1993).

LiD systems that display the highest accuracy rates generally make use of phonotactic rules, approaching the individual modelling of each phoneme in a signal and assessing its location amongst others to give a picture of whether that sequence is admissible in a given language. Such approaches however require somewhat expert linguistic knowledge (Lin & Wang 2005) and the modelling and implementation of such rules is time-consuming and beyond the scope of the current project.

Although some success in identification has been shown using raw waveforms and recurrent neural networks (Kwasny *et.al* 1992, 1993) it is possible to use acoustical features of speech signals in order to make reasonable assumptions about the identity of a given language. Acoustical features were first explored in 1982 with pattern matching techniques applied to the obtained data and the overall accuracy achieved was 84%, with individual languages scoring between 76.8% for American English to 94% for Korean (Cimarusti & Ives 1982).

This study was however only tested on five separate speakers which somewhat negated the system's ability in speaker independence.

With regards to the above studies it is fairly problematic to be able to compare them to each other directly. This is due to the widely different nature of their approaches and issues generated by each of these methods. Matějka states that the term 'automatic' implies that '*the process is independent of content, task or vocabulary and robust with regard to speaker identity, sex, age as well as to noise and distortion introduced by the communication channel*' (Matějka 2004, p.112). It is not until recently that we see a renewed interest in the problem of LiD, partly due to the growing availability of suitable speech data such as the OGI Multi-Language Telephone Speech Corpus (OGLTS). In 1993 the National Institute of Standards and Technology (NIST) chose OGLTS as the standard for evaluating LiD systems (Muthusamy 1994).

The LiD problem can be seen as an extension of ASR research, as both traditionally make use of MFCC vectors and phoneme modelling. LiD has recently been expanded into the domains of dialect and accent discrimination within spoken languages (Chen *et.al,* 2001). Many previous methods rely for the most part on involved statistical models. This project shall attempt to make successful identification based on acoustical properties of the speech signal alone and not delve into more complicated statistical methods of discrimination such as phoneme modelling and phonotactics.

Chapter Three

Professional Considerations

As I am using speech samples I must be mindful of copyright considerations. I have however restricted my corpus to free online radio podcasts, and as such believe I am utilising the speech samples I have obtained according to fair academic use.

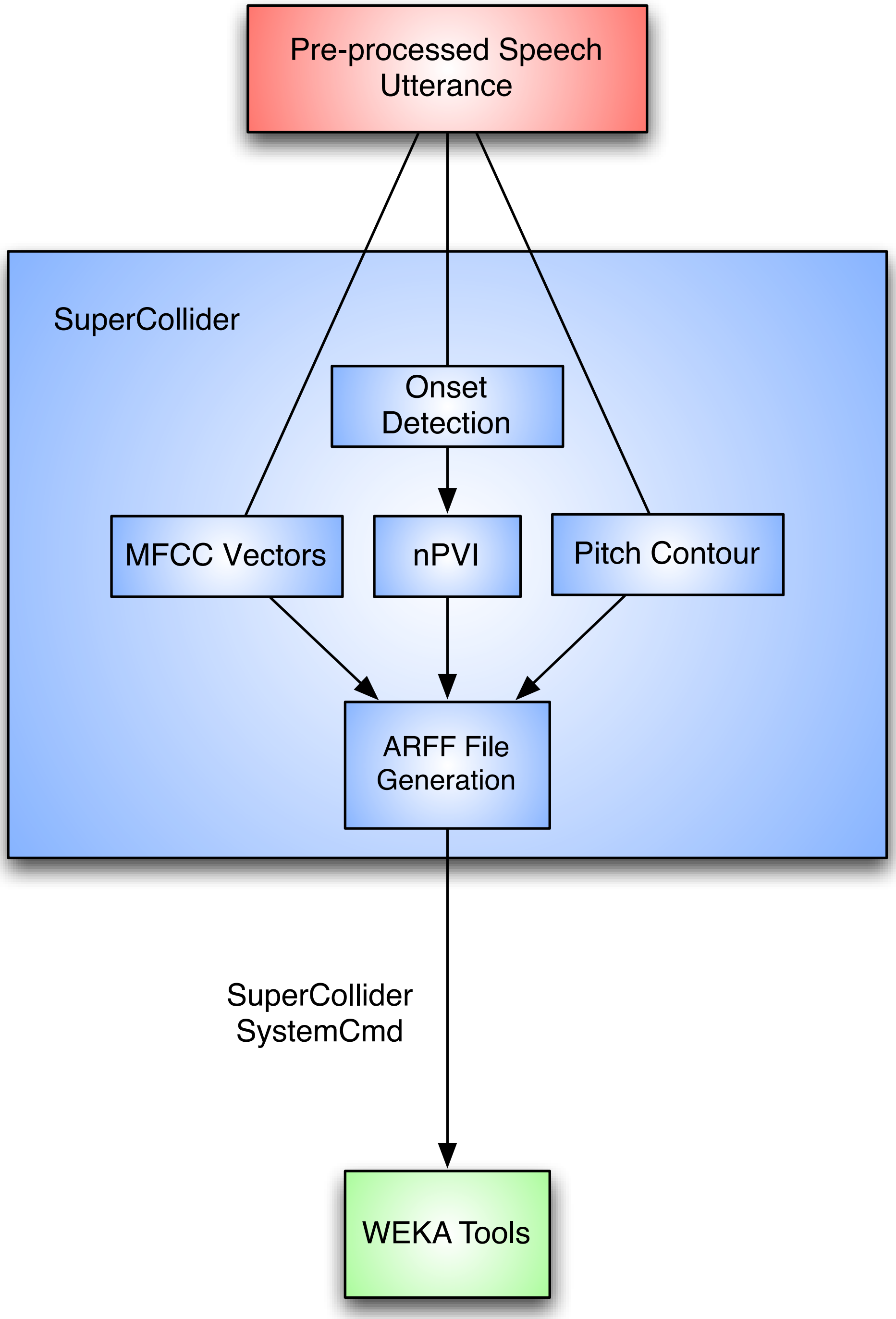
In accordance with Section 6 of the Code of Conduct for the British Computer Society 2006, *"You shall carry out work or study with due care and diligence in accordance with the relevant authority's requirements, and the interests of system users. If your professional judgment is overruled, you shall indicate the likely risks and consequences,"* I will therefore ensure that I conduct my studies according to the procedures laid out by the University of Sussex and the School of Informatics.

In accordance with Section 14, *"You shall seek to upgrade your professional knowledge and skill, and shall maintain awareness of technological developments, procedures and standards which are relevant to your field, and encourage your subordinates to do likewise,"* I intend to broaden my knowledge on the investigated subjects to the best of my ability.

In accordance with Section 15, "*You shall not claim any level of competence that you do not possess. You shall only offer to do work or provide a service that is within your professional competence,"* I acknowledge that I shall not plagiarize either research or code and that any research will be properly referenced in my work and that third-party code used in the completion of my project is stated as such and not passed off as my own.

Chapter Four

System Overview



Chapter Five

Implementation

The following chapter details steps taken in order to generate data from speech signals for further investigation. For feature extraction, the system utilises the SuperCollider Music Information Retrieval Library (SCMIR) authored by Nicholas Collins (Collins 2010a). This library allows for the extraction of meaningful properties from audio signals such as MFCCs, chromagrams, loudness plots and others so they may be used for further processing and investigation.

5.1 Pre-processing of Speech

The speech data that is used for the project is obtained from online radio podcasts, downloaded from the iTunes© library. Radio broadcasts are a good example of the type of input that an LiD system may be required to work on; it is semi-spontaneous and also very colloquial in nature. The real world applications of LiD systems suggest that it is unrealistic that well-organised databases such as academic linguistic corpora would ever be examined.

Audio files were converted to 16-bit, stereo AIFF files with a sample rate of 44,100 Hz and split into smaller utterances between the lengths of twenty and thirty seconds. Where possible, phrasing began and ended at the onset and termination of sentences.

For each language ten utterances were utilised, split equally between gender, in the hope that any acoustical differences arising from the differences between male and female voicing were accounted for in the training process. It was attempted to obtain samples of at least three different speakers where possible in the training data, so that data would not overfit in the case of one speaker being particularly typical of that language's features. These steps were taken in order to satisfy Matějka's criteria for autonomy in terms of independence regarding speaker independence and autonomy (Matějka 2004).

5.2 MFCC Vectors

Due to their ability to show the amplitude spectrum of an audio signal in a compact form, Mel Frequency Cepstral Coefficients (MFCCs) have long been one of the most prominent features in ASR and LiD (Logan 2000, Ganchev *et.al* 2005, Jurafsky 2009 p.329). Figure 7 shows the process of MFCC generation from a given audio signal. Roads (Roads, 1996 p.516) describes the cepstrum as a way of separating a strongly pitched component from the rest of the spectral data. For speech, cepstral analysis can be seen to separate two features, the glottal pulse excitation of the vocal chord and the vocal tract resonances.

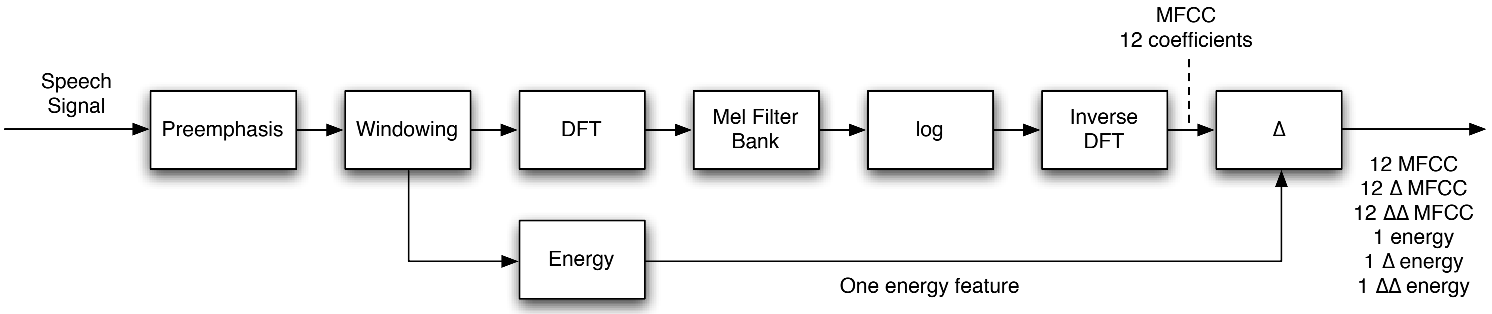


Figure 7 - The MFCC generation process (Jurafsky 2009)

Jurafsky notes that speech is a non-stationary signal and as such a function must occur in order to make the statistical properties of the utterance static. Therefore, a windowing function is applied to break the utterance into segments for further processing. It is preferred that a non-rectangular window is applied; such windowing functions can cause problems when abruptly cutting off signals at their boundaries. The most common window used in the extraction of MFCCs is the Hamming window (Figure 8a). The FFT function in the SCMIR library makes use of the Hann window (Figure 8b) as default, however, these are functionally and perceptually similar and the effect this has on results is negligible.

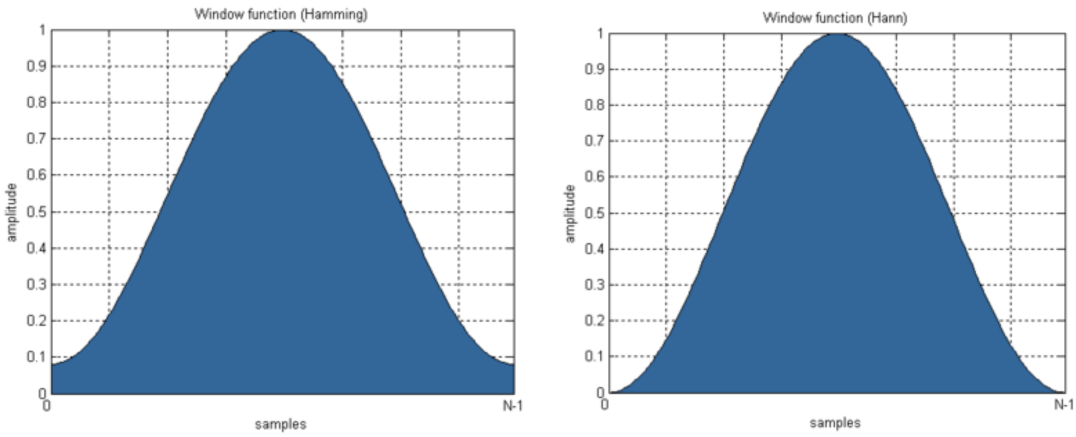


Figure 8(a) & 8(b) - The Hamming and the Hann windowing functions (Wikipedia)

In order to calculate the amount of energy that the signal contains in each frequency band, the windowed signal *x[n]…x[m]* is used as the input of a Discrete Fourier Transform (DFT) and the output, for *N* frequency bands is a complex number *X[k]* that represents both the magnitude and phase of that frequency component in the original signal.

The results of the FFT will show information concerning the amount of energy at each frequency band. However psychoacoustically, human beings are not equally sensitive to all frequencies and display a logarithmic sensitivity, which is reduced above circa 1kHz.

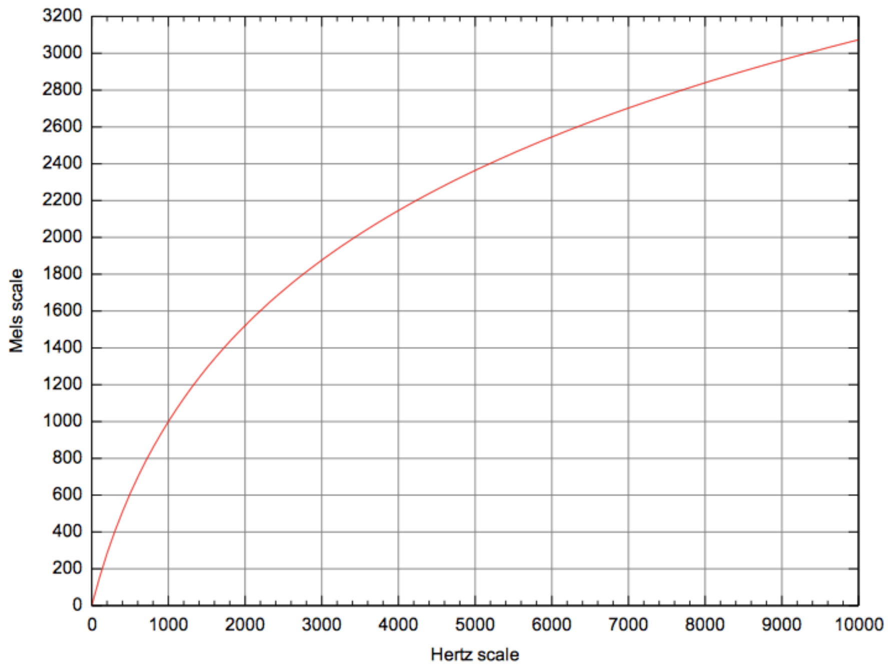


Figure 9 - The Mel Scale (Wikipedia)

It has been shown that modelling this property of human hearing (Jurafsky p.332) improves the performance of speech recognition algorithms and therefore the output of the FFT is wrapped onto the Mel scale. A Mel is a unit of frequency that is roughly linear below 1,000 Hz and logarithmic above (Figure 9). To compute the Mel frequency *m* from the raw acoustic frequency the following function is applied;

$$Mel(f) = 1127\ln\left(1 + \left(\frac{f}{700}\right)\right)$$

Figure 10- Calculating the Mel frequency from a given frequency (Jurafsky 2009)

5.3 Pitch Contours

Prosody between languages varies greatly and one prosodic cue that can be taken from speech signals is that of pitch contour. Figures 11 through 14 show a selection of pitch contours in German, Mandarin, Italian and Japanese that have been generated in the Praat© software. Mandarin and Japanese are considered tonal languages and this can be seen clearly in the pitch contour diagrams; tonal

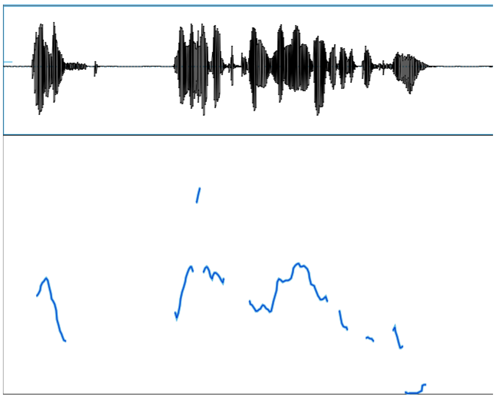


Figure 11 - German pitch contour

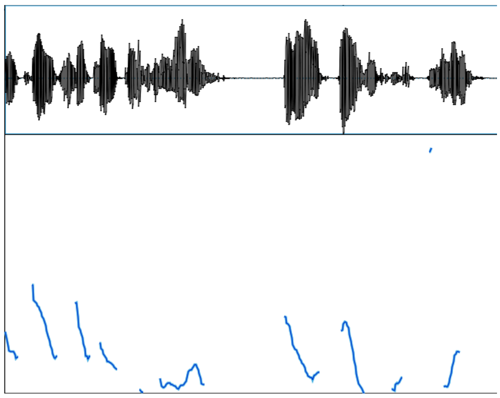


Figure 12 - Mandarin pitch contour

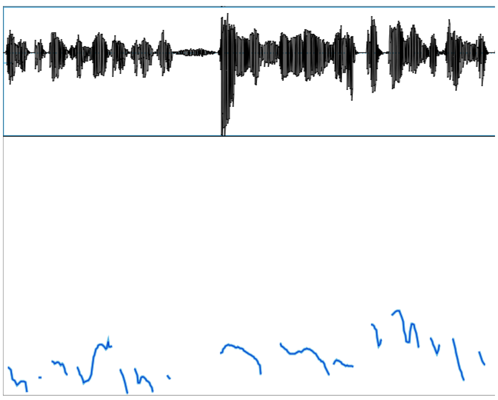


Figure 13 – Italian pitch contour

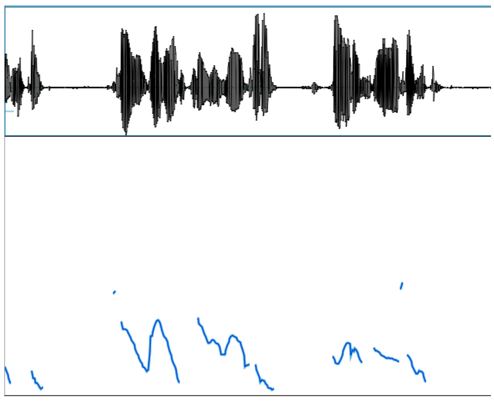


Figure 14 - Japanese pitch contour

changes tend to be tied to individual phonemes, which is in contrast to the contour displayed by German, whose tonal changes generally are greater within phonemes and vary far more. The pitch contour for Italian displays a characteristic of this language; that the pitch of phonemes tends to drop at the end of syllabic structures and rises at the beginning.

Traditionally, LiD systems that attempt to discriminate based solely on the information of the pitch contour have performed poorly in contrast to other methods. Recent attempts in using pitch contours to train Gaussian Mixture Models (Lin & Wang, 2005) and by coupling them with MFCC Vectors (Ezzaidi 2001) have proved more fruitful. Pitch contours are extracted by way of the 'Tartini' uGen within SuperCollider by the SCMIR library. The Tartini uGen is modelled on the 'McLeod Pitch Method' (McLeod & Wyvill, 2005) and generates two features; a fundamental frequency trail and a measure of confidence in the fundamental frequency of the pitch contour in the range of 0 to 1.

5.4 Speech Rhythm

Onsets are detected to gain a meaningful feature based on the acoustic-phonetic rhythm of the speech signals. The languages of the world differ in their rhythm and therefore cues for identification may be taken from these features (Ling *et.al* 2000, Farinas & Pellegrino 2001, Ramus 2002). Gibbon & Gut describe rhythm as '*the recurrence of a perceivable temporal patterning of strongly marked (focal)... and weakly marked (non-focal) values of some parameter... of a constant temporal domain'* (Gibbon & Gut 2001, p.95).

Languages traditionally have been categorised as either stress-timed, which refers to regularly occurring beats or stresses such as in British English and German, or syllable-timed which depends on regularly timed syllables such as in French. Recently, a third basis of timing has been proposed, that of the 'mora', a subsyllabic timing unit that occurs in Japanese and Estonian among others (Gibbon & Gut 2001, Grabe & Low 2002, Port *et.al* 1987).

Previous methods to assess the validity of temporal structure of speech signals have relied on the identification of individual phonemes and their corresponding syllables in order to classify a language. Durational differences can however be associated with vowels rather than syllables and likewise a raw measure of onsets may be used in order to calculate a single meaningful value for this purpose.

In order to capture the rhythmic properties of a language, a measure known as the 'Normalised Pairwise Variability Index' (nPVI) (Grabe & Low 2002) was used in order to generate a single number that could characterise the temporal features of a speech signal. This approach differs to previous methods in that rhythmic units are not treated phonologically, requiring the deconstruction of a phrase into identified phonemes; rather the duration between each acoustic event is used in order to calculate a value. Grabe & Low define the acoustic event as the onset of each vowel event; my implementation however uses the onset of a speech signal to the same effect.

$$PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

Figure 15 - The normalised pairwise variability index function (taken from Low & Grabe 2002)

The nPVI calculation is shown in Figure 15. Onsets are detected and extracted from the signal; the difference in duration between each pair of sequential measurements is then calculated. The absolute value of the difference is then divided by the mean duration of the pair. Finally the index is normalised by summing all of the differences and dividing this value by the number of durations that were observed. In Low & Grabe's implementation, the nPVI produces fractional values and is for this reason multiplied by 100. To avoid any issues with weighting this final step was omitted from my function as the WEKA machine learning tools function best when presented with values between 0 & 1.

5.5 ARFF File Generation

The Attribute-Relation file format (ARFF) is a dataset and does not specify which of the data is to be classified (Witten & Frank 2005). This allows for numerous machine-learning approaches from the same file and to compare the usefulness of any given data type. All feature data that is extracted is made available for further investigation in this format. This is the acceptable structure for the WEKA machine learning tools and the SCMIR library includes a function that writes out to the ARFF format.

This function, however, was not adequate for the requirements of the project and as such it was required that this be modified in order to correctly output all feature types from a speech utterance in the correct format. This function was placed in the main body of the feature extraction files and called for each bank of languages. The function takes in as parameters an SCMIR audio file and writes to the file that has been previously opened in the patch. For the comparison of language pairs, a SuperCollider patch exists that automatically generates the ARFF file for each of the 66 pairs, over 8 feature sets that contain an increasing amount of information. For the comparison of language families and for the final experiments in which all languages are compared, individual patches were created for the creation of the ARFF files necessary for the machine learning stage of the project.

Feature comparisons carried out on the language combinations are depicted in Figure 16. All language combinations are tested on with a minimal amount of features, 4 MFCC vectors, a medium amount, 13 MFCC vectors and a pitch

Feature Sets / Comparison Type	4 MFCC	4 MFCC Tartini	4 MFCC nPVI	4 MFCC Tartini nPVI	13 MFCC	13 MFCC Tartini	13 MFCC nPVI	41 MFCC	41 MFCC Tartini	41 MFCC Tartini nPVI
Language Pair	✔	✔	✔		✔	✔		✔	✔	✔
Language Family	✔					✔			✔	✔
All Languages	✔	✔		✔	✔	✔	✔	✔	✔	✔

Figure 16 - Feature Sets extracted by comparison type

contour, and a larger amount of features, 41 MFCC Vectors, a pitch contour and a measure of the nPVI. Within language pair comparisons, a pitch contour and an nPVI measure augment the minimum amount of features in separate experiments to assess their advantages.

5.6 The WEKA Environment

The machine learning aspect of my project was made possible through the use of the WEKA workbench, a collection of algorithms and visualisation tools used for data classification and machine learning. WEKA is accessible through a GUI environment that is useful for visually representing data in any given ARFF file it is provided with, however, it also allows for command line functionality. A SuperCollider patch was written that allows the chosen machine-learning algorithm to be specified along with its parameters and then recursively called on a selection of ARFF files in an attempt to classify the data therein.

Two algorithms were chosen for the classification of data – the Naïve Bayes algorithm and WEKA's built in Multilayer Perceptron (MLP). Naïve Bayes was chosen as a benchmark for the tests as it is a fast way to classify data and also seen as 'lazy'. As such this function is generally unsuitable for the task of LiD and results obtained from this run of tests would be able to confirm or deny the complexity of the problem (Witten & Frank 2005).

The MLP was chosen in reference to previous studies that had utilised artificial neural networks and has the following parameters; a learning rate of 0.3 for updating the weights, a momentum of 0.2 and 2000 epochs to be run. The number of hidden nodes utilised was dependent on the number of features generated by feature extraction; this was the sum total of the amount of attributes (features) and the amount of classes (languages to be compared) within each ARFF file. For example, a comparison of all twelve languages with 41 MFCC Vectors, a pitch contour (two features) and an nPVI measure would equate to fifty-six hidden nodes.

Chapter Six

Evaluation

The following chapter outlines a selection of results obtained from the system; full results are made available in Appendix II. For conducted experiments, features were extracted twice; once with values averaged over the entire length of the speech signal and secondly, features were gathered into 5 second segments as segmentation of the speech signal is common in many previous approaches to the LiD problem (Zissman 1996, Wu 2006, Muthusamy 1993).

Two tables are provided; one for averaged values and a second for segmentation. Note the difference in the number of instances between these two comparisons; this is due to the varying length of the speech signal used and ranges between 64 and 69 generated instances for each file. Within tables, the left column is the feature set extracted for each pair, the language columns display how many instances of that language were correctly classified and the right column shows the accuracy for the system on that particular feature set. A mean performance is shown for all feature sets extracted in the blue cell. Cells highlighted in yellow display where the system performed worse than chance.

For five-second segments nPVI results are omitted, as this was not functioning at the time of the report completion. Also shown are corresponding bar graphs; due to aforementioned differences in instance quantity this is displayed on a logarithmic scale. Finally a line graph is presented depicting mean performance of the system for each feature set, and a mean across all feature sets. For all charts red & blue lines depict the results for the averaged data and green

& purple show the five-second segmentation results. All values on the left-hand axis are percentages.

6.1 Language Pairs

Comparisons were made on the audio files for each of the twelve languages against each other, equating to sixty-six language pairs, depicted in Figure 17. It is with these comparisons that I am able to assess the true ability to discriminate, as by choosing which languages to compare the distance between languages can be controlled and the system's response to a wide range of language similarities evaluated.

English, German	English, Dutch	English, French
English, Polish	English, Russian	English, Mandarin
German, French	German, Spanish	German, Italian
German, Mandarin	German, Korean	German, Japanese
Dutch, Czech	Dutch, Polish	Dutch, Russian
French, Spanish	French, Italian	French, Czech
French, Korean	French, Japanese	Spanish, Italian
Spanish, Mandarin	Spanish, Korean	Spanish, Japanese
Italian, Mandarin	Italian, Korean	Italian, Japanese
Czech, Korean	Czech, Japanese	Polish, Russian
Russian, Mandarin	Russian, Korean	Russian, Japanese
English, Spanish	English, Italian	English, Czech
English, Korean	English, Japanese	German, Dutch
German, Czech	German, Polish	German, Russian
Dutch, French	Dutch, Spanish	Dutch, Italian
Dutch, Mandarin	Dutch, Korean	Dutch, Japanese
French, Polish	French, Russian	French, Mandarin
Spanish, Czech	Spanish, Polish	Spanish, Russian
Italian, Czech	Italian, Polish	Italian, Russian
Czech, Polish	Czech, Russian	Czech, Mandarin
Polish, Mandarin	Polish, Korean	Polish, Japanese
Mandarin, Korean	Mandarin, Japanese	Korean, Japanese

Figure 17 - Feature Sets extracted by comparison type

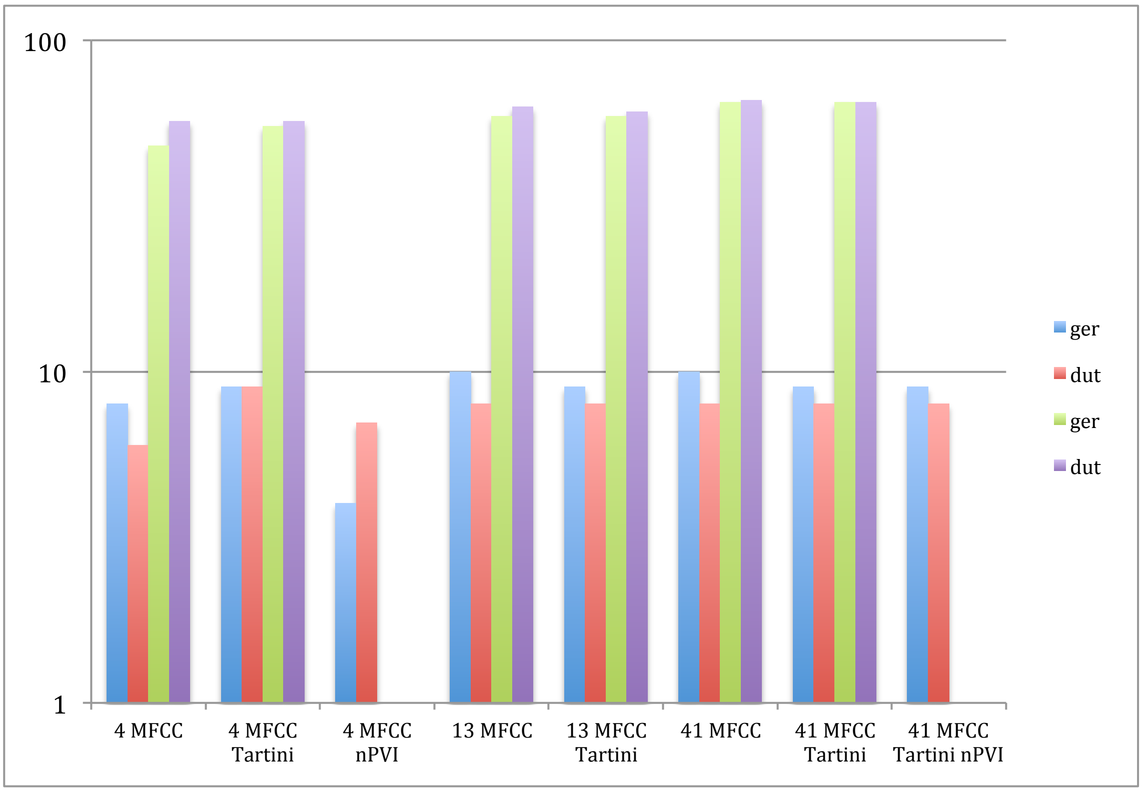
Presented first are three sets of results from language pairs whose distance is short – that is they are not far removed from each other on their respective language trees. After, I present a further three sets from language pairs whose distance is much greater. Where phonetic similarities between languages are described, '/ /' denotes the corresponding symbol in the International Phonetic Alphabet.

1) German & Dutch

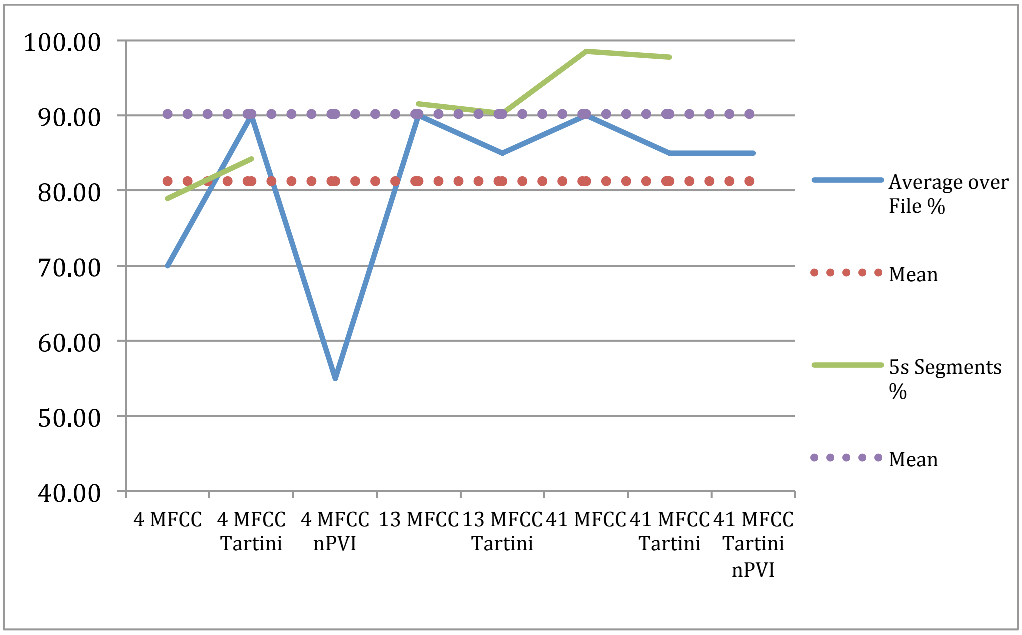
Averaged	Ger	Dut	%
4 MFCC	8	6	70
4 MFCC Tartini	9	9	90
4 MFCC nPVI	4	7	55
13 MFCC	10	8	90
13 MFCC Tartini	9	8	85
41 MFCC	10	8	90
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	9	8	85
			81.25

Instances ➤	66	67	
5s Segments	Ger	Dut	%
4 MFCC	48	57	78.95
4 MFCC Tartini	55	57	84.21
4 MFCC nPVI			
13 MFCC	59	63	91.50
13 MFCC Tartini	59	61	90.23
41 MFCC	65	66	98.50
41 MFCC Tartini	65	65	97.74
41 MFCC Tartini nPVI			
			90.19

German and Dutch are separated by a degree of about thirteen to eighteen hundred years however they are very homogenous in terms of grammar and phonetics. Both are stress timed which may account for accuracy drops when introducing the nPVI measure into the system. When utilising a low number of MFCCs (4) averaging across the entire file, the addition of a pitch contour improved the performance however this increase was not observed when a



greater number of spectral bins were used. When the signals were segmented, accuracy improves by approximately 9% with a small drop corresponding with the addition of a pitch contour at 13 MFCC vectors.



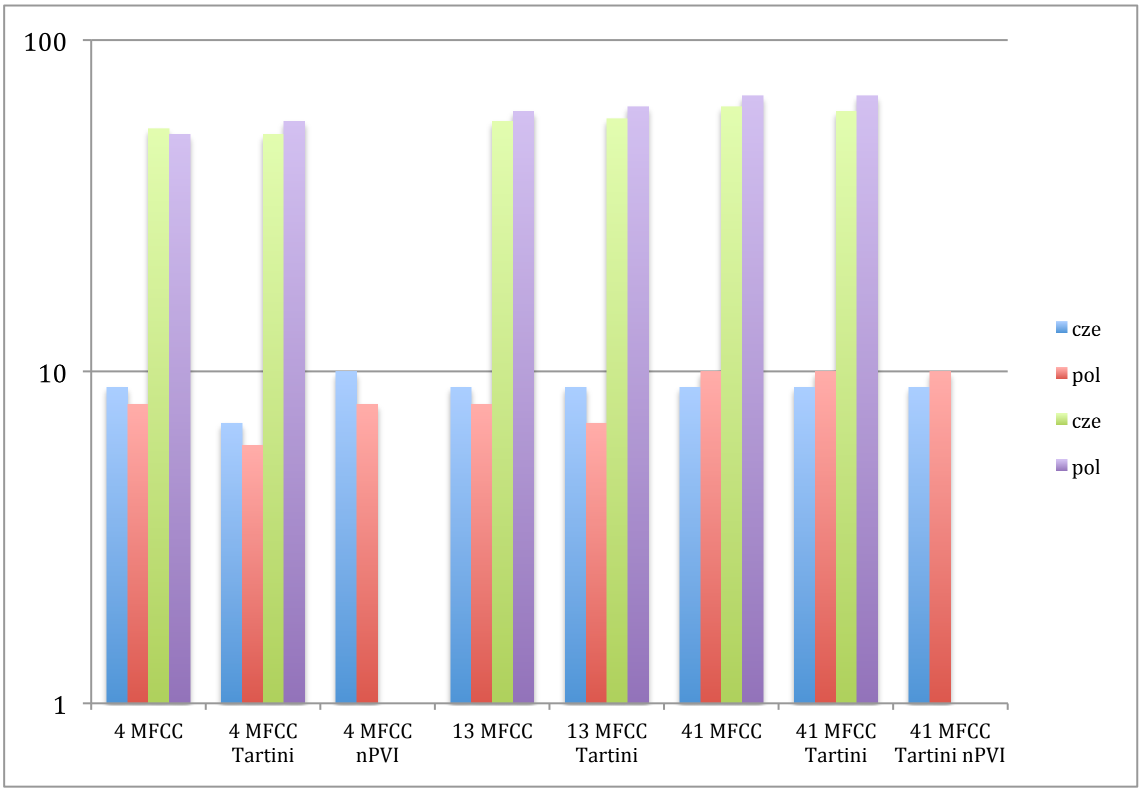
German and Dutch both make great use of velar fricatives, such as the 'ch', /x/, in 'Dach' (roof, German) and in 'goed' (good, Dutch); also voiceless retroflex fricatives such as 'sch', /ʂ/, in the German 'Schade' (shame). Such sounds produce a large amount of noise energy across the entire spectrum and it is possible that when comparing a greater number of spectral bins that this distributed energy across the bins is confusing the system slightly under the averaged comparison.

2) Czech & Polish

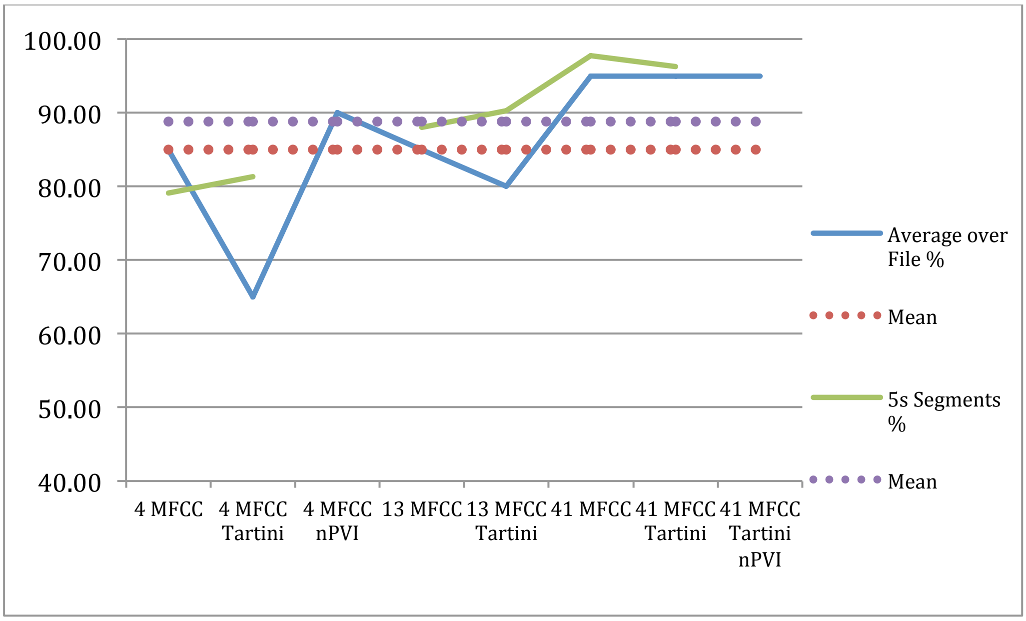
Averaged	Cze	Pol	%
4 MFCC	9	8	85
4 MFCC Tartini	7	6	65
4 MFCC nPVI	10	8	90
13 MFCC	9	8	85
13 MFCC Tartini	9	7	80
41 MFCC	9	10	95
41 MFCC Tartini	9	10	95
41 MFCC Tartini nPVI	9	10	95
			86.25

Instances ➤	65	69	
5s Segments	Cze	Pol	%
4 MFCC	54	52	79.10
4 MFCC Tartini	52	57	81.34
4 MFCC nPVI			
13 MFCC	57	61	88.06
13 MFCC Tartini	58	63	90.30
41 MFCC	63	68	97.76
41 MFCC Tartini	61	68	96.27
41 MFCC Tartini nPVI			
			88.81

Czech and Polish are both Western Slavic languages and share many similarities in vocabulary and grammar, however they possess very different and distinctive acoustic traits. Czech is generally a softer language, with a greater ratio of vowels to consonants; Czech possesses ten separate vowel sounds whereas Polish contains only seven, two of which are nasal.



Although more closely related to the Czech language, phonetically Polish is more similar to Russian and this is reflected in the full results with a lower accuracy for this comparison. Of interest here is the accuracy drop with the introduction of pitch contour information, this suggests that the contours of Czech and Polish are quite similar and are causing some confusion.

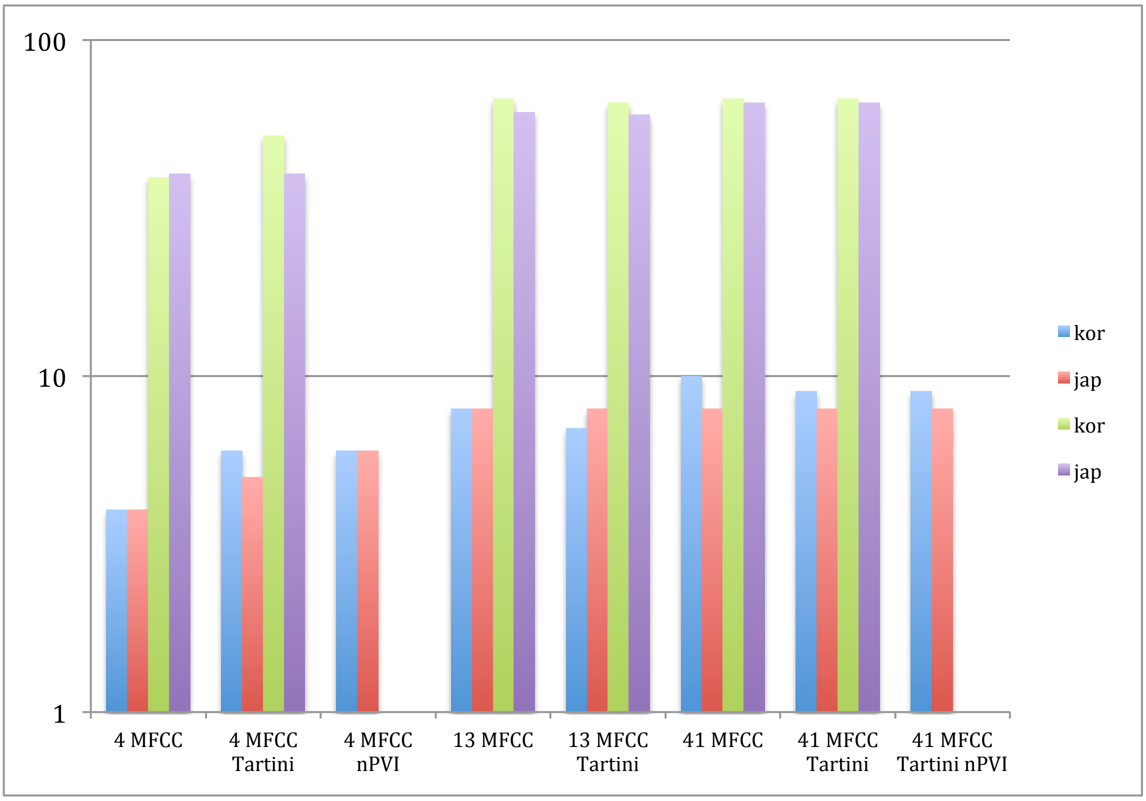


3) Korean & Japanese

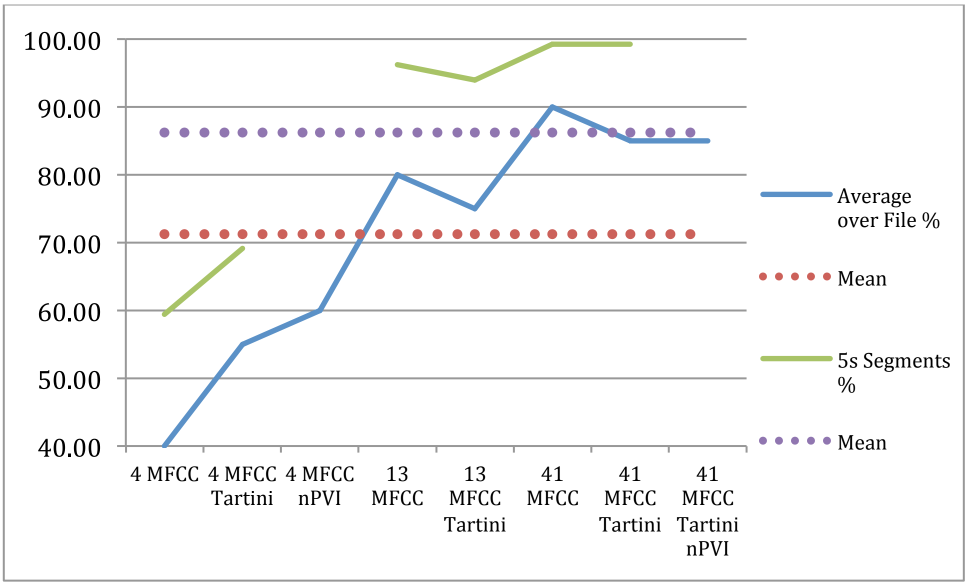
Averaged	Kor	Jap	%
4 MFCC	4	4	40
4 MFCC Tartini	6	5	55
4 MFCC nPVI	6	6	60
13 MFCC	8	8	80
13 MFCC Tartini	7	8	75
41 MFCC	10	8	90
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	9	8	85
			71.25

Instances ➤	67	66	
5s Segments	Kor	Jap	%
4 MFCC	39	40	59.40
4 MFCC Tartini	52	40	69.17
4 MFCC nPVI			
13 MFCC	67	61	96.24
13 MFCC Tartini	65	60	93.99
41 MFCC	67	65	99.25
41 MFCC Tartini	67	65	99.25
41 MFCC Tartini nPVI			
			86.22

Korean and Japanese, although both treated as isolate languages, share many traits in grammar and sentence construction. When listening to the languages separately, these similarities manifest themselves in a likeness of rhythm and to the untrained human ear they are often confused. Phonetic similarities include shared vowels and also the use of only one liquid consonant, a flap that varies between the lateral /ɹ/ and the central /r/ (Ingram & Park 1998).



The obtained results show that with a minimal number of features the system performs just below chance, however, this accuracy shows a general rise with an increasing number of features. This output is the preferred output of the system in that by providing more information the task of discrimination improves. With the addition of a pitch contour the accuracy does drop marginally, suggesting that the prosodic contour of these two languages is relatively homologous.

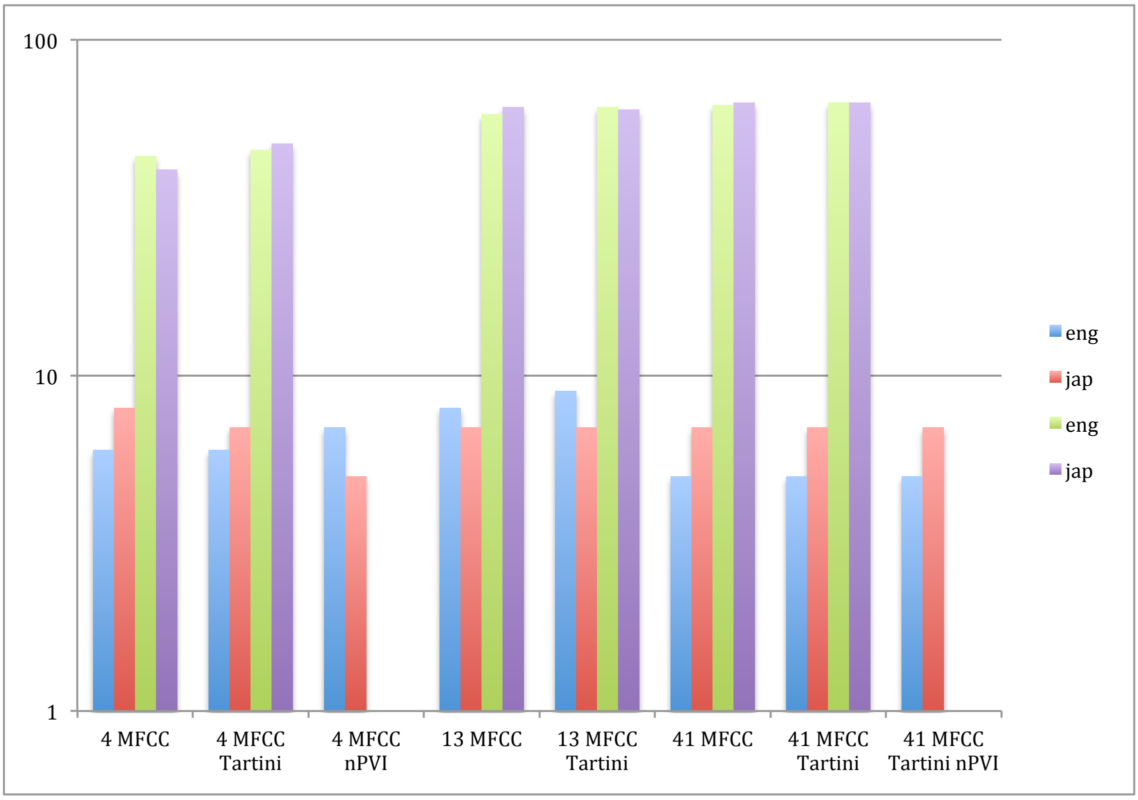


4) English & Japanese

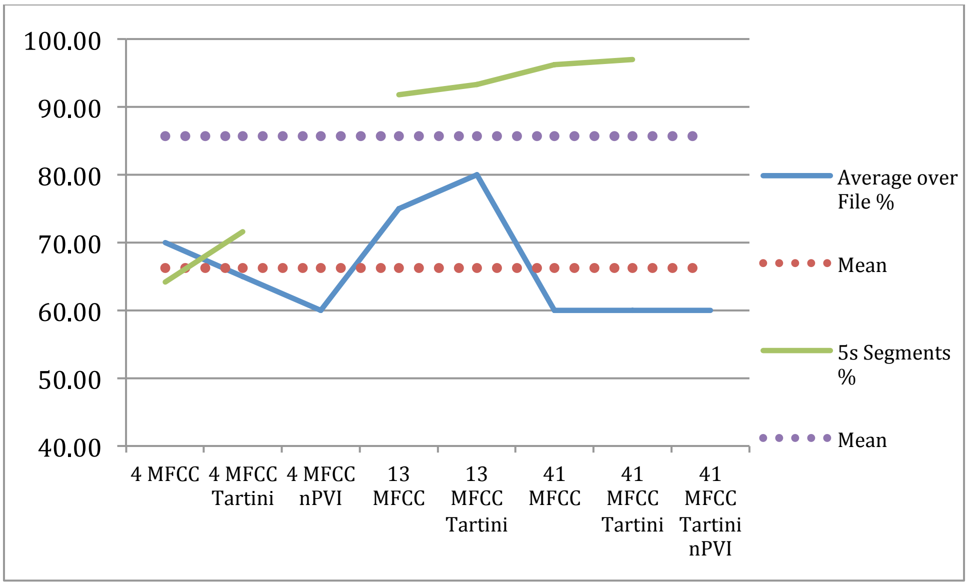
Averaged	Eng	Jap	%
4 MFCC	6	8	70
4 MFCC Tartini	6	7	65
4 MFCC nPVI	7	5	60
13 MFCC	8	7	75
13 MFCC Tartini	9	7	80
41 MFCC	5	7	60
41 MFCC Tartini	5	7	60
41 MFCC Tartini nPVI	5	7	60
			66.25

Instances ➤	68	66	
5s Segments	Eng	Jap	%
4 MFCC	45	41	64.18
4 MFCC Tartini	47	49	71.64
4 MFCC nPVI			
13 MFCC	60	63	91.79
13 MFCC Tartini	63	62	93.28
41 MFCC	64	65	96.27
41 MFCC Tartini	65	65	97.01
41 MFCC Tartini nPVI			
			85.70

These are interesting results considering that these two languages have possibly the greatest distance between each other from all others included in this study. Phonetically there are relatively few differences between English and Japanese, the latter only possesses two unique phonemes that do not occur in the English language; a lengthened 'o' vowel /oː/ and the aforementioned single liquid /ɺ-r/ as found in the syllables 'ra-ri-ru-re-ro'.



The system shows best performance when presented with a medium amount of features with a significant increase in accuracy after the introduction of pitch contour information. This is to be expected as Japanese possesses a much more tonal character across individual phonemes and its timing differs from the English stress system in the form of syllables and the 'mora'.



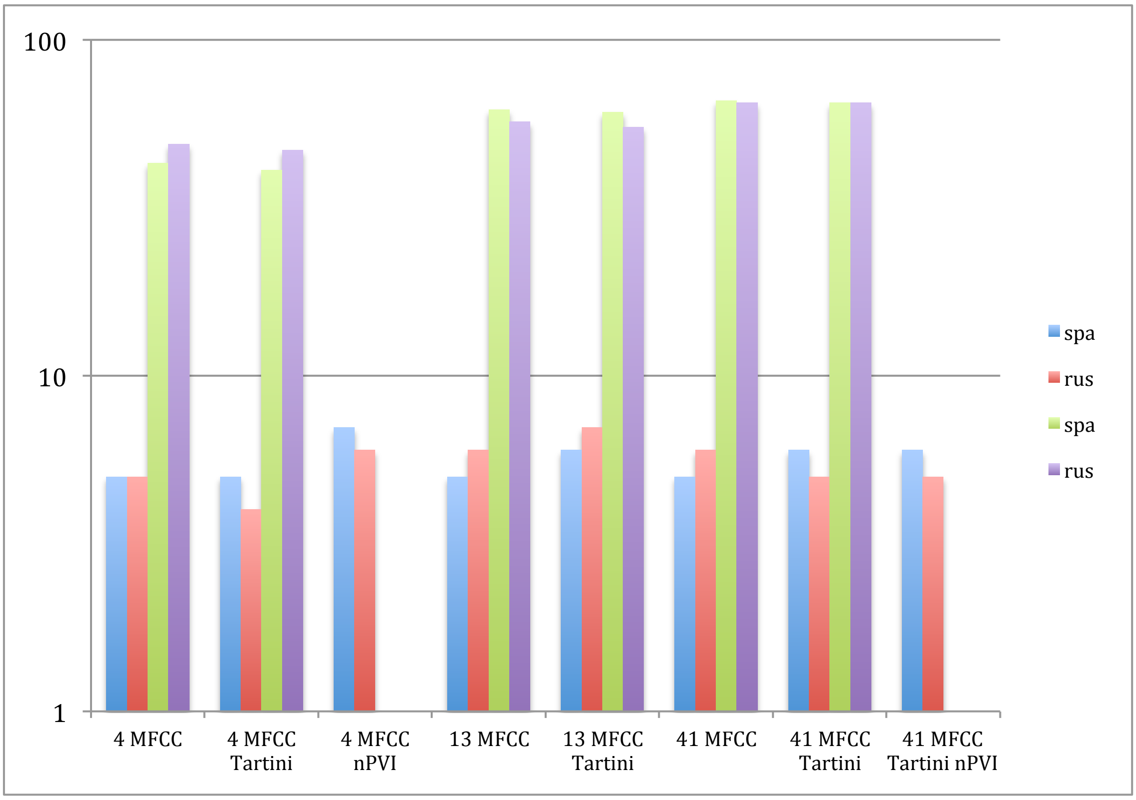
When dealing with averaged data, the system performs relatively poorly with a minimum feature set actually performing better than a maximum feature set. Performance is markedly improved with the segmentation of data. Within Japanese there is a correlation between a phoneme, the smallest unit of speech data, and the syllabic structure of the language, which may account for the great increase in accuracy seen when breaking up the speech signal.

5) Spanish & Russian

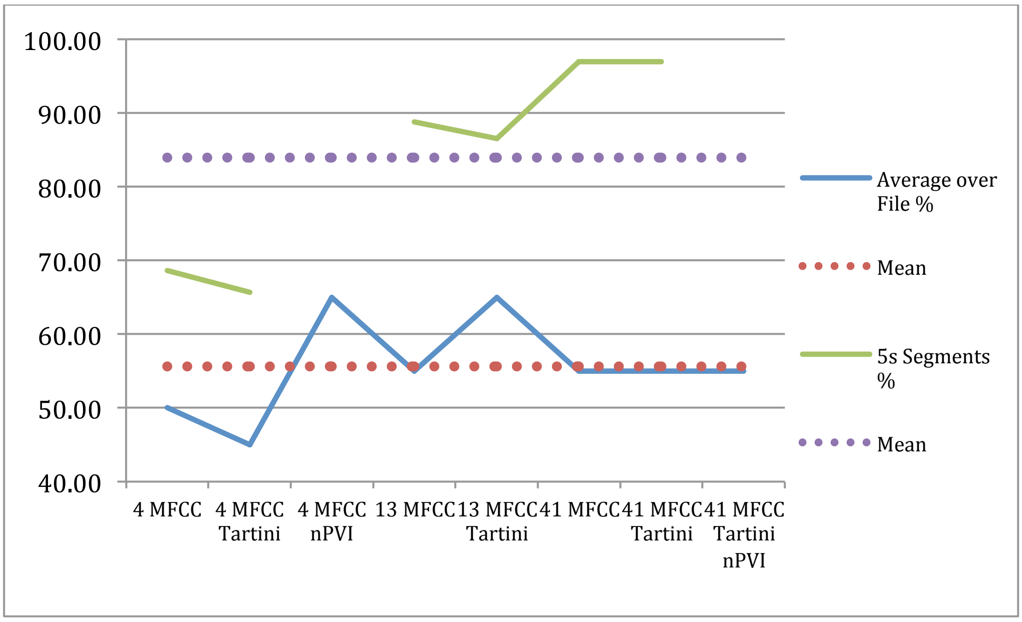
Averaged	Spa	Rus	%
4 MFCC	5	5	50
4 MFCC Tartini	5	4	45
4 MFCC nPVI	7	6	65
13 MFCC	5	6	55
13 MFCC Tartini	6	7	65
41 MFCC	5	6	55
41 MFCC Tartini	6	5	55
41 MFCC Tartini nPVI	6	5	55
			55.63

Instances ➤	66	68	
5s Segments	Spa	Rus	%
4 MFCC	43	49	68.66
4 MFCC Tartini	41	47	65.67
4 MFCC nPVI			
13 MFCC	62	57	88.81
13 MFCC Tartini	61	55	86.57
41 MFCC	66	65	97.01
41 MFCC Tartini	65	65	97.01
41 MFCC Tartini nPVI			
			83.96

Although these two languages occupy quite different spaces on the Indo-European language tree and are a significant distance from each other, the system appears to be experiencing a high level of incertitude between the two when using averaged data.



Similarities are sparse; Spanish is generally acccepted to be a syllable-timed tongue whose prosodic curve can be shown to be relatively discrete in terms of rising and falling at the onset and termination of word segments (Delattre 1965). In contrast Russian is stress-timed whose pitch contour varies more across word segments. Phonetic content differs also, with Russian possessing at least four liquid consonants depending upon dialect (Ladefoged & Maddieson 1996, Jones & Ward 1969). It is a possibility that although very dissimilar, the differences between the two are not being picked up by the system due to segmental information being lost by using averaged data.



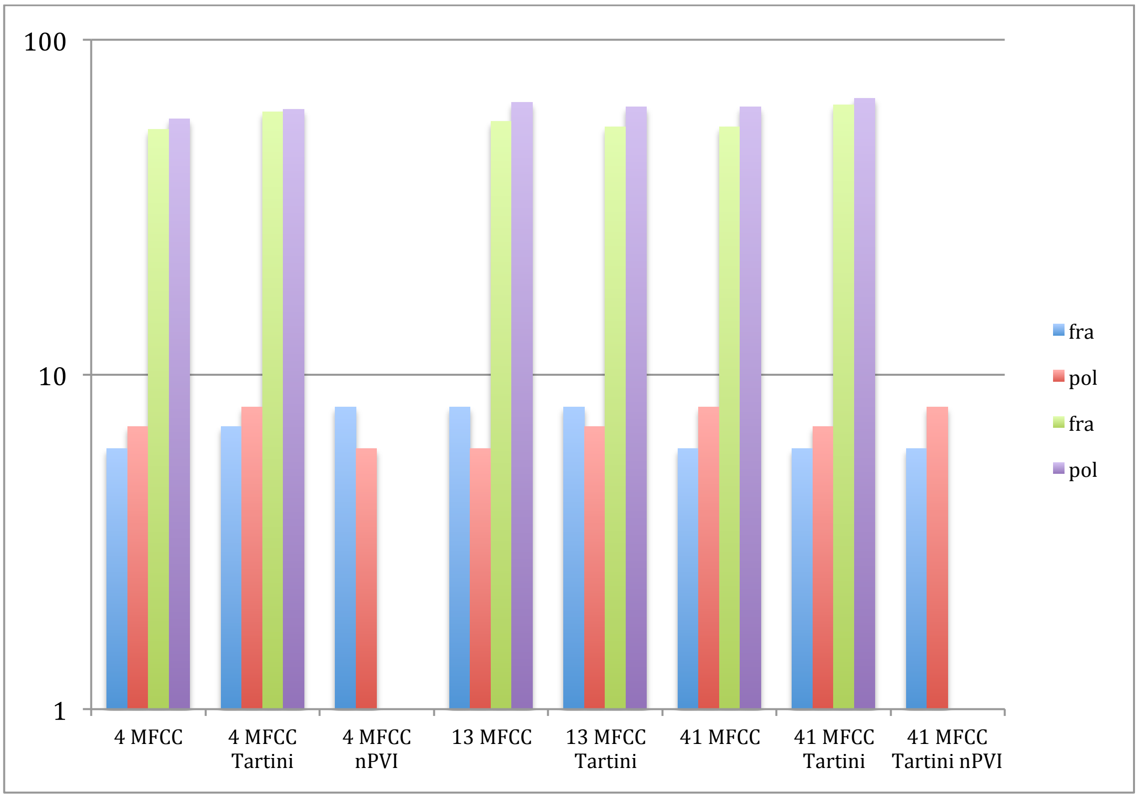
Segmentation indeed appears to remove a great deal of the confusion, although for a lower number of features the performance is still low. When a pitch contour is added to 13 MFCC vectors the rate of discrimination falls. This is in contrast to the averaged data where the performance increases under the same conditions.

6) French & Polish

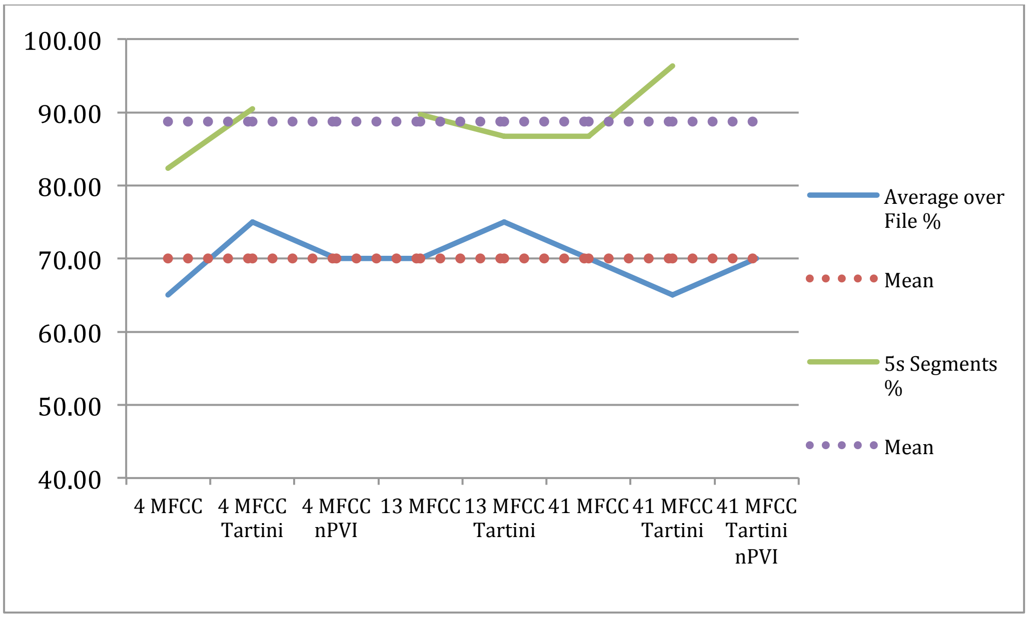
Averaged	Fra	Pol	%
4 MFCC	6	7	65
4 MFCC Tartini	7	8	75
4 MFCC nPVI	8	6	70
13 MFCC	8	6	70
13 MFCC Tartini	8	7	75
41 MFCC	6	8	70
41 MFCC Tartini	6	7	65
41 MFCC Tartini nPVI	6	8	70
			70.00

Instances ➤	67	69	
5s Segments	Fra	Pol	%
4 MFCC	54	58	82.35
4 MFCC Tartini	61	62	90.44
4 MFCC nPVI			
13 MFCC	57	65	89.71
13 MFCC Tartini	55	63	86.76
41 MFCC	55	63	86.76
41 MFCC Tartini	64	67	96.32
41 MFCC Tartini nPVI			
			88.73

This comparison shows relatively good results although I speculate that there exist various factors that could cause a lower rate of accuracy. Both French and Polish, although quite distant from each other, share one phoneme that is relative common in both languages, /ʒ/. This appears very commonly in French such as 'jour' (day) and the Polish 'jeść' (eat).



The output of the system on this comparison remains relatively stable with small fluctuations, regardless of the number of features that are provided. Of interest is the rise in accuracy for the first two instances of Tartini addition; French is generally a more animated language than Polish and this could account for the rise. However, the third addition of pitch contour information contrastingly causes a drop in accuracy.



As is seen in all other language pair comparisons segmentation causes a marked increase in performance with the system outputting close to 100% accuracy for a large feature set. Even when presented with only 4 MFCC vectors the performance is past 80%.

6.2 Comparison of Languages by Family

It was felt pertinent to assess the capability of the system to discriminate within language families, as this was the area where most localised confusion was expected, given the relatively close distance between compared languages at this level. For these results tables and charts are presented in an identical fashion to the language pair comparisons – with the addition of confusion matrices to indicate exactly which language the system thought it was auditioning. Confusion matrices are shown for the lowest feature set, 4 MFCC Vectors, as this is where the most confusion is likely to occur.

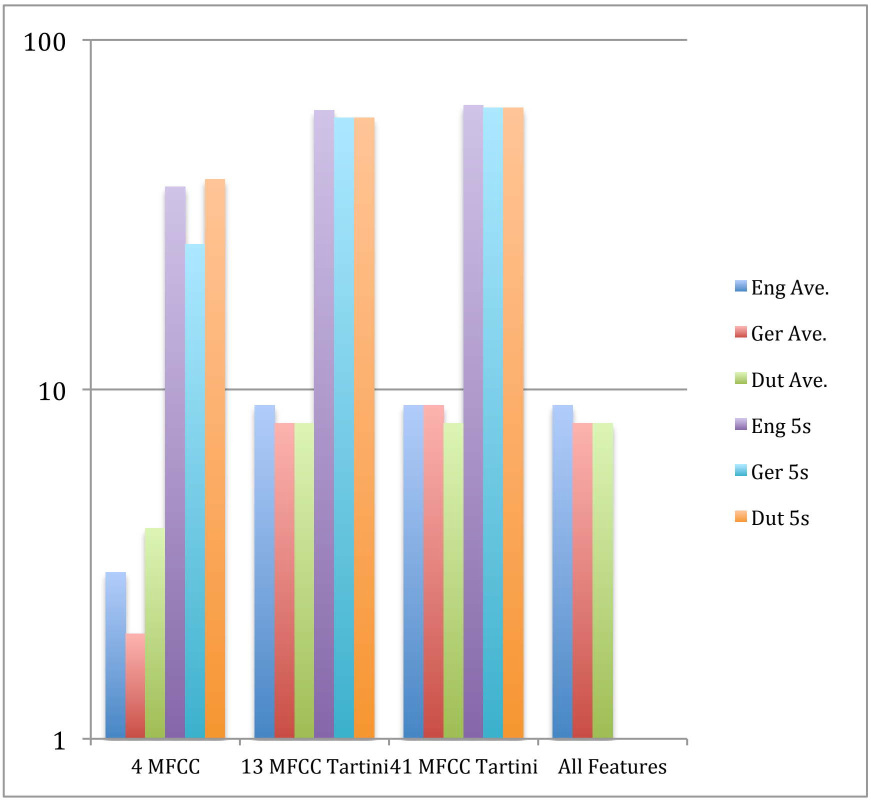
A) Germanic

Averaged	Eng	Ger	Dut	%
4 MFCC	3	1	4	26.67
13 MFCC Tartini	9	8	8	83.33
41 MFCC Tartini	9	9	8	86.67
All Features	9	8	8	83.33
				70.00

Instances ➤	68	66	67	
5s Segments	Eng	Ger	Dut	%
4 MFCC	38	26	40	51.74
13 MFCC Tartini	63	60	60	91.04
41 MFCC Tartini	65	64	64	96.02
All Features				
				79.60

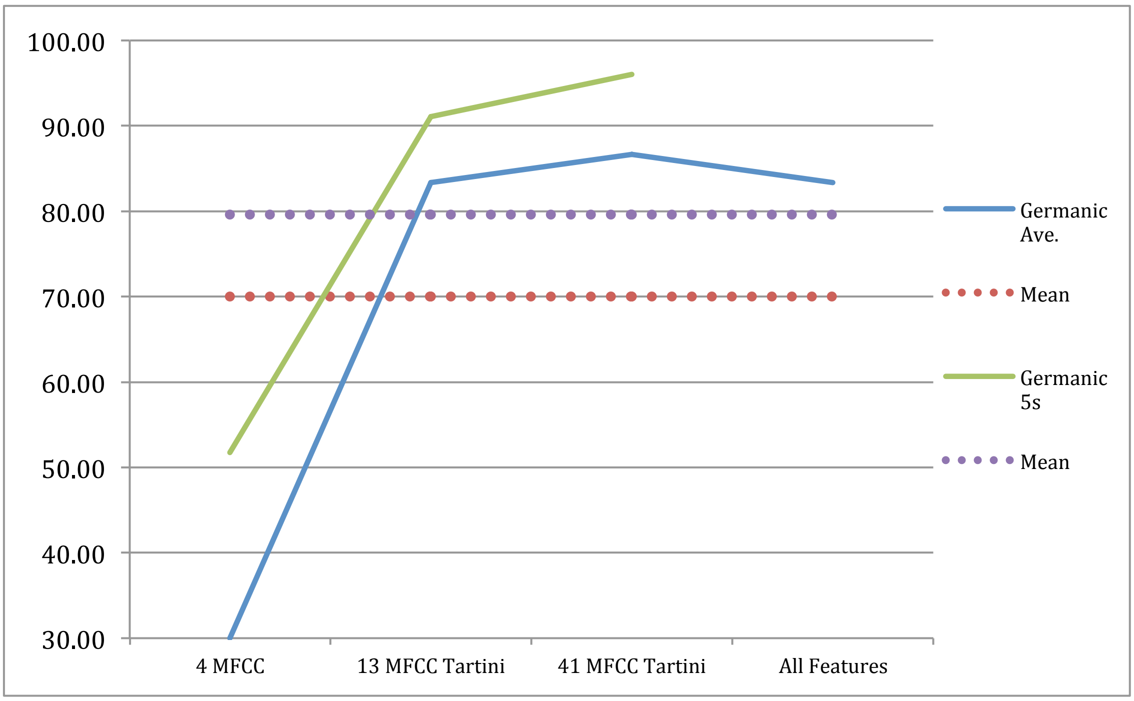
Within this language family Modern English displays the greatest distance from the others, it has developed from the Anglo-Frisian and Saxon languages (Chambers & Wilkie 1970) and also has been highly influenced by French. Given this contrast to German and Dutch, whose development has been rather uninfluenced by outside sources, unsurprisingly within this comparison it is for English that the system displays the greatest accuracy, although it also scores

Confusion Matrix - Germanic						
Averaged			5s Segments			
Eng	Ger	Dut	Eng	Ger	Dut	← Classified As
3	2	5	38	17	13	Eng
5	1	4	22	26	18	Ger
4	2	4	17	10	40	Dut



highly on German and Dutch when presented with a large number of features.

It is interesting to note that the confusion matrix tells us that when uncertain, the system is most likely to classify the speech as English and rather surprisingly there is relatively little confusion between German and Dutch, which is displayed only for averaged data.

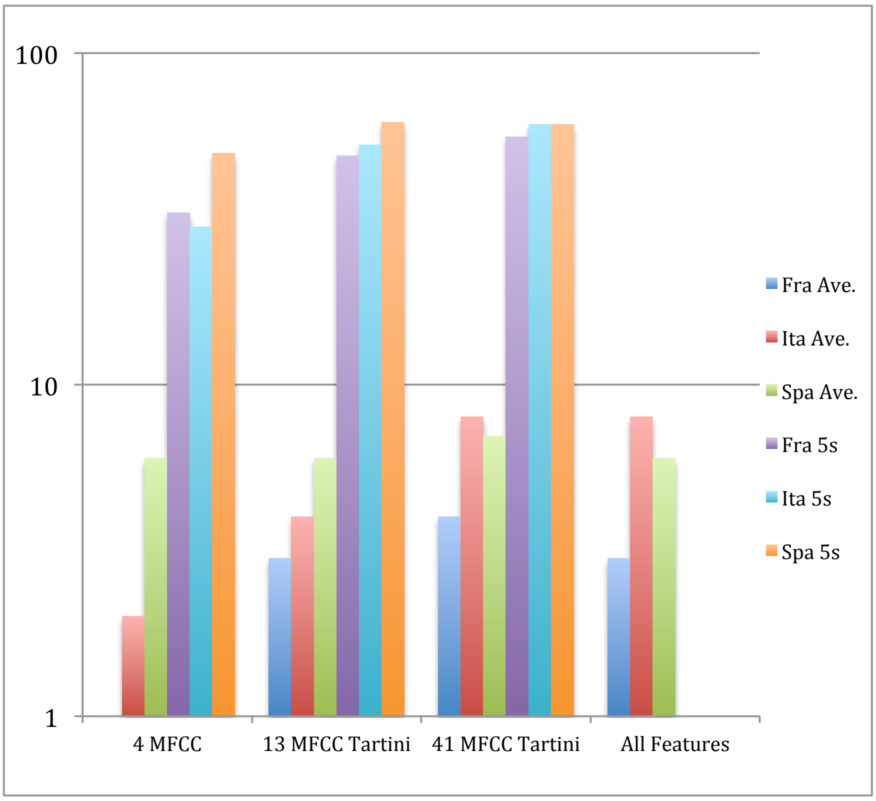


B) Romantic

Averaged	Fra	Ita	Spa	%
4 MFCC	4	6	6	53.33
13 MFCC Tartini	2	6	9	56.67
41 MFCC Tartini	2	6	7	50.00
All Features	2	7	7	53.33
				53.33

Instances ➤	67	66	66	
5s Segments	Fra	Ita	Spa	%
4 MFCC	33	30	50	56.78
13 MFCC Tartini	49	53	62	82.41
41 MFCC Tartini	56	61	61	89.45
All Features				
				76.21

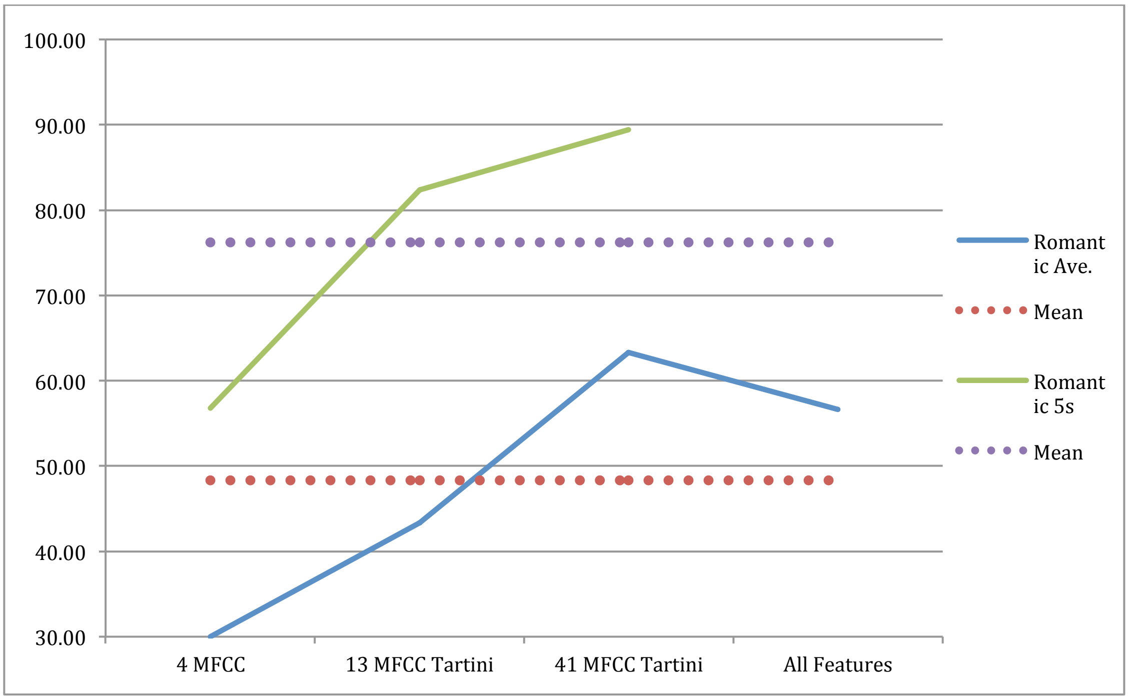
Within the Romantic language family the distance between French, Italian and Spanish is relatively small; Spanish & French are both Western Romance languages and only two major branches separate Italian & Spanish. When looking at averaged data the accuracy is, in terms of performance related to other comparisons in this study, quite low.



The system seems to display relatively little confusion when testing on averaged data save for the case of Spanish which for

Confusion Matrix - Romantic						
Averaged			5s Segments			
Fra	Ita	Spa	Fra	Ita	Spa	← Classified As
7	1	2	33	18	16	Fra
1	6	3	24	30	12	Ita
2	5	3	8	8	50	Spa

the most part is categorised as Italian. When segmenting the data it is Spanish that performs the best and Italian the worst, with a large number of Italian instances being identified as French.

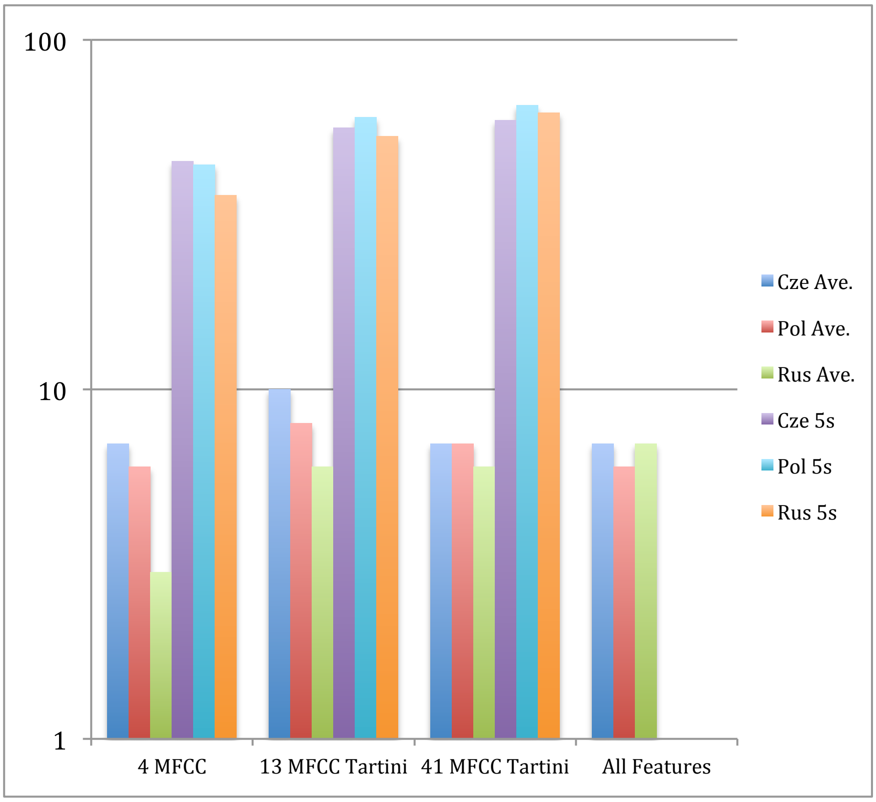


C) Slavic

Averaged	Cze	Pol	Rus	%
4 MFCC	7	4	2	43.33
13 MFCC Tartini	10	6	5	70.00
41 MFCC Tartini	9	6	5	66.67
All Features	10	7	5	73.33
				63.33

Instances ➤	65	69	68	
5s Segments	Cze	Pol	Rus	%
4 MFCC	45	44	36	61.88
13 MFCC Tartini	56	60	53	83.66
41 MFCC Tartini	59	65	62	92.08
All Features				
				79.21

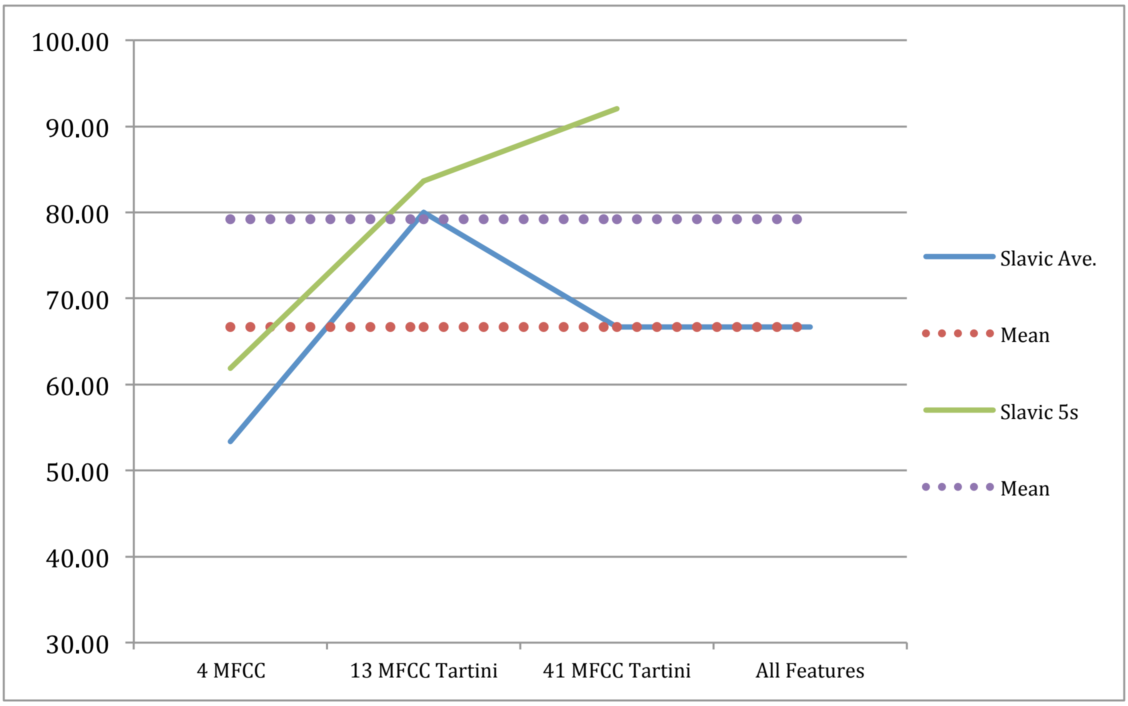
From the Slavic languages compared, Czech is the one that displays the most acoustic individuality, whereas quite a degree of phonetic similarity exists between Polish and Russian. It displays less harsh characteristics and possesses less extreme cases of stress and intonation. It does not feature /ʒ/ as frequently as Russian or Polish and has less vowel sounds than either.



It is the most readily identified language when dealing with averaged data and even with a low feature set the confusion matrix tells us that it is the top

Confusion Matrix - Slavic						
Averaged			5s Segments			
Cze	Pol	Rus	Cze	Pol	Rus	← Classified As
7	1	2	45	11	9	Cze
1	6	3	8	44	17	Pol
2	5	3	15	17	36	Rus

performer. Russian on the other hand, is most readily confused with Polish and vice versa, even when the data is segmented.

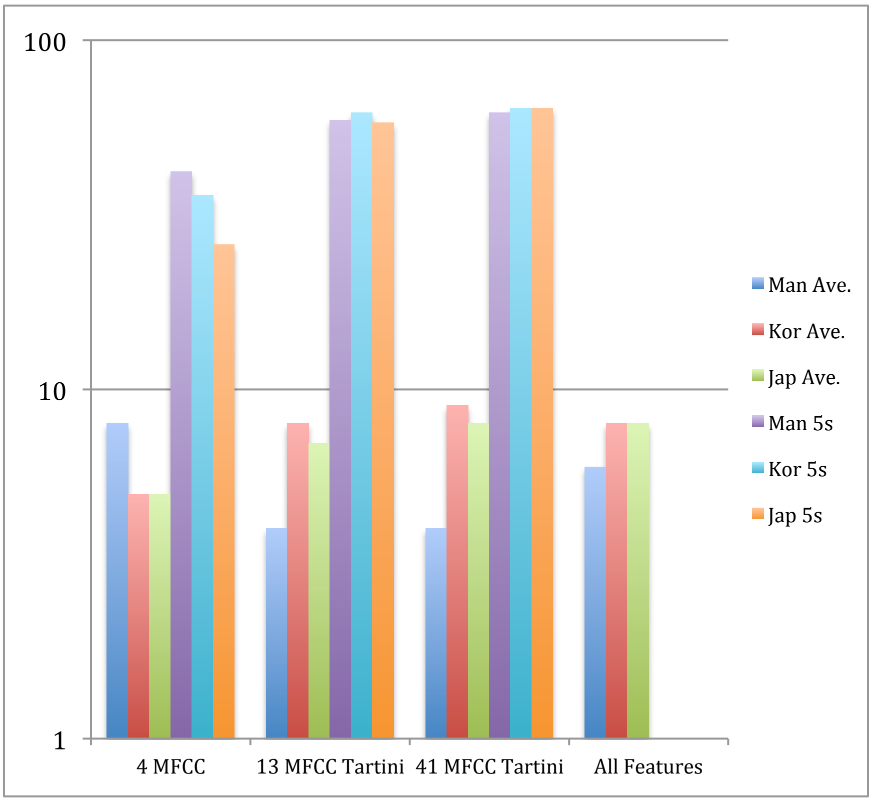


For averaged data the performance drops when presented with a large feature set, however, the segmented results show a continuous rise in accuracy. Next to the Germanic languages compared in this project, these three Slavic tongues are probably the closest to each other in terms of phonetic and prosodic cues.

D) Sino-Tibetan / Macro-Altaic

Averaged	Man	Kor	Jap	%
4 MFCC	7	5	3	50.00
13 MFCC Tartini	7	7	6	66.67
41 MFCC Tartini	7	7	7	70.00
All Features	7	9	6	73.33
				65.00

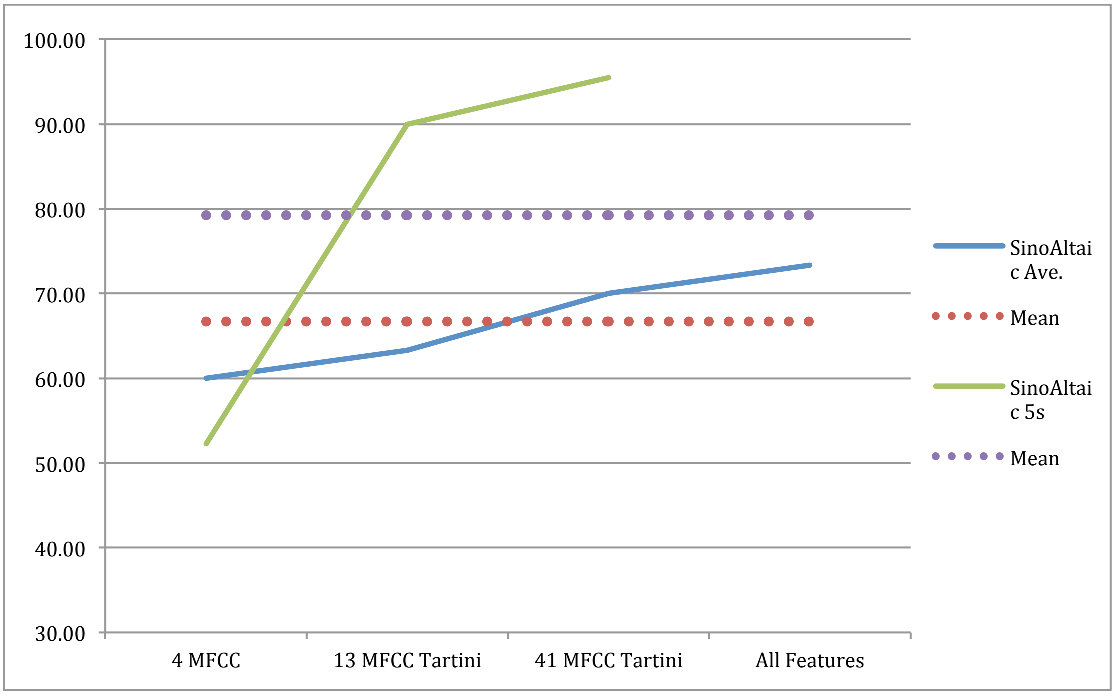
Instances ➤	66	67	66	
5s Segments	Man	Kor	Jap	%
4 MFCC	42	36	26	52.26
13 MFCC Tartini	59	62	58	89.95
41 MFCC Tartini	62	64	64	95.48
All Features				
				79.23



This technically is not a comparison within a language family as the three compared languages here span three distinct languages groups, due to the traditional placement of Korean and Japanese in their own isolated families.

Confusion Matrix - Sino-Altaic						
Averaged			5s Segments			
Man	Kor	Jap	Man	Kor	Jap	← Classified As
8	0	2	42	15	9	Man
3	5	2	11	36	20	Kor
1	4	5	8	32	26	Jap

A comparison between these languages was performed for three reasons; i) because of the recent speculation into the genetic relationship between Korean and Japanese, ii) the recent grouping of Korean and Japanese under the Macro-Altaic language branch and also, iii) with reference to Navrátil's (Navrátil 2001) studies into human language identification abilities, whose results showed a much lower success rate for identifying non-European languages by the tested listeners.



With a low number of spectral bins, Mandarin performs the best suggesting that its general spectral characteristics are significantly different from either Korean or Japanese.

6.3 Comparison Between All Languages

I finally present the results obtained from presenting the system with all of the twelve languages in this project. Naïve Bayes was seen to produce a mean accuracy of 30.28% on averaged data and 45.05% on segmented data, with maximum feature set accuracy of 34.17% and 69.16% respectively. The MLP performed only marginally better on averaged data at a mean of 30.38% and with a rise in accuracy to 58.45% when using segmented data.

MLP - Averaged	Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	%
4 MFCC	2	1	3	0	2	2	6	0	1	1	1	4	19.17
4 MFCC Tartini	1	0	6	0	1	0	6	0	1	0	0	5	16.67
4 MFCC Tartini nPVI	2	4	8	3	1	1	5	1	0	2	1	3	25.83
13 MFCC	1	6	3	2	2	1	4	5	0	1	4	3	26.67
13 MFCC Tartini	2	7	6	0	3	2	7	4	1	2	3	3	33.33
13 MFCC Tartini nPVI	3	6	5	3	5	1	4	4	1	2	2	5	34.17
41 MFCC	4	6	5	3	6	1	8	2	2	2	7	5	42.50
41 MFCC Tartini	3	6	5	3	5	1	7	2	3	4	5	6	41.67
41 MFCC Tartini nPVI	3	6	5	3	4	1	7	2	1	2	7	6	39.17
													31.02

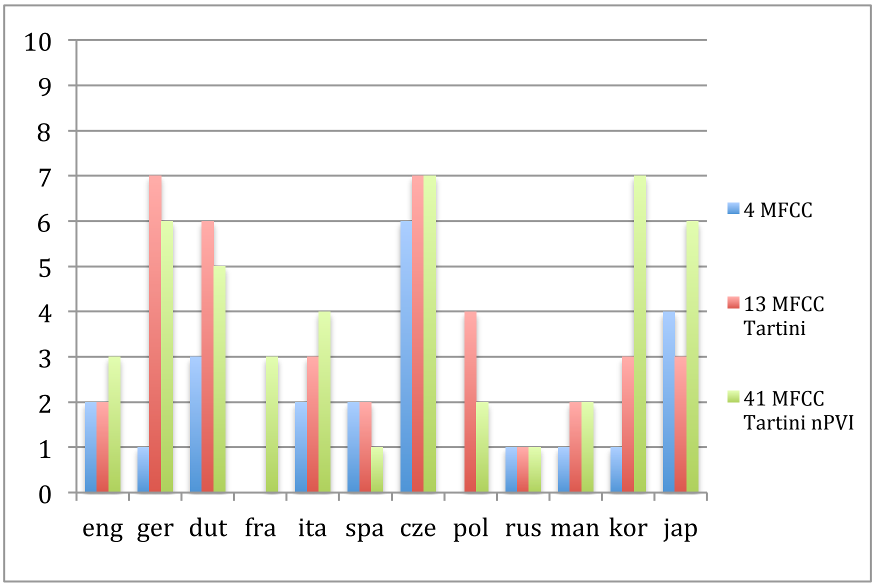
However, when using segmented data and a large features set, that of 41 MFCC vectors and a pitch contour, the system was able to achieve 89.01% accuracy.

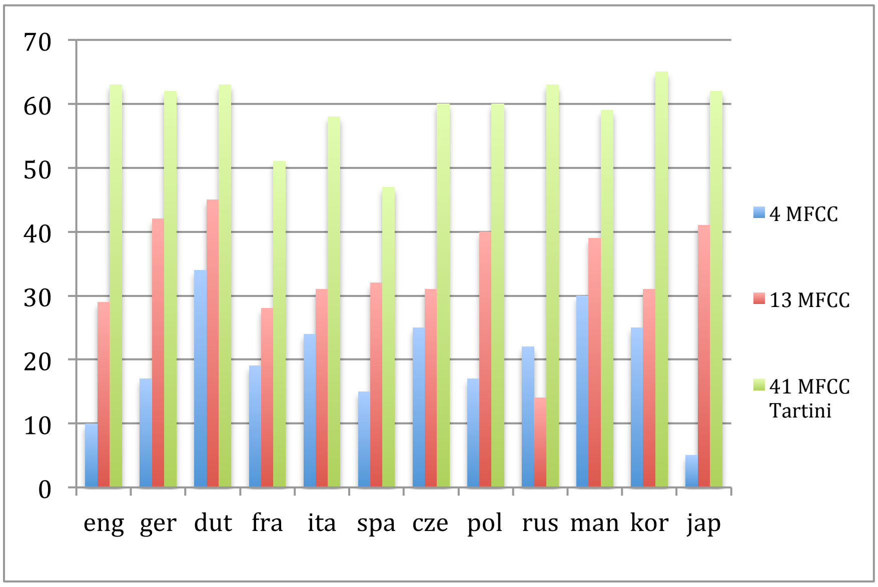
Instances ➤	68	66	67	67	68	66	65	67	68	66	68	66	
MLP - 5s Segments	Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	%
4 MFCC	10	17	34	19	24	15	25	17	22	30	25	5	30.34
4 MFCC Tartini	20	25	40	26	22	24	35	16	21	31	25	12	37.08
4 MFCC Tartini nPVI													
13 MFCC	29	42	45	28	31	32	31	40	14	39	31	41	50.31
13 MFCC Tartini	43	48	43	31	34	27	29	43	36	40	33	43	56.18
13 MFCC Tartini nPVI													
41 MFCC	58	61	63	53	58	48	58	59	60	56	65	64	87.77
41 MFCC Tartini	63	62	63	51	58	47	60	60	63	59	65	62	89.01
41 MFCC Tartini nPVI													
													58.45

Presented below are confusion matrices for the twelve-language tests. I also include the confusion matrix for the highest feature set for comparison.

All Languages - 4 MFCC																								
Averaged												5s Segments												
Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	← As
2	1	2	0	2	1	0	1	1	0	0	0	10	5	1	9	7	5	4	7	3	3	9	5	Eng
0	1	1	0	2	0	2	1	2	1	0	0	1	17	7	11	10	6	2	6	0	0	2	4	Ger
0	0	3	2	0	1	0	0	1	0	2	1	2	3	34	3	6	3	0	6	0	2	6	2	Dut
0	1	1	0	2	0	1	0	2	3	0	0	1	4	6	19	12	5	4	1	1	11	0	3	Fra
0	0	0	2	2	0	2	0	1	3	0	0	4	5	8	9	24	4	4	3	1	3	0	1	Ita
2	1	2	0	0	2	0	1	0	1	1	0	3	6	10	5	1	15	1	5	0	0	11	9	Spa
0	1	0	0	2	0	6	0	1	0	0	0	8	8	1	6	2	4	25	8	0	3	0	0	Cze
1	0	2	1	0	1	1	0	0	0	1	3	4	5	6	1	1	2	1	17	3	12	12	5	Pol
1	4	1	0	1	0	1	1	1	0	0	0	2	6	12	3	1	3	1	5	22	6	5	2	Rus
0	1	1	2	1	0	1	1	2	1	0	0	7	4	1	6	5	4	0	8	0	30	0	1	Man
0	1	1	0	0	3	0	2	0	0	1	2	2	6	12	4	0	3	1	5	0	5	25	4	Kor
0	1	1	0	0	1	0	3	0	0	0	4	1	11	7	4	5	7	2	12	0	2	10	5	Jap

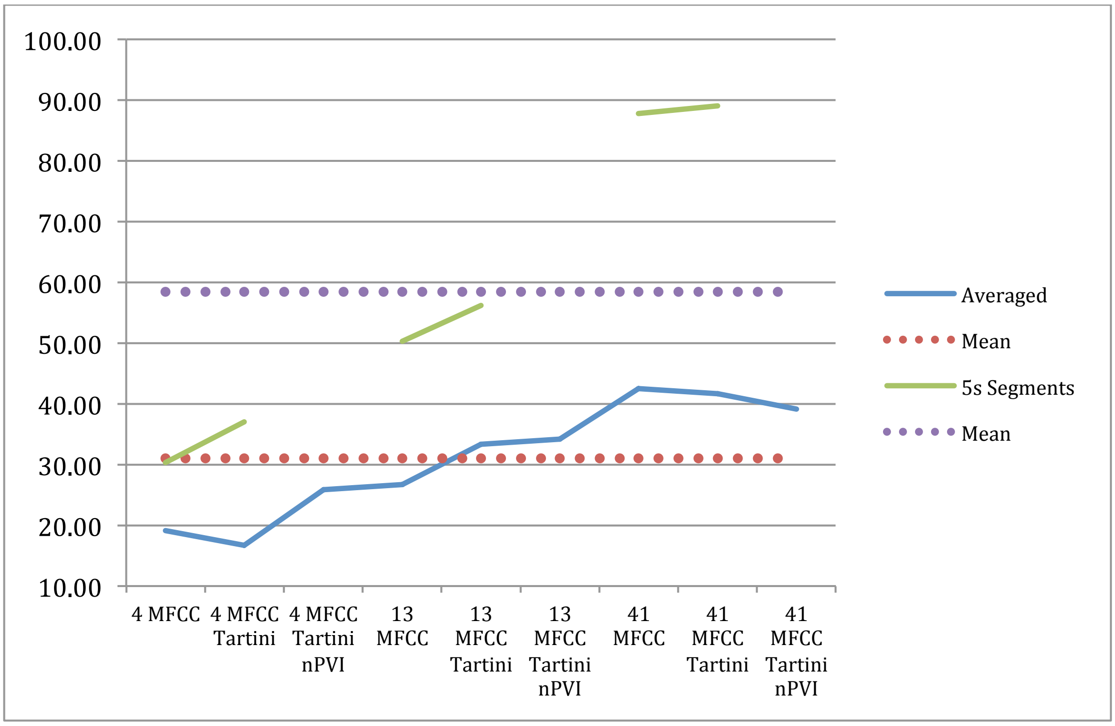
All Languages - 41 MFCC + Pitch Contour																								
Averaged												5s Segments												
Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	Eng	Ger	Dut	Fra	Ita	Spa	Cze	Pol	Rus	Man	Kor	Jap	← As
3	0	0	1	1	3	0	0	2	0	0	0	63	2	0	0	0	0	0	2	0	0	0	1	Eng
0	6	0	0	0	0	1	1	1	0	1	0	0	62	0	0	1	0	0	1	0	1	1	0	Ger
0	1	5	0	0	0	0	2	0	2	0	0	0	0	63	0	0	2	0	1	0	1	0	0	Dut
1	0	0	3	1	2	0	1	0	1	0	1	1	3	2	51	3	1	0	3	1	1	0	1	Fra
1	0	1	1	5	0	0	1	0	1	0	0	1	1	1	1	58	0	0	1	2	0	1	0	Ita
0	1	0	3	0	1	0	1	1	1	1	1	0	1	2	4	3	47	2	1	2	0	4	0	Spa
0	1	0	0	0	1	7	0	0	0	0	1	0	1	0	0	0	1	60	1	1	0	0	1	Cze
0	0	1	1	1	0	0	2	0	0	4	1	0	0	1	0	0	0	1	60	0	1	2	4	Pol
0	1	1	0	0	1	1	1	3	0	1	1	1	1	0	0	1	0	1	1	63	0	0	0	Rus
0	0	1	1	0	0	0	0	1	4	1	2	0	0	0	1	2	3	0	0	0	59	1	0	Man
0	0	0	0	0	1	0	4	0	0	5	0	0	1	0	0	0	0	0	1	0	0	65	0	Kor
1	0	0	1	0	0	0	0	1	1	0	6	0	0	0	0	1	0	0	1	0	2	0	62	Jap





When dealing with the lowest feature set, one can see a great deal of confusion which is to be expected when attempting to discriminate between all twelve languages. Of interest here is the fact that confusion occurs not only within language families, where similar languages may be expected to occupy, rather a broad spectrum of confusion exists across language families. Not only are there similarities between language families but different languages of the world may share some features – even though their distance is great.

The graph below confirms the data in the confusion matrices, that is, that the accuracy of my system is greatly improved when dealing with segmented data.



6.4 Results Summary

Within language pairs results were encouraging, especially between languages whose distance is small and are often confused by human listeners. It was interesting to see less accurate results for more distant language pairs. This is a discrimination that may be easy for a human to make but may not be so clear-cut for an automatic system. Humans have the benefit of some previous internal models of language and an idea of what foreign dialects may sound like.

Comparisons between language families show a general rise in accuracy despite short distance between languages. The level of confusion was at times unexpected, as shown in the confusion matrices.

A comparison between all languages reflected the previously stated assumption that increasing task complexity would occur in the case that the system was presented with a greater number of languages. When taking average values from speech signals it is apparent that the system is losing a great deal of resolution as the majority of useful cues in language can be seen to occur at the segmental level.

The act of segmenting the speech data before presenting it to the system was shown to vastly increase the accuracy of the system, with a maximum accuracy of 89.01% on twelve languages, for a feature set comprising of 41 MFCC Vectors and a pitch contour.

Chapter Seven

Conclusions

7.1 Achievement of Objectives

The stated aim of this project was to construct a system that could successfully discriminate between a set of twelve languages with an error rate of less than 20%. My results show that this aim has been achieved through the use of an MLP with a comparatively large feature set. The system however is not optimal as the steps that must be taken to repeat such an output are not part of the same program. Separate libraries handle the two main functions of the system, feature extraction and machine learning – as such the process is rather laborious.

Also of issue is the length of time that the system takes to generate such results. For single speech files, the feature extraction is relatively fast, with the SCMIR library efficiently gathering meaningful information at an acceptable rate. For each language pair, where ten speech files are analysed and a corresponding ARFF file generated, this process takes between twenty and ninety seconds dependent on the feature set requested – more features take more time to extract. Given that 528 comparisons were made, the batch call to produce each run of these ARFF files took approximately two hours. When the files were setup to extract segmented information, this time approximately doubled. The batch call to invoke the WEKA machine learning tools also took a significant amount of time. When presented with 528 ARFF files the system took roughly one and a half hours to produce all the outputs. I believe that some of this time was due to the fact that I had specifically requested that output files be created for each comparison made. Had the data been passed directly onto another module, say a neural net within SuperCollider, the time taken by WEKA to classify data would have been lessened.

Several problems were encountered during this study. The nPVI function, although working and giving correct values, was causing issues when extracting features from segmented data in the guise of an extra feature. This was due to the ARFF generation patch needing to know from which segment it was currently extracting features and which onsets to select for the nPVI calculation. I believe that the addition of nPVI information to segmented data could lead to even higher accuracy. For the batch calling of the WEKA functions, I was unable to get the '`.pathMatch`' syntax in SuperCollider to function properly. This would have allowed me to iterate over a folder containing a large number of ARFF files and call the WEKA toolkit on them. As this was not working in time I had to hard code the paths of all ARFFs to be classified to ensure functionality.

7.2 Further Work

The nPVI function could be improved upon by using values that are taken from vowel onsets, as the onset currently detected in the speech signal are not exclusively formant onsets; they are possibly stops, fricative sounds, labial, glottal and dental sounds and this could decrease the function's effectiveness.

If I were to extend this work I would first explore the use of the neural network within SuperCollider. It is my belief that by not having to port the obtained feature data to another platform such as WEKA the process would not only be greatly sped up but also enable to whole LiD process to be contiguous and within the same environment.

The system could benefit from a greater amount of data. Previous studies have varied greatly in their approaches concerning speaker and gender independence (Muthusamy 1993) and the testing of the system on data that has been obtained from a greater number of speakers both male and female is warranted. With regards to speech signals from which features are extracted, it may be of merit to investigate the use of noisier signals, as this is more representative of the real world situations in which such a system would be of benefit. I would very much like to investigate the possibilities of a real-time system that accepts microphone input as an acceptable signal.

The segmentation of data was shown to greatly improve the accuracy of the system in all language pair comparisons and also when comparing all twelve languages. In addition to separating a speech signal into segments of a set length, it would most likely benefit future versions of this system if signals were split into segments equating to individual phonemes. Spectral and prosodic features at the phonemic level would present any classification algorithm with a better model of a given language (Hazen & Zue, 1997, Matějka 2004b) and suprasegmental information could be used to augment this information (Ramus & Mehler 1999).

7.3 Executive Summary

This project has shown that the LiD problem can be tackled through the investigation of acoustical features alone without resorting to more complex statistical methods. I believe this reflects the assumption that the human ability to make reasonable estimates, given only a short period of audition, is predominantly based on information contained in acoustic cues.

Bibliography

Adda-Decker M (2008) "Automatic Language Identification" in 'Language and Speech Processing', Mariani J [ed.] (2009), Wiley ISTE, http://www.limsi.fr/Individu/madda/publications/PDF/wiley2009languageID.pdf, Retrieved 9th November 2010

Aitchinson J (1976) "The Articulate Mammal", London: Hutchinson Press

Allen F, Ambikairajah E & Epps J (2005) "Language identification using warping and the shifted delta cepstrum", in 'IEEE 7th Workshop on Multimedia Signal Processing', pp.1–4

Arai T (1995) "Automatic Language Identification Using Sequential Information of Phonemes" in 'Proceedings of IEICE Transactions on Information and Systems' Vol.E78-D, pp.705-711

ARFF File Format (online), http://www.cs.waikato.ac.nz/~ml/weka/arff.html, retrieved 27th March 2011

Beckwith, C I (2004) "Koguryo: The Language of Japan's Continental Relatives: An Introduction to the Historical-Comparative Study of the Japanese-Koguryoic Languages", Leiden: Brill

Bee S & Cooke M [ed.] (1993) "Visual Representations of Speech Signals", Chichester: John Wiley & Sons

De Boysson-Bardies B & Vihman M (1991) "Adaptation to Language: Evidence From Babbling and First Word From Four Languages", in 'Language' 67(2), pp.297-319

British Computer Society Code of Good Conduct for BCS Members. BCS, (online): http://www.bcs.org/server.php?show=nav.6030, retrieved 15th November 2010

Campbell W M, Gleason T P, Reynolds D A, Singer E & Torres-Carrasquillo P A (2003) "Acoustic, Phonetic and Discriminative Approaches to Automatic Language Identification", in 'Proceedings of Eighth European Conference on Speech Communication and Technology', pp.1345-1348

Chambers WW & Wilkie J R (1970) "A Short History of the German Language", Bungay: Chaucer Press Ltd

Chen T, Liu Z & Wang Y (1998) "Audio Feature Extraction and Analysis for Scene Segmentation and Classification", in 'The Journal of VLSI Signal Processing' 20(1), pp.61-79

Chen T (2001) "Automatic accent identification using Gaussian mixture models", in 'IEEE workshop on Automatic Speech Recognition and Understanding'

Coblin W S (2000) "A Brief History of Mandarin," Journal of the American Oriental Society 120(4), pp.537–552

Collins N (2010a) "SuperCollider Music Information Retrieval Library", http://www.informatics.sussex.ac.uk/users/nc81/code.html, retrieved October 2010

Collins N. (2010b) "An Introduction to Computer Music", pp.248-262, Chichester: John Wiley & Sons

Delattre P (1965) "Comparing the Phonetic Features of English, French, German and Spanish: An Interim Report", London: George G. Harrap & Company Limited

Ethnologue, 'Statistical Summaries – Summary by language size', http://www.ethnologue.org/ethno_docs/distribution.asp?by=size - Retrieved Monday 15th November 2010

Ezzaidi H, Rouat J & O'Shaughnessy D (2001) "Towards Combining Pitch and MFCC for Speaker Identification Systems," in 'Proceedings of EuroSpeech 2001', pp.2825–2828

Farinas J & Pellegrino F (2001) "Automatic Rhythm Modelling for Language Identification", in 'Seventh European Conference on Speech Communication and Technology'

Ganchev T, Fakotakis N & Kokkinakis G (2005) "Comparative Evaluation of Various MFCC Implementations On the Speaker Verification Task", in 'Proceedings of the SPECOM-2005', pp.191–194

Gibbon D & Gut U (2001) "Measuring Speech Rhythm" in 'Seventh European Conference on Speech Communication and Technology', pp.95-98

Gold B, Morgan N (2000) "Speech and Audio Signal Processing: Processing and Perception of Speech and Music", Chichester: John Wiley & Sons

Grabe E & Low E L (2002) "Durational Variability in Speech and the Rhythm Class Hypothesis", in 'Papers in Laboratory Phonology' 7, pp.515-546

Greenberg J H (2000) "Indo-European and its Closest Relatives. The Eurasiatic Language Family", Stanford: Stanford University Press

Hallé P A, de Boysson-Bardies B & Vihman M (1991) "Beginnings of Prosodic Organisation: Intonation and Duration Patterns of Disyllables produced by Japanese and French Infants", in 'Language and Speech' 34, pp.299-318

Hazen T J & Zue V W (1997) "Segment-Based Automatic Language Identification", in 'Journal of the Acoustical Society of America' 101(4), pp.2323-2331

Hermansky H (1989) "Perceptual Linear Predictive (PLP) Analysis of Speech" in 'Journal of the Acoustical Society of America' 87(4), pp.1738-1752

Hoboken N J (2006) "Computational Auditory Scene Analysis: Principles, Algorithms and Applications", in 'Journal of the Acoustical Society of America', 124(1), pp.13

House A S & Neuburg E P (1977) "Toward Automatic Identification of the Language of an Utterance. Preliminary Methodological Considerations", in 'Journal of the Acoustical Society of America', 62(3), pp.708-13

Port R F, Dalby J & O'Dell M (1987) "Evidence for Mora-timing in Japanese", in 'Journal of the Acoustical Society of America' 81(5), pp.1574-1585

Ingram J C L & Park S G (1997) “Cross-Language Vowel Perception and Production by Japanese and Korean Learners of English”, in ‘Journal of Phonetics’ 25, pp.343–370

Ingram J C L & Park S G (1998) “Language, context, and speaker effects in the identification and discrimination of English r and l by Japanese and Korean listeners”, in ‘Journal of the Acoustical Society of America’ 103(2), pp.1161-1174

Jelinek F (1997) “Statistical Methods for Speech Recognition”, Massachusetts: MIT Press

Jones D & Ward D (1969) “The Phonetics of Russian”, Cambridge: Cambridge University Press

Joos M (1948) “Acoustic Phonetics – Supplement to ‘Language – Journal of the Linguistic Society of America’”, Baltimore: Waverly Press

Jurafsky D & Martin J H (2009) “Speech and Language Processing – An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition”, 2nd ed., New Jersey: Pearson Education, Inc.

Kirchhoff K, Parandekar S & Bilmes J (1993) “Mixed-memory Markov Models for Automatic Language Identification”, in ‘Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing’, vol.1

Kirchhoff K & Parandekar S (2001) “Multi-stream Statistical N-gram Modelling with Application to Automatic Language Identification” in ‘Proceedings of EuroSpeech 2001’, pp.803–806

Kwasny S, Wu W, Kalman B & Engebretson E (1993) “Identifying Language from Raw Speech – An Application of Recurrent Neural Networks”, in ‘Proceedings of 5th Midwest Artificial Intelligence and Cognitive Science Conference”

Ladefoged P & Maddieson I (1996) “The Sounds of the World’s Languages”, Oxford: Blackwell Publishers Ltd

Ladefoged P (2001) “A Course In Phonetics”, 5th Ed., London: Harcourt College

Levitt A & Wang Q (1991) “Evidence for Language-specific Rhythmic Influences in the Reduplicative Babbling of French and English-Learning Infants” in ‘Language and Speech’ 34, pp.235-249

Li K P & Edwards T J (1980) “Statistical Models For Automatic Language Identification”, in ‘Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing’

Lin C Y & Wang H C (2005) “Language Identification Using Pitch Contour Information”, in ‘ Proceedings of International Conference on Acoustics, Speech and Signal Processing’, pp.601-604

Ling L E, Grabe E, & Nolan F (2000) “Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English”, in ‘Language and Speech’ 43(4), pp.377-401

Logan B (2000) “Mel Frequency Cepstral Coefficients for Musical Modelling”, in ‘International Symposium on Music Information Retrieval’, vol.28

Matějka P, Černocky J & Sigmund M (2004a) “Introduction to Automatic Language Identification”, in ‘Conference Proceedings of Radioelektronika’, pp.112-115

Matějka P (2004b) “Automatic Language Identification Using Phoneme and Automatically derived unit strings,” in ‘Text, Speech and Dialogue, Proceedings’, pp.147–153

McLeod P & Wyvill G (2005) “A Smarter Way to Find Pitch”, in ‘Proceedings of International Computer Music Conference’

McNeill D (1970) “The Acquisition of Language: The Study of Developmental Psycholinguistics”, New York: Harper & Row

Moore B C J (2003), “An Introduction to the Psychology of Hearing”, London: Elsevier Academic Press

Muthusamy Y K (1993) “A Segmental Approach to Language Identification”, PhD Thesis, Oregon Graduate Institute of Science & Technology

Muthusamy Y K, Barnard E & Cole R A (1994) “Automatic Language Identification: A Review/Tutorial", in ‘IEEE Signal Processing Magazine, 11(4), pp.33-41

Nakagawa S, Ueda Y & Seino T (1992) “Speaker-independent, Text-independent Language Identification by HMM”, in ‘Proceedings of International Conference on Spoken Language Processing’, pp.1011-1014

Navrátil J (2006) “Automatic Language Identification”, in ‘Multilingual Speech Processing’, Schultz T & Kirchhoff K [eds.], London: Elsevier, pp.233–272

Navrátil J (2001) “Spoken Language Recognition - A Step Toward Multilinguality in Speech Processing”, in ‘IEEE Transactions on Speech and Audio Processing’ 9(6), pp.678-685

Otake T & Cutler A [ed.] (1996) “Phonological Structure and Language Processing: Cross Linguistic Studies”, Berlin: Mouton de Gruyter

Picton P (1994) “Introduction to Neural Networks”, Basingstoke: MacMillian Press Ltd

Ramat A G & Ramat P (1998) “The Indo-European Languages”, London: Routledge

Ramus F (2002) “Acoustic Correlates of Linguistic Rhythm: Perspectives”, in ‘Proceedings of the International Conference of Speech & Prosody’, pp.115–120

Ramus F & Mehler J (1999) “Language Identification With Suprasegmental Cues: A Study Based on Speech Resynthesis”, in ‘The Journal of the Acoustical Society of America’ 105(1), p.512

Roads C (1996) “The Computer Music Tutorial”, Massachusetts: MIT Press, pp.514-516

Rogers H E (2000) “The Sounds of Language: An Introduction to Phonetics”, Harlow: Pearson Education Ltd

Rouas J L, Farinas J & Pellegrino F (2003) “Automatic Modelling of Rhythm and Intonation for Language Identification”, in ‘Proceedings of the 15th International Congress of Phonetic Sciences’, pp.567–570

Rouas J L (2005) “Rhythmic Unit Extraction and Modelling for Automatic Language Identification”, in ‘Speech Communication’ 47(4), pp.436–456

Schouten M E H (1992) “The Auditory Processing of Speech: From Sounds to Words”, Berlin: Mouton de Gruyter

Slaney E S M (1997) “Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator”, in ‘Proceedings of International Conference on Acoustics, Speech and Signal Processing’, vol.97

Torres-Carrasquillo P A, Reynolds D A & Deller J R (1993) “Language Identification using Gaussian Mixture Model Tokenization”, in ‘IEEE International Conference on Acoustics, Speech, and Signal Processing’

Torres-Carrasquillo P A (2002) “Approaches to Language Identification using Gaussian Mixture Models and Shifted Delta Cepstral Features”, in Proceedings of International Conference of Spoken Language Processing’ vol.2, pp.89-92

White L & Mattys S L (2007) “Calibrating Rhythm: First Language and Second Language Studies”, in ‘Journal of Phonetics’ 35(4), pp.501–522

Witten I H & Frank E (2005) “Data Mining: Practical Machine Learning Tools and Techniques”, 2nd Ed, San Francisco: Morgan Kaufman

Wu C H (2006) “Automatic Segmentation and Identification of Mixed-language Speech Using Delta-BIC and LSA-based GMMs”, in ‘IEEE Transactions on Audio, Speech and Language Processing’ 14(1), pp.266-276

Yan Y & Barnard E (1995) “A Comparison of Neural Net and Linear Classifiers as the Pattern Recogniser in Automatic Language Identification”, in ‘Proceedings of International Conference on Neural Networks and Signal Processing’

Yan Y, Barnard E & Cole R (1996) “Development of an Approach to Automatic Language Identification Based on Phone Recognition”, in ‘Computer Speech and Language’ 10, pp.37-54

Zissman M A (1996) “Comparison of Four Approaches to Automatic Language Identification of Telephone Speech”, in ‘IEEE Transactions on Speech and Audio Processing’ no.1, pp.31

Zissman M A & Berkling K M (2001) “Automatic Language Identification”, in ‘Speech Communication’ 35, pp.115-124

Zölzer U [ed.] (2002) “DAFX; Digital Audio Effects”, Chichester: John Wiley & Sons

Appendix I

Project Logs

July 2010

- Background reading into phonetics and linguistics.

September 2010

- Background reading into Automatic Speech Recognition and Language Identification.

October 2010

- Project proposal submitted.
- First Supervisor meeting with Dr. N Collins – discussion of reading and MFCC vectors.
- Investigation of available speech corpora.
- First successful real-time extraction of MFCC vectors on basic corpus – two samples each of English and Japanese.

November 2010

- Storing of MFCC information in a separate file for further processing.
- Began work on Interim Report.
- Successful offline batch extraction of SCMIR features.
- Addition of pitch contour information.
- First version of ARFF file generation.
- Addition of French and German into the system.

December 2010

- First investigations into the WEKA environment.
- First run of Naïve Bayes vs. MLP to assess effectiveness on small language set – 72.5% accuracy obtained on four languages with medium feature set.
- Decision taken to split up comparisons into feature sets to be able to assess feature importance.

January 2011

- Reading into speech rhythm and prosodic cues.
- Investigation into onset extraction.

February 2011

- First version of nPVI function.
- Looked into calling WEKA functions from the command line.
- Addition of remaining languages into the system.

March 2011

- Began work on draft report – template produced.
- SuperCollider patch for WEKA command line functionality produced.
- Amplitudes of speech signal files normalised.

April 2011

- Correction of nPVI function.
- SuperCollider patch for batch creation of ARFF files produced.
- Segmented data generated.
- Batch tests of WEKA run for all extracted datasets.

- Draft Report completed.

May 2011

- Finalisation of project report.

Appendix II

Full Results

I present the full results obtained from the system. First I show the language pair discriminations, the first set from tests performed on unnormalised audio files, which extracted data averaged over the speech signals and contained an incorrect nPVI function. The second set shows results for both averaged data and segmented data, with a corrected nPVI function. For the purposes of space the full results for Naïve Bayes are not shown for language pairs.

Next the language family comparisons are shown and finally the all language comparisons. Both of these sets contain results as above, with the addition of Naïve Bayes for comparison. The results are labelled according to which data they were obtained from and which machine-learning algorithm was used to carry out the classification.

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

engGer	eng	ger	✔
4 MFCC	5	8	65
4 MFCC Tartini	6	9	75
4 MFCC nPVI	4	6	50
13 MFCC	9	8	85
13 MFCC Tartini	9	8	85
41 MFCC	8	9	85
41 MFCC Tartini	8	9	85
41 MFCC Tartini nPVI	8	9	85
			76.88

engDut	eng	dut	✔
4 MFCC	5	5	50
4 MFCC Tartini	8	8	80
4 MFCC nPVI	5	7	60
13 MFCC	8	5	65
13 MFCC Tartini	10	9	95
41 MFCC	9	8	85
41 MFCC Tartini	10	10	100
41 MFCC Tartini nPVI	10	9	95
			78.75

gerIta	ger	ita	✔
4 MFCC	8	8	80
4 MFCC Tartini	9	9	90
4 MFCC nPVI	9	8	85
13 MFCC	10	9	95
13 MFCC Tartini	10	9	95
41 MFCC	10	10	100
41 MFCC Tartini	10	9	95
41 MFCC Tartini nPVI	10	9	95
			91.88

dutSpa	dut	spa	✔
4 MFCC	5	4	45
4 MFCC Tartini	4	5	45
4 MFCC nPVI	7	6	65
13 MFCC	9	9	90
13 MFCC Tartini	9	8	85
41 MFCC	10	9	95
41 MFCC Tartini	10	10	100
41 MFCC Tartini nPVI	10	10	100
			78.13

fraMan	fra	man	✔
4 MFCC	4	5	45
4 MFCC Tartini	3	5	40
4 MFCC nPVI	5	5	50
13 MFCC	5	6	55
13 MFCC Tartini	4	7	55
41 MFCC	8	9	95
41 MFCC Tartini	10	9	95
41 MFCC Tartini nPVI	9	9	90
			65.63

fraKor	fra	kor	✔
4 MFCC	6	8	70
4 MFCC Tartini	7	7	70
4 MFCC nPVI	7	7	0
13 MFCC	8	9	95
13 MFCC Tartini	8	8	80
41 MFCC	9	9	90
41 MFCC Tartini	8	9	85
41 MFCC Tartini nPVI	8	9	95
			73.13

polRus	pol	rus	✔
4 MFCC	5	4	45
4 MFCC Tartini	6	5	55
4 MFCC nPVI	5	7	60
13 MFCC	8	7	75
13 MFCC Tartini	8	7	75
41 MFCC	7	8	75
41 MFCC Tartini	7	6	65
41 MFCC Tartini nPVI	7	6	65
			64.38

polMan	pol	man	✔
4 MFCC	6	7	65
4 MFCC Tartini	5	9	70
4 MFCC nPVI	7	9	80
13 MFCC	7	8	75
13 MFCC Tartini	7	6	65
41 MFCC	7	9	80
41 MFCC Tartini	7	8	75
41 MFCC Tartini nPVI	7	8	75
			73.13

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

gerCze	ger	cze	✔
4 MFCC	8	7	75
4 MFCC Tartini	8	7	75
4 MFCC nPVI	8	6	70
13 MFCC	10	9	95
13 MFCC Tartini	10	9	95
41 MFCC	9	9	90
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	8	85
			84.38

gerPol	ger	pol	✔
4 MFCC	10	5	75
4 MFCC Tartini	6	6	60
4 MFCC nPVI	7	6	65
13 MFCC	10	8	90
13 MFCC Tartini	9	8	85
41 MFCC	10	7	85
41 MFCC Tartini	9	6	75
41 MFCC Tartini nPVI	9	6	75
			76.25

dutJap	dut	jap	✔
4 MFCC	5	5	50
4 MFCC Tartini	10	10	100
4 MFCC nPVI	8	5	65
13 MFCC	8	8	80
13 MFCC Tartini	9	9	90
41 MFCC	9	8	85
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	9	8	85
			80.00

fraJap	fra	jap	✔
4 MFCC	7	8	75
4 MFCC Tartini	7	8	75
4 MFCC nPVI	7	6	65
13 MFCC	7	7	70
13 MFCC Tartini	7	7	70
41 MFCC	9	8	85
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	9	8	85
			76.25

itaJap	ita	jap	✔
4 MFCC	10	8	90
4 MFCC Tartini	10	8	90
4 MFCC nPVI	10	8	90
13 MFCC	8	8	80
13 MFCC Tartini	9	8	85
41 MFCC	10	9	95
41 MFCC Tartini	10	9	95
41 MFCC Tartini nPVI	10	9	95
			90.00

czePol	cze	pol	✔
4 MFCC	10	8	90
4 MFCC Tartini	7	7	70
4 MFCC nPVI	9	8	85
13 MFCC	9	8	85
13 MFCC Tartini	9	8	85
41 MFCC	10	10	100
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	9	90
			86.88

engKor	eng	kor	✔
4 MFCC	6	5	55
4 MFCC Tartini	7	8	75
4 MFCC nPVI	5	5	50
13 MFCC	9	8	85
13 MFCC Tartini	10	9	95
41 MFCC	8	9	85
41 MFCC Tartini	9	10	95
41 MFCC Tartini nPVI	9	9	90
			78.75

engJap	eng	jap	✔
4 MFCC	6	8	70
4 MFCC Tartini	5	8	65
4 MFCC nPVI	7	6	65
13 MFCC	8	7	75
13 MFCC Tartini	9	7	80
41 MFCC	6	7	65
41 MFCC Tartini	6	6	60
41 MFCC Tartini nPVI	6	5	55
			66.88

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

fraCze	fra	cze	✔
4 MFCC	7	10	85
4 MFCC Tartini	10	10	100
4 MFCC nPVI	7	8	75
13 MFCC	9	10	95
13 MFCC Tartini	9	10	95
41 MFCC	10	10	100
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	9	90
			91.25

spaRus	spa	rus	✔
4 MFCC	5	6	55
4 MFCC Tartini	6	1	35
4 MFCC nPVI	6	6	60
13 MFCC	5	6	55
13 MFCC Tartini	6	7	65
41 MFCC	5	6	55
41 MFCC Tartini	7	6	65
41 MFCC Tartini nPVI	6	5	55
			55.63

manKor	man	kor	✔
4 MFCC	9	8	85
4 MFCC Tartini	7	8	75
4 MFCC nPVI	5	5	50
13 MFCC	8	6	70
13 MFCC Tartini	5	6	55
41 MFCC	7	8	75
41 MFCC Tartini	7	7	70
41 MFCC Tartini nPVI	7	9	80
			70.00

manJap	man	jap	✔
4 MFCC	6	4	65
4 MFCC Tartini	8	5	65
4 MFCC nPVI	6	6	60
13 MFCC	7	7	70
13 MFCC Tartini	7	7	70
41 MFCC	6	8	70
41 MFCC Tartini	8	8	80
41 MFCC Tartini nPVI	8	8	80
			70.00

gerSpa	ger	spa	✔
4 MFCC	8	8	80
4 MFCC Tartini	7	7	70
4 MFCC nPVI	8	7	75
13 MFCC	10	8	90
13 MFCC Tartini	10	8	90
41 MFCC	9	5	70
41 MFCC Tartini	6	9	75
41 MFCC Tartini nPVI	9	5	70
			77.50

czeJap	cze	jap	✔
4 MFCC	10	10	100
4 MFCC Tartini	9	9	90
4 MFCC nPVI	10	9	95
13 MFCC	10	10	100
13 MFCC Tartini	10	9	95
41 MFCC	9	10	95
41 MFCC Tartini	9	10	95
41 MFCC Tartini nPVI	8	10	90
			95.00

fraRus	fra	rus	✔
4 MFCC	6	6	60
4 MFCC Tartini	5	4	45
4 MFCC nPVI	7	7	70
13 MFCC	8	7	75
13 MFCC Tartini	6	7	65
41 MFCC	10	9	95
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	8	9	95
			74.38

engRus	eng	rus	✔
4 MFCC	4	6	50
4 MFCC Tartini	5	5	50
4 MFCC nPVI	6	3	45
13 MFCC	6	6	60
13 MFCC Tartini	6	5	55
41 MFCC	7	8	75
41 MFCC Tartini	7	8	75
41 MFCC Tartini nPVI	7	8	75
			60.63

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

engFra	eng	fra	✔
4 MFCC	6	6	60
4 MFCC Tartini	6	5	55
4 MFCC nPVI	4	4	40
13 MFCC	8	7	75
13 MFCC Tartini	6	6	60
41 MFCC	4	7	55
41 MFCC Tartini	5	5	50
41 MFCC Tartini nPVI	4	5	45
			55.00

engSpa	eng	spa	✔
4 MFCC	4	6	50
4 MFCC Tartini	8	6	70
4 MFCC nPVI	6	7	65
13 MFCC	7	5	60
13 MFCC Tartini	8	6	70
41 MFCC	7	5	60
41 MFCC Tartini	8	5	65
41 MFCC Tartini nPVI	8	5	65
			63.13

dutIta	dut	ita	✔
4 MFCC	7	10	85
4 MFCC Tartini	7	8	75
4 MFCC nPVI	7	9	80
13 MFCC	10	9	95
13 MFCC Tartini	9	8	85
41 MFCC	9	9	90
41 MFCC Tartini	10	9	95
41 MFCC Tartini nPVI	10	9	95
			87.50

dutCze	dut	cze	✔
4 MFCC	10	10	100
4 MFCC Tartini	9	10	95
4 MFCC nPVI	10	10	100
13 MFCC	10	10	100
13 MFCC Tartini	8	10	90
41 MFCC	9	10	95
41 MFCC Tartini	9	10	95
41 MFCC Tartini nPVI	9	10	95
			96.25

spaJap	spa	jap	✔
4 MFCC	5	7	60
4 MFCC Tartini	7	7	70
4 MFCC nPVI	5	5	50
13 MFCC	6	7	65
13 MFCC Tartini	6	7	65
41 MFCC	5	8	65
41 MFCC Tartini	5	8	65
41 MFCC Tartini nPVI	5	7	60
			62.50

itaCze	ita	cze	✔
4 MFCC	8	7	75
4 MFCC Tartini	8	8	80
4 MFCC nPVI	7	6	65
13 MFCC	9	10	95
13 MFCC Tartini	10	10	100
41 MFCC	9	8	85
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	9	90
			85.00

polKor	pol	kor	✔
4 MFCC	1	3	20
4 MFCC Tartini	3	4	35
4 MFCC nPVI	4	4	40
13 MFCC	7	5	60
13 MFCC Tartini	7	5	60
41 MFCC	6	6	60
41 MFCC Tartini	7	6	65
41 MFCC Tartini nPVI	8	6	70
			51.25

engIta	eng	ita	✔
4 MFCC	5	3	40
4 MFCC Tartini	5	5	50
4 MFCC nPVI	6	4	50
13 MFCC	7	7	70
13 MFCC Tartini	8	7	75
41 MFCC	6	7	65
41 MFCC Tartini	7	6	65
41 MFCC Tartini nPVI	6	7	65
			60.00

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

gerRus	ger	rus	✔
4 MFCC	5	5	50
4 MFCC Tartini	4	4	40
4 MFCC nPVI	4	1	25
13 MFCC	8	7	75
13 MFCC Tartini	8	6	70
41 MFCC	10	8	90
41 MFCC Tartini	10	8	90
41 MFCC Tartini nPVI	10	8	90
			66.25

gerMan	ger	man	✔
4 MFCC	5	7	60
4 MFCC Tartini	4	5	45
4 MFCC nPVI	5	7	60
13 MFCC	9	9	90
13 MFCC Tartini	9	9	90
41 MFCC	10	10	100
41 MFCC Tartini	10	10	100
41 MFCC Tartini nPVI	10	10	100
			80.63

spaIta	spa	ita	✔
4 MFCC	8	8	80
4 MFCC Tartini	8	7	75
4 MFCC nPVI	7	8	75
13 MFCC	8	9	95
13 MFCC Tartini	9	10	95
41 MFCC	8	9	85
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	9	90
			85.63

spaCze	spa	cze	✔
4 MFCC	9	10	95
4 MFCC Tartini	8	10	90
4 MFCC nPVI	9	10	95
13 MFCC	10	10	100
13 MFCC Tartini	10	10	100
41 MFCC	7	8	75
41 MFCC Tartini	6	8	70
41 MFCC Tartini nPVI	6	8	70
			86.88

polJap	pol	jap	✔
4 MFCC	3	7	50
4 MFCC Tartini	4	6	50
4 MFCC nPVI	5	7	60
13 MFCC	8	7	75
13 MFCC Tartini	8	7	75
41 MFCC	8	8	80
41 MFCC Tartini	8	7	75
41 MFCC Tartini nPVI	8	7	75
			67.50

rusMan	rus	man	✔
4 MFCC	5	8	65
4 MFCC Tartini	7	5	60
4 MFCC nPVI	8	5	65
13 MFCC	6	7	65
13 MFCC Tartini	7	7	70
41 MFCC	7	7	70
41 MFCC Tartini	6	7	65
41 MFCC Tartini nPVI	6	7	65
			65.63

gerKor	ger	kor	✔
4 MFCC	8	7	75
4 MFCC Tartini	8	7	75
4 MFCC nPVI	8	2	50
13 MFCC	9	7	80
13 MFCC Tartini	9	8	85
41 MFCC	10	9	95
41 MFCC Tartini	10	9	95
41 MFCC Tartini nPVI	10	9	95
			81.25

gerJap	ger	jap	✔
4 MFCC	7	7	70
4 MFCC Tartini	8	6	70
4 MFCC nPVI	6	7	65
13 MFCC	10	9	95
13 MFCC Tartini	10	8	90
41 MFCC	10	8	90
41 MFCC Tartini	10	8	90
41 MFCC Tartini nPVI	10	8	90
			82.50

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

spaMan	spa	man	✔
4 MFCC	6	8	70
4 MFCC Tartini	5	5	50
4 MFCC nPVI	6	5	55
13 MFCC	4	6	50
13 MFCC Tartini	4	4	40
41 MFCC	3	6	45
41 MFCC Tartini	4	6	50
41 MFCC Tartini nPVI	4	6	50
			51.25

spaKor	spa	kor	✔
4 MFCC	4	6	50
4 MFCC Tartini	2	4	30
4 MFCC nPVI	7	3	50
13 MFCC	6	8	70
13 MFCC Tartini	7	8	75
41 MFCC	6	10	80
41 MFCC Tartini	5	9	70
41 MFCC Tartini nPVI	4	9	65
			61.25

korJap	kor	jap	✔
4 MFCC	6	4	50
4 MFCC Tartini	6	4	50
4 MFCC nPVI	4	6	55
13 MFCC	9	9	90
13 MFCC Tartini	9	8	85
41 MFCC	10	8	90
41 MFCC Tartini	10	8	90
41 MFCC Tartini nPVI	10	8	90
			75.00

engMan	eng	man	✔
4 MFCC	5	5	50
4 MFCC Tartini	4	7	55
4 MFCC nPVI	4	7	55
13 MFCC	6	7	65
13 MFCC Tartini	7	7	70
41 MFCC	7	7	70
41 MFCC Tartini	8	8	80
41 MFCC Tartini nPVI	8	8	80
			65.63

dutKor	dut	kor	✔
4 MFCC	5	3	40
4 MFCC Tartini	7	6	65
4 MFCC nPVI	7	5	60
13 MFCC	10	10	100
13 MFCC Tartini	10	10	100
41 MFCC	10	9	95
41 MFCC Tartini	10	10	100
41 MFCC Tartini nPVI	10	10	100
			82.50

fraIta	fra	ita	✔
4 MFCC	5	4	45
4 MFCC Tartini	5	6	55
4 MFCC nPVI	6	5	55
13 MFCC	6	6	60
13 MFCC Tartini	7	4	55
41 MFCC	7	7	70
41 MFCC Tartini	7	7	70
41 MFCC Tartini nPVI	7	8	75
			60.63

itaKor	ita	kor	✔
4 MFCC	10	9	95
4 MFCC Tartini	10	9	95
4 MFCC nPVI	10	9	95
13 MFCC	9	9	90
13 MFCC Tartini	10	10	100
41 MFCC	9	10	95
41 MFCC Tartini	9	10	95
41 MFCC Tartini nPVI	9	10	95
			95.00

czeKor	cze	kor	✔
4 MFCC	9	8	85
4 MFCC Tartini	10	8	90
4 MFCC nPVI	9	8	85
13 MFCC	9	8	85
13 MFCC Tartini	10	10	100
41 MFCC	10	10	100
41 MFCC Tartini	10	10	100
41 MFCC Tartini nPVI	10	10	100
			93.13

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

gerDut	ger	dut	✔
4 MFCC	8	5	65
4 MFCC Tartini	9	9	90
4 MFCC nPVI	5	5	50
13 MFCC	10	8	90
13 MFCC Tartini	9	8	85
41 MFCC	9	7	80
41 MFCC Tartini	10	8	90
41 MFCC Tartini nPVI	9	8	85
			79.38

gerFra	ger	fra	✔
4 MFCC	8	4	60
4 MFCC Tartini	7	6	65
4 MFCC nPVI	7	4	55
13 MFCC	9	10	95
13 MFCC Tartini	9	10	95
41 MFCC	9	7	80
41 MFCC Tartini	9	7	80
41 MFCC Tartini nPVI	9	7	80
			76.25

dutPol	dut	pol	✔
4 MFCC	5	2	35
4 MFCC Tartini	7	6	65
4 MFCC nPVI	4	7	55
13 MFCC	6	6	60
13 MFCC Tartini	7	6	65
41 MFCC	8	7	75
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	8	85
			66.25

fraPol	fra	pol	✔
4 MFCC	6	7	65
4 MFCC Tartini	8	6	70
4 MFCC nPVI	7	7	70
13 MFCC	6	6	60
13 MFCC Tartini	7	7	70
41 MFCC	7	8	75
41 MFCC Tartini	7	7	70
41 MFCC Tartini nPVI	7	7	70
			68.75

itaPol	ita	pol	✔
4 MFCC	9	8	85
4 MFCC Tartini	9	8	85
4 MFCC nPVI	9	7	80
13 MFCC	7	7	70
13 MFCC Tartini	7	8	75
41 MFCC	8	9	85
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	8	9	85
			81.88

itaRus	ita	rus	✔
4 MFCC	6	7	65
4 MFCC Tartini	8	7	75
4 MFCC nPVI	5	7	60
13 MFCC	8	8	80
13 MFCC Tartini	8	8	80
41 MFCC	8	10	90
41 MFCC Tartini	7	10	85
41 MFCC Tartini nPVI	8	10	90
			78.13

engCze	eng	cze	✔
4 MFCC	8	9	85
4 MFCC Tartini	7	9	80
4 MFCC nPVI	8	8	80
13 MFCC	10	10	100
13 MFCC Tartini	10	10	100
41 MFCC	9	9	90
41 MFCC Tartini	0	9	5
41 MFCC Tartini nPVI	10	9	95
			79.38

engPol	eng	pol	✔
4 MFCC	7	5	60
4 MFCC Tartini	5	5	50
4 MFCC nPVI	8	5	65
13 MFCC	8	8	80
13 MFCC Tartini	10	7	85
41 MFCC	8	8	80
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	9	8	85
			73.75

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

dutRus	dut	rus	✔
4 MFCC	7	5	60
4 MFCC Tartini	9	8	85
4 MFCC nPVI	5	5	50
13 MFCC	6	4	50
13 MFCC Tartini	7	6	65
41 MFCC	5	7	60
41 MFCC Tartini	9	7	80
41 MFCC Tartini nPVI	9	7	80
			66.25

dutMan	dut	man	✔
4 MFCC	8	6	70
4 MFCC Tartini	8	7	75
4 MFCC nPVI	7	7	70
13 MFCC	9	9	90
13 MFCC Tartini	9	8	95
41 MFCC	8	9	85
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	8	10	90
			82.50

spaPol	spa	pol	✔
4 MFCC	5	5	50
4 MFCC Tartini	5	5	50
4 MFCC nPVI	7	6	65
13 MFCC	6	9	75
13 MFCC Tartini	5	9	70
41 MFCC	6	7	65
41 MFCC Tartini	6	7	65
41 MFCC Tartini nPVI	6	8	70
			63.75

itaMan	ita	man	✔
4 MFCC	6	7	65
4 MFCC Tartini	6	7	65
4 MFCC nPVI	6	8	70
13 MFCC	7	7	70
13 MFCC Tartini	6	9	75
41 MFCC	6	8	70
41 MFCC Tartini	6	8	70
41 MFCC Tartini nPVI	6	8	70
			69.38

rusKor	rus	kor	✔
4 MFCC	6	7	65
4 MFCC Tartini	8	7	75
4 MFCC nPVI	6	6	60
13 MFCC	7	9	80
13 MFCC Tartini	9	9	90
41 MFCC	9	9	90
41 MFCC Tartini	9	8	85
41 MFCC Tartini nPVI	8	10	90
			79.38

rusJap	rus	jap	✔
4 MFCC	6	7	65
4 MFCC Tartini	5	5	50
4 MFCC nPVI	5	8	65
13 MFCC	4	5	45
13 MFCC Tartini	4	5	45
41 MFCC	6	7	65
41 MFCC Tartini	5	7	60
41 MFCC Tartini nPVI	6	6	60
			56.88

dutFra	dut	fra	✔
4 MFCC	8	5	65
4 MFCC Tartini	8	8	80
4 MFCC nPVI	8	5	65
13 MFCC	9	8	95
13 MFCC Tartini	7	9	80
41 MFCC	9	9	90
41 MFCC Tartini	9	9	90
41 MFCC Tartini nPVI	9	8	85
			81.25

fraSpa	fra	spa	✔
4 MFCC	5	3	40
4 MFCC Tartini	7	6	65
4 MFCC nPVI	5	6	55
13 MFCC	8	7	75
13 MFCC Tartini	7	7	70
41 MFCC	6	7	65
41 MFCC Tartini	7	7	70
41 MFCC Tartini nPVI	7	7	70
			63.75

Language Pairs, Unnormalised Audio, Average over File, Incorrect nPVI Function

czeRus	cze	rus	✔
4 MFCC	7	7	70
4 MFCC Tartini	7	8	75
4 MFCC nPVI	8	7	75
13 MFCC	7	7	70
13 MFCC Tartini	10	7	85
41 MFCC	6	8	70
41 MFCC Tartini	7	8	75
41 MFCC Tartini nPVI	7	9	80
			75.00

czeMan	cze	man	✔
4 MFCC	10	9	95
4 MFCC Tartini	10	8	90
4 MFCC nPVI	10	9	95
13 MFCC	9	9	90
13 MFCC Tartini	9	10	95
41 MFCC	8	9	95
41 MFCC Tartini	8	9	85
41 MFCC Tartini nPVI	8	9	85
			91.25

MLP - Language Pairs - Normalised Audio

Average over File

engGer	eng	ger	✔
4 MFCC	6	8	70.00
4 MFCC Tartini	6	7	65.00
4 MFCC nPVI	5	4	45.00
13 MFCC	8	8	80.00
13 MFCC Tartini	9	8	85.00
41 MFCC	8	9	85.00
41 MFCC Tartini	8	9	85.00
41 MFCC Tartini nPVI	7	10	85.00
			75.00

gerIta	ger	ita	✔
4 MFCC	8	8	80.00
4 MFCC Tartini	9	9	90.00
4 MFCC nPVI	7	8	75.00
13 MFCC	10	9	95.00
13 MFCC Tartini	10	9	95.00
41 MFCC	10	10	100.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	10	9	95.00
			90.63

fraMan	fra	man	✔
4 MFCC	4	3	35.00
4 MFCC Tartini	4	3	35.00
4 MFCC nPVI	7	8	75.00
13 MFCC	6	6	60.00
13 MFCC Tartini	4	7	55.00
41 MFCC	8	9	85.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	8	9	85.00
			65.00

polRus	pol	rus	✔
4 MFCC	4	6	50.00
4 MFCC Tartini	4	2	30.00
4 MFCC nPVI	4	5	45.00
13 MFCC	8	7	75.00
13 MFCC Tartini	7	7	70.00
41 MFCC	7	7	70.00
41 MFCC Tartini	7	6	65.00
41 MFCC Tartini nPVI	7	6	65.00
			58.75

5s Segments

Instances ➤	68	66	
engGer	eng	ger	✔
4 MFCC	47	41	65.67
4 MFCC Tartini	50	51	75.37
4 MFCC nPVI			
13 MFCC	61	63	92.54
13 MFCC Tartini	62	61	91.79
41 MFCC	66	66	98.51
41 MFCC Tartini	67	66	99.25
41 MFCC Tartini nPVI			
			87.19

Instances ➤	66	66	
gerIta	ger	ita	✔
4 MFCC	53	43	72.73
4 MFCC Tartini	53	59	69.70
4 MFCC nPVI			
13 MFCC	62	59	91.67
13 MFCC Tartini	60	60	90.91
41 MFCC	62	63	94.70
41 MFCC Tartini	64	61	94.70
41 MFCC Tartini nPVI			
			85.73

Instances ➤	67	66	
fraMan	fra	man	✔
4 MFCC	46	48	70.68
4 MFCC Tartini	55	54	81.95
4 MFCC nPVI			
13 MFCC	61	62	92.48
13 MFCC Tartini	60	62	91.73
41 MFCC	67	66	100.00
41 MFCC Tartini	66	66	99.25
41 MFCC Tartini nPVI			
			89.35

Instances ➤	69	68	
polRus	pol	rus	✔
4 MFCC	50	51	73.72
4 MFCC Tartini	53	49	74.45
4 MFCC nPVI			
13 MFCC	61	59	87.59
13 MFCC Tartini	65	57	89.05
41 MFCC	68	66	97.81
41 MFCC Tartini	68	64	96.35
41 MFCC Tartini nPVI			
			86.50

MLP - Language Pairs - Normalised Audio

Average over File

gerCze	ger	cze	✔
4 MFCC	8	9	85.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	9	8	85.00
13 MFCC	10	10	100.00
13 MFCC Tartini	10	9	95.00
41 MFCC	9	9	90.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	10	9	95.00
			90.00

dutJap	dut	jap	✔
4 MFCC	5	7	60.00
4 MFCC Tartini	9	10	95.00
4 MFCC nPVI	7	6	65.00
13 MFCC	9	8	85.00
13 MFCC Tartini	9	8	85.00
41 MFCC	9	9	90.00
41 MFCC Tartini	8	9	85.00
41 MFCC Tartini nPVI	9	9	90.00
			81.88

itaJap	ita	jap	✔
4 MFCC	10	8	90.00
4 MFCC Tartini	10	7	85.00
4 MFCC nPVI	10	8	90.00
13 MFCC	8	8	80.00
13 MFCC Tartini	9	7	80.00
41 MFCC	10	10	100.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	9	9	90.00
			88.75

engKor	eng	kor	✔
4 MFCC	6	6	60.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	7	8	75.00
13 MFCC	9	8	85.00
13 MFCC Tartini	9	9	90.00
41 MFCC	8	9	85.00
41 MFCC Tartini	9	10	95.00
41 MFCC Tartini nPVI	9	10	95.00
			82.50

5s Segments

Instances ➤	66	65	
gerCze	ger	cze	✔
4 MFCC	55	40	72.52
4 MFCC Tartini	54	55	83.21
4 MFCC nPVI			
13 MFCC	64	56	91.60
13 MFCC Tartini	61	54	87.79
41 MFCC	66	64	99.24
41 MFCC Tartini	66	63	98.47
41 MFCC Tartini nPVI			
			88.80

Instances ➤	67	66	
dutJap	dut	jap	✔
4 MFCC	55	46	75.94
4 MFCC Tartini	56	47	77.44
4 MFCC nPVI			
13 MFCC	63	62	93.99
13 MFCC Tartini	62	62	93.23
41 MFCC	66	65	98.50
41 MFCC Tartini	66	65	98.50
41 MFCC Tartini nPVI			
			89.60

Instances ➤	66	66	
itaJap	ita	jap	✔
4 MFCC	53	50	78.03
4 MFCC Tartini	51	51	77.27
4 MFCC nPVI			
13 MFCC	55	60	87.12
13 MFCC Tartini	57	59	87.88
41 MFCC	65	66	99.24
41 MFCC Tartini	65	66	99.24
41 MFCC Tartini nPVI			
			88.13

Instances ➤	68	67	
engKor	eng	kor	✔
4 MFCC	45	50	70.37
4 MFCC Tartini	59	61	88.89
4 MFCC nPVI			
13 MFCC	64	65	95.56
13 MFCC Tartini	66	67	98.52
41 MFCC	66	66	97.78
41 MFCC Tartini	67	66	98.52
41 MFCC Tartini nPVI			
			91.60

MLP - Language Pairs - Normalised Audio

Average over File

fraCze	fra	cze	✔
4 MFCC	9	10	95.00
4 MFCC Tartini	10	10	100.00
4 MFCC nPVI	9	10	95.00
13 MFCC	9	10	95.00
13 MFCC Tartini	9	10	95.00
41 MFCC	10	10	100.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	9	9	90.00
			95.00

manKor	man	kor	✔
4 MFCC	8	7	75.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	6	6	60.00
13 MFCC	6	6	60.00
13 MFCC Tartini	5	6	55.00
41 MFCC	7	8	75.00
41 MFCC Tartini	7	8	75.00
41 MFCC Tartini nPVI	8	10	90.00
			70.63

gerSpa	ger	spa	✔
4 MFCC	9	8	85.00
4 MFCC Tartini	7	7	70.00
4 MFCC nPVI	8	6	70.00
13 MFCC	10	8	90.00
13 MFCC Tartini	10	8	90.00
41 MFCC	10	6	80.00
41 MFCC Tartini	9	5	70.00
41 MFCC Tartini nPVI	9	5	70.00
			78.13

fraRus	fra	rus	✔
4 MFCC	7	7	70.00
4 MFCC Tartini	5	5	50.00
4 MFCC nPVI	7	4	55.00
13 MFCC	7	8	75.00
13 MFCC Tartini	6	7	65.00
41 MFCC	10	9	95.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	9	9	90.00
			73.75

5s Segments

Instances ➤	67	65	
fraCze	fra	cze	✔
4 MFCC	56	52	81.82
4 MFCC Tartini	58	55	85.61
4 MFCC nPVI			
13 MFCC	60	56	87.88
13 MFCC Tartini	60	57	88.64
41 MFCC	65	63	96.97
41 MFCC Tartini	66	64	98.48
41 MFCC Tartini nPVI			
			89.90

Instances ➤	66	67	
manKor	man	kor	✔
4 MFCC	56	51	80.45
4 MFCC Tartini	59	61	90.23
4 MFCC nPVI			
13 MFCC	58	64	91.73
13 MFCC Tartini	61	65	94.74
41 MFCC	65	66	98.50
41 MFCC Tartini	64	65	96.99
41 MFCC Tartini nPVI			
			92.11

Instances ➤	66	66	
gerSpa	ger	spa	✔
4 MFCC	42	51	70.45
4 MFCC Tartini	54	55	82.58
4 MFCC nPVI			
13 MFCC	62	63	94.70
13 MFCC Tartini	64	63	96.21
41 MFCC	65	63	96.97
41 MFCC Tartini	63	61	93.94
41 MFCC Tartini nPVI			
			89.14

Instances ➤	67	68	
fraRus	fra	rus	✔
4 MFCC	44	56	74.07
4 MFCC Tartini	51	53	77.04
4 MFCC nPVI			
13 MFCC	58	62	88.89
13 MFCC Tartini	54	59	83.70
41 MFCC	66	68	99.26
41 MFCC Tartini	65	68	98.52
41 MFCC Tartini nPVI			
			86.91

MLP - Language Pairs - Normalised Audio

Average over File

engDut	eng	dut	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	8	8	80.00
4 MFCC nPVI	6	6	60.00
13 MFCC	7	5	60.00
13 MFCC Tartini	10	8	90.00
41 MFCC	9	9	90.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	9	9	90.00
			76.25

dutSpa	dut	spa	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	5	5	50.00
4 MFCC nPVI	7	6	65.00
13 MFCC	9	8	85.00
13 MFCC Tartini	9	8	85.00
41 MFCC	10	9	95.00
41 MFCC Tartini	10	10	100.00
41 MFCC Tartini nPVI	10	9	100.00
			78.75

fraKor	fra	kor	✔
4 MFCC	6	8	70.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	7	7	70.00
13 MFCC	8	9	85.00
13 MFCC Tartini	8	8	80.00
41 MFCC	9	9	90.00
41 MFCC Tartini	8	9	85.00
41 MFCC Tartini nPVI	8	9	95.00
			80.00

polMan	pol	man	✔
4 MFCC	6	6	60.00
4 MFCC Tartini	5	8	65.00
4 MFCC nPVI	4	8	60.00
13 MFCC	7	7	70.00
13 MFCC Tartini	8	6	70.00
41 MFCC	6	8	70.00
41 MFCC Tartini	6	7	65.00
41 MFCC Tartini nPVI	6	7	65.00
			65.63

5s Segments

Instances ➤	68	67	
engDut	eng	dut	✔
4 MFCC	49	55	77.04
4 MFCC Tartini	64	62	93.33
4 MFCC nPVI			
13 MFCC	66	60	93.33
13 MFCC Tartini	65	63	94.81
41 MFCC	67	66	98.52
41 MFCC Tartini	68	65	98.52
41 MFCC Tartini nPVI			
			92.59

Instances ➤	67	66	
dutSpa	dut	spa	✔
4 MFCC	50	44	70.68
4 MFCC Tartini	52	49	75.94
4 MFCC nPVI			
13 MFCC	59	57	87.22
13 MFCC Tartini	58	59	87.97
41 MFCC	65	65	97.74
41 MFCC Tartini	65	64	96.99
41 MFCC Tartini nPVI			
			86.09

Instances ➤	67	67	
fraKor	fra	kor	✔
4 MFCC	52	54	79.10
4 MFCC Tartini	58	56	85.07
4 MFCC nPVI			
13 MFCC	56	64	89.55
13 MFCC Tartini	61	64	93.28
41 MFCC	64	67	97.76
41 MFCC Tartini	65	67	98.51
41 MFCC Tartini nPVI			
			90.55

Instances ➤	69	66	
polMan	pol	man	✔
4 MFCC	51	44	70.37
4 MFCC Tartini	58	53	82.22
4 MFCC nPVI			
13 MFCC	65	61	93.33
13 MFCC Tartini	62	57	88.15
41 MFCC	67	64	97.04
41 MFCC Tartini	67	64	97.04
41 MFCC Tartini nPVI			
			88.02

MLP - Language Pairs - Normalised Audio

Average over File

gerPol	ger	pol	✔
4 MFCC	10	5	75.00
4 MFCC Tartini	6	5	55.00
4 MFCC nPVI	8	6	70.00
13 MFCC	10	8	90.00
13 MFCC Tartini	9	8	85.00
41 MFCC	10	7	85.00
41 MFCC Tartini	9	7	80.00
41 MFCC Tartini nPVI	10	6	80.00
			77.50

fraJap	fra	jap	✔
4 MFCC	7	8	75.00
4 MFCC Tartini	6	8	70.00
4 MFCC nPVI	7	7	70.00
13 MFCC	7	8	75.00
13 MFCC Tartini	7	7	70.00
41 MFCC	9	8	85.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	8	9	85.00
			76.88

czePol	cze	pol	✔
4 MFCC	9	8	85.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	10	8	90.00
13 MFCC	9	8	85.00
13 MFCC Tartini	9	7	80.00
41 MFCC	9	10	95.00
41 MFCC Tartini	9	10	95.00
41 MFCC Tartini nPVI	9	10	95.00
			86.25

engJap	eng	jap	✔
4 MFCC	6	8	70.00
4 MFCC Tartini	6	7	65.00
4 MFCC nPVI	7	5	60.00
13 MFCC	8	7	75.00
13 MFCC Tartini	9	7	80.00
41 MFCC	5	7	60.00
41 MFCC Tartini	5	7	60.00
41 MFCC Tartini nPVI	5	7	60.00
			66.25

5s Segments

Instances ➤	66	69	
gerPol	ger	pol	✔
4 MFCC	51	59	81.48
4 MFCC Tartini	53	56	80.74
4 MFCC nPVI			
13 MFCC	62	64	93.33
13 MFCC Tartini	60	63	91.11
41 MFCC	65	67	97.78
41 MFCC Tartini	65	66	97.04
41 MFCC Tartini nPVI			
			90.25

Instances ➤	67	66	
fraJap	fra	jap	✔
4 MFCC	53	48	75.94
4 MFCC Tartini	54	47	75.94
4 MFCC nPVI			
13 MFCC	61	61	91.73
13 MFCC Tartini	61	62	92.48
41 MFCC	64	64	96.24
41 MFCC Tartini	64	65	96.99
41 MFCC Tartini nPVI			
			88.22

Instances ➤	65	69	
czePol	cze	pol	✔
4 MFCC	54	52	79.10
4 MFCC Tartini	52	57	81.34
4 MFCC nPVI			
13 MFCC	57	61	88.06
13 MFCC Tartini	58	63	90.30
41 MFCC	63	68	97.76
41 MFCC Tartini	61	68	96.27
41 MFCC Tartini nPVI			
			88.81

Instances ➤	68	66	
engJap	eng	jap	✔
4 MFCC	45	41	64.18
4 MFCC Tartini	47	49	71.64
4 MFCC nPVI			
13 MFCC	60	63	91.79
13 MFCC Tartini	63	62	93.28
41 MFCC	64	65	96.27
41 MFCC Tartini	65	65	97.01
41 MFCC Tartini nPVI			
			85.70

MLP - Language Pairs - Normalised Audio

Average over File

spaRus	spa	rus	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	5	4	45.00
4 MFCC nPVI	7	6	65.00
13 MFCC	5	6	55.00
13 MFCC Tartini	6	7	65.00
41 MFCC	5	6	55.00
41 MFCC Tartini	6	5	55.00
41 MFCC Tartini nPVI	6	5	55.00
			55.63

manJap	man	jap	✔
4 MFCC	7	7	70.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	8	5	65.00
13 MFCC	6	7	65.00
13 MFCC Tartini	7	7	70.00
41 MFCC	6	8	70.00
41 MFCC Tartini	6	8	70.00
41 MFCC Tartini nPVI	5	8	65.00
			67.50

czeJap	cze	jap	✔
4 MFCC	10	10	100.00
4 MFCC Tartini	9	9	90.00
4 MFCC nPVI	10	10	100.00
13 MFCC	10	10	100.00
13 MFCC Tartini	10	8	90.00
41 MFCC	8	10	90.00
41 MFCC Tartini	8	10	90.00
41 MFCC Tartini nPVI	8	10	90.00
			93.75

engRus	eng	rus	✔
4 MFCC	5	4	45.00
4 MFCC Tartini	4	5	45.00
4 MFCC nPVI	4	2	30.00
13 MFCC	6	5	65.00
13 MFCC Tartini	5	4	45.00
41 MFCC	7	9	80.00
41 MFCC Tartini	7	9	80.00
41 MFCC Tartini nPVI	7	8	75.00
			58.13

5s Segments

Instances ➤	66	68	
spaRus	spa	rus	✔
4 MFCC	43	49	68.66
4 MFCC Tartini	41	47	65.67
4 MFCC nPVI			
13 MFCC	62	57	88.81
13 MFCC Tartini	61	55	86.57
41 MFCC	66	65	97.01
41 MFCC Tartini	65	65	97.01
41 MFCC Tartini nPVI			
			83.96

Instances ➤	66	66	
manJap	man	jap	✔
4 MFCC	48	52	75.76
4 MFCC Tartini	53	49	77.27
4 MFCC nPVI			
13 MFCC	62	63	94.70
13 MFCC Tartini	62	65	96.21
41 MFCC	64	65	97.73
41 MFCC Tartini	64	64	96.97
41 MFCC Tartini nPVI			
			89.77

Instances ➤	62	66	
czeJap	cze	jap	✔
4 MFCC	51	51	77.86
4 MFCC Tartini	50	52	77.86
4 MFCC nPVI			
13 MFCC	60	63	93.89
13 MFCC Tartini	62	65	96.95
41 MFCC	62	66	97.71
41 MFCC Tartini	62	66	97.71
41 MFCC Tartini nPVI			
			90.33

Instances ➤	68	68	
engRus	eng	rus	✔
4 MFCC	40	49	65.44
4 MFCC Tartini	58	56	83.82
4 MFCC nPVI			
13 MFCC	63	65	87.50
13 MFCC Tartini	67	60	93.38
41 MFCC	68	66	98.53
41 MFCC Tartini	68	66	98.53
41 MFCC Tartini nPVI			
			87.87

MLP - Language Pairs - Normalised Audio

Average over File

engFra	eng	fra	✔
4 MFCC	6	6	60.00
4 MFCC Tartini	6	5	55.00
4 MFCC nPVI	7	6	65.00
13 MFCC	8	6	70.00
13 MFCC Tartini	7	6	65.00
41 MFCC	5	7	60.00
41 MFCC Tartini	5	5	50.00
41 MFCC Tartini nPVI	5	5	50.00
			59.38

dutIta	dut	ita	✔
4 MFCC	7	9	80.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	9	7	80.00
13 MFCC	10	9	95.00
13 MFCC Tartini	10	9	95.00
41 MFCC	10	9	95.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	10	9	95.00
			88.75

spaJap	spa	jap	✔
4 MFCC	2	7	45.00
4 MFCC Tartini	6	6	60.00
4 MFCC nPVI	4	5	45.00
13 MFCC	6	7	65.00
13 MFCC Tartini	6	7	65.00
41 MFCC	6	8	70.00
41 MFCC Tartini	5	8	65.00
41 MFCC Tartini nPVI	5	8	65.00
			60.00

polKor	pol	kor	✔
4 MFCC	1	2	15.00
4 MFCC Tartini	2	4	30.00
4 MFCC nPVI	4	6	45.00
13 MFCC	7	5	60.00
13 MFCC Tartini	7	5	60.00
41 MFCC	6	6	60.00
41 MFCC Tartini	6	7	65.00
41 MFCC Tartini nPVI	6	7	65.00
			50.00

5s Segments

engFra	eng	fra	✔
Instances ➤	68	67	
4 MFCC	35	60	70.37
4 MFCC Tartini	55	55	81.48
4 MFCC nPVI			
13 MFCC	63	56	88.15
13 MFCC Tartini	61	58	88.15
41 MFCC	66	64	96.30
41 MFCC Tartini	66	64	96.30
41 MFCC Tartini nPVI			
			86.79

dutIta	dut	ita	✔
Instances ➤	67	66	
4 MFCC	56	54	82.71
4 MFCC Tartini	53	48	75.94
4 MFCC nPVI			
13 MFCC	60	61	90.98
13 MFCC Tartini	57	58	86.47
41 MFCC	67	65	99.25
41 MFCC Tartini	66	65	98.50
41 MFCC Tartini nPVI			
			88.97

spaJap	spa	jap	✔
Instances ➤	66	66	
4 MFCC	41	34	56.82
4 MFCC Tartini	46	36	62.12
4 MFCC nPVI			
13 MFCC	58	63	91.67
13 MFCC Tartini	59	61	90.91
41 MFCC	64	66	98.48
41 MFCC Tartini	64	66	98.48
41 MFCC Tartini nPVI			
			83.08

polKor	pol	kor	✔
Instances ➤	69	67	
4 MFCC	47	30	56.62
4 MFCC Tartini	48	48	70.59
4 MFCC nPVI			
13 MFCC	65	63	94.12
13 MFCC Tartini	61	67	91.91
41 MFCC	65	64	94.85
41 MFCC Tartini	66	65	69.32
41 MFCC Tartini nPVI			
			79.57

MLP - Language Pairs - Normalised Audio

Average over File

gerRus	ger	rus	✔
4 MFCC	5	4	45.00
4 MFCC Tartini	4	6	50.00
4 MFCC nPVI	5	4	45.00
13 MFCC	8	7	75.00
13 MFCC Tartini	8	7	75.00
41 MFCC	10	8	90.00
41 MFCC Tartini	10	8	90.00
41 MFCC Tartini nPVI	10	7	85.00
			69.38

spaIta	spa	ita	✔
4 MFCC	8	8	80.00
4 MFCC Tartini	9	7	80.00
4 MFCC nPVI	7	8	75.00
13 MFCC	8	9	95.00
13 MFCC Tartini	9	10	95.00
41 MFCC	9	10	95.00
41 MFCC Tartini	10	10	100.00
41 MFCC Tartini nPVI	10	9	95.00
			89.38

polJap	pol	jap	✔
4 MFCC	4	5	45.00
4 MFCC Tartini	5	6	55.00
4 MFCC nPVI	5	8	65.00
13 MFCC	8	8	80.00
13 MFCC Tartini	8	7	75.00
41 MFCC	8	8	80.00
41 MFCC Tartini	8	8	80.00
41 MFCC Tartini nPVI	8	8	80.00
			70.00

gerKor	ger	kor	✔
4 MFCC	8	8	80.00
4 MFCC Tartini	8	7	75.00
4 MFCC nPVI	6	4	50.00
13 MFCC	9	7	80.00
13 MFCC Tartini	9	9	90.00
41 MFCC	10	10	100.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	10	9	95.00
			83.13

5s Segments

gerRus	ger	rus	✔
Instances ➤	66	68	
4 MFCC	53	47	74.63
4 MFCC Tartini	50	54	77.61
4 MFCC nPVI			
13 MFCC	61	60	90.30
13 MFCC Tartini	61	60	90.30
41 MFCC	65	67	98.51
41 MFCC Tartini	65	66	97.76
41 MFCC Tartini nPVI			
			88.18

spaIta	spa	ita	✔
Instances ➤	66	66	
4 MFCC	57	54	84.09
4 MFCC Tartini	56	53	82.58
4 MFCC nPVI			
13 MFCC	62	62	93.94
13 MFCC Tartini	60	61	91.67
41 MFCC	66	65	99.24
41 MFCC Tartini	66	64	98.48
41 MFCC Tartini nPVI			
			91.67

polJap	pol	jap	✔
Instances ➤	69	66	
4 MFCC	42	42	62.22
4 MFCC Tartini	50	41	67.41
4 MFCC nPVI			
13 MFCC	61	57	87.41
13 MFCC Tartini	65	62	94.07
41 MFCC	64	65	95.56
41 MFCC Tartini	64	65	95.56
41 MFCC Tartini nPVI			
			83.70

gerKor	ger	kor	✔
Instances ➤	66	67	
4 MFCC	51	49	75.19
4 MFCC Tartini	58	63	90.98
4 MFCC nPVI			
13 MFCC	61	63	93.23
13 MFCC Tartini	60	61	90.98
41 MFCC	66	66	99.25
41 MFCC Tartini	66	67	100.00
41 MFCC Tartini nPVI			
			91.60

MLP - Language Pairs - Normalised Audio

Average over File

spaMan	spa	man	✔
4 MFCC	6	8	70.00
4 MFCC Tartini	6	5	50.00
4 MFCC nPVI	5	6	55.00
13 MFCC	4	6	50.00
13 MFCC Tartini	4	4	40.00
41 MFCC	2	6	40.00
41 MFCC Tartini	4	5	45.00
41 MFCC Tartini nPVI	5	6	55.00
			50.63

korJap	kor	jap	✔
4 MFCC	4	4	40.00
4 MFCC Tartini	6	5	55.00
4 MFCC nPVI	6	6	60.00
13 MFCC	8	8	80.00
13 MFCC Tartini	7	8	75.00
41 MFCC	10	8	90.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	8	85.00
			71.25

dutKor	dut	kor	✔
4 MFCC	4	4	40.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	8	7	75.00
13 MFCC	10	10	100.00
13 MFCC Tartini	10	10	100.00
41 MFCC	10	9	95.00
41 MFCC Tartini	10	10	100.00
41 MFCC Tartini nPVI	9	10	95.00
			83.75

itaKor	ita	kor	✔
4 MFCC	10	9	95.00
4 MFCC Tartini	10	9	95.00
4 MFCC nPVI	9	9	90.00
13 MFCC	9	9	90.00
13 MFCC Tartini	10	10	100.00
41 MFCC	9	10	95.00
41 MFCC Tartini	9	10	95.00
41 MFCC Tartini nPVI	9	10	95.00
			94.38

5s Segments

Instances ➤	66	66	
spaMan	spa	man	✔
4 MFCC	47	52	75.00
4 MFCC Tartini	55	55	83.33
4 MFCC nPVI			
13 MFCC	60	56	87.88
13 MFCC Tartini	61	61	92.42
41 MFCC	64	64	96.97
41 MFCC Tartini	64	63	96.21
41 MFCC Tartini nPVI			
			88.64

Instances ➤	67	66	
korJap	kor	jap	✔
4 MFCC	39	40	59.40
4 MFCC Tartini	52	40	69.17
4 MFCC nPVI			
13 MFCC	67	61	96.24
13 MFCC Tartini	65	60	93.99
41 MFCC	67	65	99.25
41 MFCC Tartini	67	65	99.25
41 MFCC Tartini nPVI			
			86.22

Instances ➤	67	67	
dutKor	dut	kor	✔
4 MFCC	46	48	70.15
4 MFCC Tartini	47	44	67.91
4 MFCC nPVI			
13 MFCC	63	65	95.52
13 MFCC Tartini	60	65	93.28
41 MFCC	66	67	99.25
41 MFCC Tartini	66	67	99.25
41 MFCC Tartini nPVI			
			87.56

Instances ➤	66	67	
itaKor	ita	kor	✔
4 MFCC	57	54	83.46
4 MFCC Tartini	55	59	85.71
4 MFCC nPVI			
13 MFCC	63	62	93.99
13 MFCC Tartini	61	63	93.23
41 MFCC	65	67	99.25
41 MFCC Tartini	65	67	99.25
41 MFCC Tartini nPVI			
			92.48

MLP - Language Pairs - Normalised Audio

Average over File

engSpa	eng	spa	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	8	5	65.00
4 MFCC nPVI	7	5	60.00
13 MFCC	7	7	70.00
13 MFCC Tartini	8	6	70.00
41 MFCC	8	5	65.00
41 MFCC Tartini	8	5	65.00
41 MFCC Tartini nPVI	8	5	65.00
			63.75

dutCze	dut	cze	✔
4 MFCC	10	10	100.00
4 MFCC Tartini	9	10	95.00
4 MFCC nPVI	10	10	100.00
13 MFCC	10	10	100.00
13 MFCC Tartini	8	10	90.00
41 MFCC	10	10	10.00
41 MFCC Tartini	9	10	95.00
41 MFCC Tartini nPVI	9	10	95.00
			85.63

itaCze	ita	cze	✔
4 MFCC	8	7	75.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	5	8	65.00
13 MFCC	9	10	95.00
13 MFCC Tartini	9	10	95.00
41 MFCC	9	8	85.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	9	9	90.00
			83.75

engIta	eng	ita	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	4	4	40.00
4 MFCC nPVI	7	6	65.00
13 MFCC	6	7	65.00
13 MFCC Tartini	7	7	70.00
41 MFCC	5	7	60.00
41 MFCC Tartini	6	7	65.00
41 MFCC Tartini nPVI	7	7	70.00
			60.63

5s Segments

Instances ➤	68	65	
engSpa	eng	spa	✔
4 MFCC	45	50	70.90
4 MFCC Tartini	50	51	75.37
4 MFCC nPVI			
13 MFCC	64	58	91.04
13 MFCC Tartini	65	61	94.03
41 MFCC	66	62	95.52
41 MFCC Tartini	67	63	97.01
41 MFCC Tartini nPVI			
			87.31

Instances ➤	67	65	
dutCze	dut	cze	✔
4 MFCC	63	60	93.18
4 MFCC Tartini	63	59	92.42
4 MFCC nPVI			
13 MFCC	67	60	96.21
13 MFCC Tartini	66	61	96.21
41 MFCC	66	64	98.48
41 MFCC Tartini	66	65	99.24
41 MFCC Tartini nPVI			
			95.96

Instances ➤	66	65	
itaCze	ita	cze	✔
4 MFCC	56	49	80.15
4 MFCC Tartini	51	50	77.10
4 MFCC nPVI			
13 MFCC	62	55	89.31
13 MFCC Tartini	60	53	86.26
41 MFCC	65	64	98.47
41 MFCC Tartini	64	64	97.71
41 MFCC Tartini nPVI			
			88.17

Instances ➤	68	66	
engIta	eng	ita	✔
4 MFCC	55	42	72.39
4 MFCC Tartini	58	54	83.58
4 MFCC nPVI			
13 MFCC	59	54	84.33
13 MFCC Tartini	64	60	92.54
41 MFCC	66	63	96.27
41 MFCC Tartini	66	63	96.27
41 MFCC Tartini nPVI			
			87.56

MLP - Language Pairs - Normalised Audio

Average over File

gerMan	ger	man	✔
4 MFCC	6	4	50.00
4 MFCC Tartini	4	5	45.00
4 MFCC nPVI	2	6	40.00
13 MFCC	9	9	90.00
13 MFCC Tartini	9	9	90.00
41 MFCC	10	9	95.00
41 MFCC Tartini	10	9	95.00
41 MFCC Tartini nPVI	10	8	90.00
			74.38

spaCze	spa	cze	✔
4 MFCC	9	10	95.00
4 MFCC Tartini	8	10	90.00
4 MFCC nPVI	9	10	95.00
13 MFCC	10	10	100.00
13 MFCC Tartini	9	10	95.00
41 MFCC	8	8	80.00
41 MFCC Tartini	6	8	70.00
41 MFCC Tartini nPVI	6	8	70.00
			86.88

rusMan	rus	man	✔
4 MFCC	7	8	75.00
4 MFCC Tartini	6	6	60.00
4 MFCC nPVI	4	6	50.00
13 MFCC	6	7	65.00
13 MFCC Tartini	7	7	70.00
41 MFCC	7	7	70.00
41 MFCC Tartini	5	7	60.00
41 MFCC Tartini nPVI	6	7	65.00
			64.38

gerJap	ger	jap	✔
4 MFCC	7	7	70.00
4 MFCC Tartini	8	6	70.00
4 MFCC nPVI	7	7	70.00
13 MFCC	10	8	90.00
13 MFCC Tartini	10	8	90.00
41 MFCC	10	8	90.00
41 MFCC Tartini	10	8	90.00
41 MFCC Tartini nPVI	10	9	95.00
			83.13

5s Segments

	66	66	
gerMan	ger	man	✔
4 MFCC	53	53	80.30
4 MFCC Tartini	51	50	76.52
4 MFCC nPVI			
13 MFCC	62	60	92.42
13 MFCC Tartini	63	61	93.94
41 MFCC	66	65	99.24
41 MFCC Tartini	66	65	99.24
41 MFCC Tartini nPVI			
			90.28

Instances ➤	66	65	
spaCze	spa	cze	✔
4 MFCC	57	56	86.26
4 MFCC Tartini	58	53	84.73
4 MFCC nPVI			
13 MFCC	63	57	91.60
13 MFCC Tartini	61	58	90.84
41 MFCC	64	63	96.95
41 MFCC Tartini	62	62	96.18
41 MFCC Tartini nPVI			
			91.09

Instances ➤	68	66	
rusMan	rus	man	✔
4 MFCC	54	43	72.39
4 MFCC Tartini	58	55	84.33
4 MFCC nPVI			
13 MFCC	53	55	80.60
13 MFCC Tartini	58	59	87.31
41 MFCC	67	63	97.01
41 MFCC Tartini	66	64	97.01
41 MFCC Tartini nPVI			
			86.44

Instances ➤	66	66	
gerJap	ger	jap	✔
4 MFCC	47	35	62.12
4 MFCC Tartini	49	52	76.52
4 MFCC nPVI			
13 MFCC	61	62	93.18
13 MFCC Tartini	65	65	98.48
41 MFCC	65	66	99.24
41 MFCC Tartini	66	66	100.00
41 MFCC Tartini nPVI			
			88.26

MLP - Language Pairs - Normalised Audio

Average over File

spaKor	spa	kor	✔
4 MFCC	6	5	55.00
4 MFCC Tartini	2	4	30.00
4 MFCC nPVI	6	7	65.00
13 MFCC	6	8	70.00
13 MFCC Tartini	7	7	70.00
41 MFCC	6	9	70.00
41 MFCC Tartini	5	8	65.00
41 MFCC Tartini nPVI	3	8	55.00
			60.00

engMan	eng	man	✔
4 MFCC	4	6	50.00
4 MFCC Tartini	4	7	55.00
4 MFCC nPVI	6	5	55.00
13 MFCC	7	7	70.00
13 MFCC Tartini	7	7	70.00
41 MFCC	7	7	70.00
41 MFCC Tartini	8	7	75.00
41 MFCC Tartini nPVI	8	7	75.00
			65.00

fraIta	fra	ita	✔
4 MFCC	6	2	40.00
4 MFCC Tartini	5	4	45.00
4 MFCC nPVI	7	8	75.00
13 MFCC	4	6	50.00
13 MFCC Tartini	6	3	45.00
41 MFCC	7	7	70.00
41 MFCC Tartini	7	8	75.00
41 MFCC Tartini nPVI	7	9	80.00
			60.00

czeKor	cze	kor	✔
4 MFCC	9	8	85.00
4 MFCC Tartini	10	8	90.00
4 MFCC nPVI	10	9	95.00
13 MFCC	9	9	90.00
13 MFCC Tartini	10	10	100.00
41 MFCC	10	10	100.00
41 MFCC Tartini	10	10	100.00
41 MFCC Tartini nPVI	10	10	100.00
			95.00

5s Segments

Instances ➤	66	67	
spaKor	spa	kor	✔
4 MFCC	44	46	67.67
4 MFCC Tartini	43	50	69.92
4 MFCC nPVI			
13 MFCC	57	62	89.47
13 MFCC Tartini	56	62	88.72
41 MFCC	62	66	96.24
41 MFCC Tartini	62	66	96.24
41 MFCC Tartini nPVI			
			84.71

Instances ➤	68	66	
engMan	eng	man	✔
4 MFCC	51	39	67.16
4 MFCC Tartini	45	44	66.42
4 MFCC nPVI			
13 MFCC	61	58	88.81
13 MFCC Tartini	59	55	85.07
41 MFCC	66	66	98.51
41 MFCC Tartini	66	65	97.76
41 MFCC Tartini nPVI			
			83.96

Instances ➤	67	66	
fraIta	fra	ita	✔
4 MFCC	48	41	66.92
4 MFCC Tartini	41	47	66.17
4 MFCC nPVI			
13 MFCC	53	58	83.46
13 MFCC Tartini	55	57	84.21
41 MFCC	63	63	94.74
41 MFCC Tartini	62	63	93.99
41 MFCC Tartini nPVI			
			81.58

Instances ➤	65	67	
czeKor	cze	kor	✔
4 MFCC	61	64	90.15
4 MFCC Tartini	53	57	83.33
4 MFCC nPVI			
13 MFCC	58	64	92.42
13 MFCC Tartini	55	64	90.15
41 MFCC	65	67	100.00
41 MFCC Tartini	64	67	99.24
41 MFCC Tartini nPVI			
			92.55

MLP - Language Pairs - Normalised Audio

Average over File

gerDut	ger	dut	✔
4 MFCC	8	6	70.00
4 MFCC Tartini	9	9	90.00
4 MFCC nPVI	4	7	55.00
13 MFCC	10	8	90.00
13 MFCC Tartini	9	8	85.00
41 MFCC	10	8	90.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	8	85.00
			81.25

dutPol	dut	pol	✔
4 MFCC	6	3	45.00
4 MFCC Tartini	8	5	65.00
4 MFCC nPVI	6	5	55.00
13 MFCC	7	5	60.00
13 MFCC Tartini	7	6	65.00
41 MFCC	8	7	75.00
41 MFCC Tartini	7	7	70.00
41 MFCC Tartini nPVI	9	8	85.00
			65.00

itaPol	ita	pol	✔
4 MFCC	8	8	80.00
4 MFCC Tartini	9	8	85.00
4 MFCC nPVI	9	7	80.00
13 MFCC	7	6	65.00
13 MFCC Tartini	7	8	75.00
41 MFCC	8	9	85.00
41 MFCC Tartini	9	9	90.00
41 MFCC Tartini nPVI	9	9	90.00
			81.25

engCze	eng	cze	✔
4 MFCC	8	10	90.00
4 MFCC Tartini	7	9	80.00
4 MFCC nPVI	8	8	80.00
13 MFCC	10	10	100.00
13 MFCC Tartini	10	10	100.00
41 MFCC	9	9	90.00
41 MFCC Tartini	10	9	90.00
41 MFCC Tartini nPVI	10	9	95.00
			90.63

5s Segments

Instances ➤	66	67	
gerDut	ger	dut	✔
4 MFCC	48	57	78.95
4 MFCC Tartini	55	57	84.21
4 MFCC nPVI			
13 MFCC	59	63	91.50
13 MFCC Tartini	59	61	90.23
41 MFCC	65	66	98.50
41 MFCC Tartini	65	65	97.74
41 MFCC Tartini nPVI			
			90.19

Instances ➤	67	69	
dutPol	dut	pol	✔
4 MFCC	49	53	75.00
4 MFCC Tartini	57	59	85.29
4 MFCC nPVI			
13 MFCC	59	62	88.97
13 MFCC Tartini	67	62	91.18
41 MFCC	65	67	97.06
41 MFCC Tartini	65	67	97.06
41 MFCC Tartini nPVI			
			89.09

Instances ➤	66	69	
itaPol	ita	pol	✔
4 MFCC	54	59	83.70
4 MFCC Tartini	54	58	82.96
4 MFCC nPVI			
13 MFCC	62	63	92.59
13 MFCC Tartini	58	62	88.89
41 MFCC	63	67	96.30
41 MFCC Tartini	62	67	95.56
41 MFCC Tartini nPVI			
			90.00

Instances ➤	68	65	
engCze	eng	cze	✔
4 MFCC	53	48	75.94
4 MFCC Tartini	63	59	91.73
4 MFCC nPVI			
13 MFCC	65	56	90.98
13 MFCC Tartini	67	58	93.99
41 MFCC	68	61	96.99
41 MFCC Tartini	68	62	97.74
41 MFCC Tartini nPVI			
			91.23

MLP - Language Pairs - Normalised Audio

Average over File

dutRus	dut	rus	✔
4 MFCC	6	4	50.00
4 MFCC Tartini	9	8	85.00
4 MFCC nPVI	6	5	55.00
13 MFCC	6	6	60.00
13 MFCC Tartini	7	6	65.00
41 MFCC	9	8	85.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	8	85.00
			71.25

spaPol	spa	pol	✔
4 MFCC	3	6	45.00
4 MFCC Tartini	4	5	45.00
4 MFCC nPVI	4	6	50.00
13 MFCC	6	9	75.00
13 MFCC Tartini	6	8	70.00
41 MFCC	6	7	65.00
41 MFCC Tartini	6	7	65.00
41 MFCC Tartini nPVI	6	7	65.00
			60.00

rusKor	rus	kor	✔
4 MFCC	5	7	60.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	4	6	50.00
13 MFCC	7	9	80.00
13 MFCC Tartini	8	9	85.00
41 MFCC	9	9	90.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	9	90.00
			75.63

dutFra	dut	fra	✔
4 MFCC	8	5	65.00
4 MFCC Tartini	8	8	80.00
4 MFCC nPVI	8	6	70.00
13 MFCC	9	8	85.00
13 MFCC Tartini	7	9	80.00
41 MFCC	9	9	90.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	9	90.00
			80.63

5s Segments

Instances ➤	67	68	
dutRus	dut	rus	✔
4 MFCC	57	40	71.85
4 MFCC Tartini	54	46	74.07
4 MFCC nPVI			
13 MFCC	58	59	86.67
13 MFCC Tartini	61	62	91.11
41 MFCC	66	67	98.52
41 MFCC Tartini	65	67	97.78
41 MFCC Tartini nPVI			
			86.67

Instances ➤	66	69	
spaPol	spa	pol	✔
4 MFCC	50	47	71.85
4 MFCC Tartini	48	45	68.89
4 MFCC nPVI			
13 MFCC	60	62	90.37
13 MFCC Tartini	57	60	86.67
41 MFCC	64	69	98.52
41 MFCC Tartini	65	68	98.52
41 MFCC Tartini nPVI			
			85.80

Instances ➤	68	67	
rusKor	rus	kor	✔
4 MFCC	48	38	63.70
4 MFCC Tartini	48	42	66.67
4 MFCC nPVI			
13 MFCC	61	66	94.07
13 MFCC Tartini	65	65	96.30
41 MFCC	67	67	99.26
41 MFCC Tartini	67	67	99.26
41 MFCC Tartini nPVI			
			86.54

Instances ➤	67	67	
dutFra	dut	fra	✔
4 MFCC	57	50	79.85
4 MFCC Tartini	59	58	86.57
4 MFCC nPVI			
13 MFCC	58	57	85.82
13 MFCC Tartini	59	58	87.31
41 MFCC	65	65	97.01
41 MFCC Tartini	65	65	97.01
41 MFCC Tartini nPVI			
			88.93

MLP - Language Pairs - Normalised Audio

Average over File

czeRus	cze	rus	✔
4 MFCC	7	7	70.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	8	7	75.00
13 MFCC	7	8	75.00
13 MFCC Tartini	10	8	90.00
41 MFCC	7	8	75.00
41 MFCC Tartini	7	8	75.00
41 MFCC Tartini nPVI	8	8	80.00
			76.88

gerFra	ger	fra	✔
4 MFCC	8	5	65.00
4 MFCC Tartini	7	7	70.00
4 MFCC nPVI	6	7	65.00
13 MFCC	9	10	95.00
13 MFCC Tartini	9	10	95.00
41 MFCC	9	7	80.00
41 MFCC Tartini	9	5	70.00
41 MFCC Tartini nPVI	9	7	80.00
			77.50

fraPol	fra	pol	✔
4 MFCC	6	7	65.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	8	6	70.00
13 MFCC	8	6	70.00
13 MFCC Tartini	8	7	75.00
41 MFCC	6	8	70.00
41 MFCC Tartini	6	7	65.00
41 MFCC Tartini nPVI	6	8	70.00
			70.00

itaRus	ita	rus	✔
4 MFCC	8	7	75.00
4 MFCC Tartini	7	8	75.00
4 MFCC nPVI	8	7	75.00
13 MFCC	8	8	80.00
13 MFCC Tartini	8	8	80.00
41 MFCC	9	10	95.00
41 MFCC Tartini	9	10	95.00
41 MFCC Tartini nPVI	9	10	95.00
			83.75

5s Segments

czeRus	cze	rus	✔
Instances ➤	65	68	
4 MFCC	52	56	81.20
4 MFCC Tartini	55	58	84.96
4 MFCC nPVI			
13 MFCC	59	63	91.73
13 MFCC Tartini	60	59	89.47
41 MFCC	63	66	96.99
41 MFCC Tartini	62	66	96.24
41 MFCC Tartini nPVI			
			90.10

gerFra	ger	fra	✔
Instances ➤	66	67	
4 MFCC	41	53	70.68
4 MFCC Tartini	42	51	69.92
4 MFCC nPVI			
13 MFCC	58	60	88.72
13 MFCC Tartini	58	59	87.97
41 MFCC	66	63	96.99
41 MFCC Tartini	66	63	96.99
41 MFCC Tartini nPVI			
			85.21

fraPol	fra	pol	✔
Instances ➤	67	69	
4 MFCC	54	58	82.35
4 MFCC Tartini	61	62	90.44
4 MFCC nPVI			
13 MFCC	57	65	89.71
13 MFCC Tartini	55	63	86.76
41 MFCC	55	63	86.76
41 MFCC Tartini	64	67	96.32
41 MFCC Tartini nPVI			
			88.73

itaRus	ita	rus	✔
Instances ➤	66	66	
4 MFCC	53	53	79.10
4 MFCC Tartini	53	56	81.34
4 MFCC nPVI			
13 MFCC	62	63	93.28
13 MFCC Tartini	60	57	87.31
41 MFCC	65	67	98.51
41 MFCC Tartini	65	67	98.51
41 MFCC Tartini nPVI			
			89.68

MLP - Language Pairs - Normalised Audio

Average over File

engPol	eng	pol	✔
4 MFCC	5	5	50.00
4 MFCC Tartini	5	5	50.00
4 MFCC nPVI	6	5	55.00
13 MFCC	8	8	80.00
13 MFCC Tartini	9	7	80.00
41 MFCC	8	8	80.00
41 MFCC Tartini	9	8	85.00
41 MFCC Tartini nPVI	9	9	90.00
			71.25

dutMan	dut	man	✔
4 MFCC	7	6	65.00
4 MFCC Tartini	8	7	75.00
4 MFCC nPVI	7	6	65.00
13 MFCC	9	9	90.00
13 MFCC Tartini	9	8	95.00
41 MFCC	9	10	95.00
41 MFCC Tartini	8	9	85.00
41 MFCC Tartini nPVI	8	9	95.00
			83.13

itaMan	ita	man	✔
4 MFCC	4	6	50.00
4 MFCC Tartini	6	7	65.00
4 MFCC nPVI	6	4	50.00
13 MFCC	7	6	65.00
13 MFCC Tartini	7	10	85.00
41 MFCC	7	7	70.00
41 MFCC Tartini	7	8	75.00
41 MFCC Tartini nPVI	7	8	75.00
			66.88

rusJap	rus	jap	✔
4 MFCC	5	7	60.00
4 MFCC Tartini	7	7	70.00
4 MFCC nPVI	5	7	60.00
13 MFCC	4	5	45.00
13 MFCC Tartini	5	5	50.00
41 MFCC	7	7	70.00
41 MFCC Tartini	5	6	55.00
41 MFCC Tartini nPVI	5	6	55.00
			58.13

5s Segments

engPol	eng	pol	✔
Instances ➤	68	69	
4 MFCC	41	56	70.80
4 MFCC Tartini	52	54	77.37
4 MFCC nPVI			
13 MFCC	62	60	89.05
13 MFCC Tartini	63	62	91.24
41 MFCC	67	69	99.27
41 MFCC Tartini	68	68	99.27
41 MFCC Tartini nPVI			
			87.83

dutMan	dut	man	✔
Instances ➤	67	66	
4 MFCC	59	57	87.22
4 MFCC Tartini	62	58	90.23
4 MFCC nPVI			
13 MFCC	61	64	93.99
13 MFCC Tartini	65	63	96.24
41 MFCC	66	64	97.74
41 MFCC Tartini	65	65	97.74
41 MFCC Tartini nPVI			
			93.86

itaMan	ita	man	✔
Instances ➤	66	66	
4 MFCC	51	45	72.73
4 MFCC Tartini	47	51	74.24
4 MFCC nPVI			
13 MFCC	55	59	86.36
13 MFCC Tartini	58	59	88.64
41 MFCC	63	65	96.97
41 MFCC Tartini	63	65	96.97
41 MFCC Tartini nPVI			
			85.98

rusJap	rus	jap	✔
Instances ➤	66	66	
4 MFCC	44	43	64.93
4 MFCC Tartini	50	44	70.15
4 MFCC nPVI			
13 MFCC	55	60	85.82
13 MFCC Tartini	55	60	85.82
41 MFCC	67	64	97.76
41 MFCC Tartini	67	64	97.76
41 MFCC Tartini nPVI			
			83.71

MLP - Language Pairs - Normalised Audio

MLP - Language Pairs - Normalised Audio

Average over File

fraSpa	fra	spa	✔
4 MFCC	5	6	55.00
4 MFCC Tartini	7	6	65.00
4 MFCC nPVI	8	6	70.00
13 MFCC	8	7	75.00
13 MFCC Tartini	7	7	70.00
41 MFCC	6	6	60.00
41 MFCC Tartini	7	6	65.00
41 MFCC Tartini nPVI	7	6	65.00
			65.63

czeMan	cze	man	✔
4 MFCC	10	9	95.00
4 MFCC Tartini	10	8	90.00
4 MFCC nPVI	10	10	100.00
13 MFCC	9	10	95.00
13 MFCC Tartini	9	10	95.00
41 MFCC	8	9	95.00
41 MFCC Tartini	8	9	85.00
41 MFCC Tartini nPVI	8	9	85.00
			92.50

5s Segments

Instances ➤	67	66	
fraSpa	fra	spa	✔
4 MFCC	47	55	76.69
4 MFCC Tartini	54	56	82.71
4 MFCC nPVI			
13 MFCC	59	64	92.48
13 MFCC Tartini	61	62	92.48
41 MFCC	61	62	92.48
41 MFCC Tartini	62	63	93.99
41 MFCC Tartini nPVI			
			88.47

Instances ➤	65	66	
czeMan	cze	man	✔
4 MFCC	56	59	87.79
4 MFCC Tartini	54	57	84.73
4 MFCC nPVI			
13 MFCC	55	56	84.73
13 MFCC Tartini	55	58	86.26
41 MFCC	61	65	96.18
41 MFCC Tartini	62	65	96.95
41 MFCC Tartini nPVI			
			89.44

Language Family Comparisons

Language Family Comparisons

MLP, Unnormalised Audio, Averaged

Germanic	eng	ger	dut	✔
4 MFCC	4	1	4	30.00
13 MFCC Tartini				
41 MFCC Tartini	9	9	8	86.67
All Features	9	8	8	83.33
				66.67

Romantic	fra	ita	spa	✔
4 MFCC	1	4	5	33.33
13 MFCC Tartini				
41 MFCC Tartini	4	8	7	63.33
All Features	4	9	7	66.67
				54.44

Slavic	cze	pol	rus	✔
4 MFCC	8	5	4	56.67
13 MFCC Tartini				
41 MFCC Tartini	7	7	6	66.67
All Features	7	6	6	63.33
				62.22

SinoAltaic	man	kor	jap	✔
4 MFCC	8	4	4	53.33
13 MFCC Tartini				
41 MFCC Tartini	4	9	8	70.00
All Features	5	9	7	70.00
				64.44

MLP, Normalised Audio, Averaged

Germanic	eng	ger	dut	✔
4 MFCC	3	1	4	26.67
13 MFCC Tartini	9	8	8	83.33
41 MFCC Tartini	9	9	8	86.67
All Features	9	8	8	83.33
				70.00

Romantic	fra	ita	spa	✔
4 MFCC	1	2	6	30.00
13 MFCC Tartini	3	4	6	43.33
41 MFCC Tartini	4	8	7	63.33
All Features	3	8	6	56.67
				48.33

Slavic	cze	pol	rus	✔
4 MFCC	7	6	3	53.33
13 MFCC Tartini	10	8	6	80.00
41 MFCC Tartini	7	7	6	66.67
All Features	7	6	7	66.67
				66.67

SinoAltaic	man	kor	jap	✔
4 MFCC	8	5	5	60.00
13 MFCC Tartini	4	8	7	63.33
41 MFCC Tartini	4	9	8	70.00
All Features	6	8	8	73.33
				66.67

MLP, Normalised Audio, 5s Segments

Instances ➤	68	66	67	
Germanic	eng	ger	dut	✔
4 MFCC	38	26	40	51.74
13 MFCC Tartini	63	60	60	91.04
41 MFCC Tartini	65	64	64	96.02
All Features				
				79.60

Instances ➤	67	66	66	
Romantic	fra	ita	spa	✔
4 MFCC	33	30	50	56.78
13 MFCC Tartini	49	53	62	82.41
41 MFCC Tartini	56	61	61	89.45
All Features				
				76.21

Instances ➤	65	69	68	
Slavic	cze	pol	rus	✔
4 MFCC	45	44	36	61.88
13 MFCC Tartini	56	60	53	83.66
41 MFCC Tartini	59	65	62	92.08
All Features				
				79.21

Instances ➤	66	67	66	
SinoAltaic	man	kor	jap	✔
4 MFCC	42	36	26	52.26
13 MFCC Tartini	59	62	58	89.95
41 MFCC Tartini	62	64	64	95.48
All Features				
				79.23

Language Family Comparisons

Language Family Comparisons

Naïve Bayes, Normalised Audio, 5s Segments

Instances ➤	68	66	67	
Germanic	eng	ger	dut	✔
4 MFCC	13	42	44	49.25
13 MFCC Tartini	55	51	55	80.10
41 MFCC Tartini	60	59	60	89.05
All Features				
				72.80
Instances ➤	67	66	66	
Romantic	fra	ita	spa	✔
4 MFCC	35	37	24	48.24
13 MFCC Tartini	50	29	43	61.31
41 MFCC Tartini	53	39	50	71.36
All Features				
				60.30
Instances ➤	65	69	68	
Slavic	cze	pol	rus	✔
4 MFCC	39	16	53	53.47
13 MFCC Tartini	40	45	51	67.33
41 MFCC Tartini	44	59	63	82.18
All Features				
				67.66
Instances ➤	66	67	66	
SinoAltaic	man	kor	jap	✔
4 MFCC	49	30	13	46.23
13 MFCC Tartini	56	55	45	78.39
41 MFCC Tartini	64	63	62	94.97
All Features				
				73.20

Naïve Bayes, Normalised Audio, Average Across File

Germanic	eng	ger	dut	✔
4 MFCC	3	5	4	40.00
13 MFCC Tartini	6	8	8	73.33
41 MFCC Tartini	4	8	7	63.33
All Features	6	9	7	73.33
				62.50
Romantic	fra	ita	spa	✔
4 MFCC	4	6	6	53.33
13 MFCC Tartini	2	6	9	56.67
41 MFCC Tartini	2	6	7	50.00
All Features	2	7	7	53.33
				53.33
Slavic	cze	pol	rus	✔
4 MFCC	7	4	2	43.33
13 MFCC Tartini	10	6	5	70.00
41 MFCC Tartini	9	6	5	66.67
All Features	10	7	5	73.33
				63.33
SinoAltaic	man	kor	jap	✔
4 MFCC	7	5	3	50.00
13 MFCC Tartini	7	7	6	66.67
41 MFCC Tartini	7	7	7	70.00
All Features	7	9	6	73.33
				65.00

Comparison Between All Languages

MLP, Unnormalised Audio, Average over File, Incorrect nPVI Function

All Languages	eng	ger	dut	fra	ita	spa	cze	pol	rus	man	kor	jap	✔
4 MFCC	0	1	2	0	0	1	8	0	1	3	1	3	16.67
4 MFCC Tartini	0	3	5	1	2	1	4	1	1	1	1	0	16.67
4 MFCC Tartini nPVI	1	2	5	0	0	0	5	0	2	3	1	2	21.00
13 MFCC	3	7	5	3	4	1	4	5	1	2	2	4	41.00
13 MFCC Tartini	2	6	5	2	3	1	6	2	2	2	3	3	37.00
13 MFCC Tartini nPVI	2	7	4	0	5	1	6	2	1	4	3	2	37.00
41 MFCC	3	5	5	3	5	1	7	2	1	4	5	5	46.00
41 MFCC Tartini	3	5	4	3	5	2	7	1	1	4	6	5	46.00
41 MFCC Tartini nPVI	3	7	5	2	3	1	8	1	1	3	7	5	46.00
													34.15

MLP, Normalised Audio, Average over File, Corrected nPVI Function

All Languages	eng	ger	dut	fra	ita	spa	cze	pol	rus	man	kor	jap	✔
4 MFCC	2	1	3	0	2	2	6	0	1	1	1	4	19.17
4 MFCC Tartini	1	0	6	0	1	0	6	0	1	0	0	5	16.67
4 MFCC Tartini nPVI	2	4	8	3	1	1	5	1	0	2	1	3	25.83
13 MFCC	1	6	3	2	2	1	4	5	0	1	4	3	26.67
13 MFCC Tartini	2	7	6	0	3	2	7	4	1	2	3	3	33.33
13 MFCC Tartini nPVI	3	6	5	3	5	1	4	4	1	2	2	5	34.17
41 MFCC	4	6	5	3	6	1	8	2	2	2	7	5	42.50
41 MFCC Tartini	3	6	5	3	5	1	7	2	3	4	5	6	41.67
41 MFCC Tartini nPVI	3	6	5	3	4	1	7	2	1	2	7	6	39.17
													31.02

MLP, Normailsed Audio, 5s Segments

Instances ➤	68	66	67	67	68	66	65	67	68	66	68	66	
All Languages	eng	ger	dut	fra	ita	spa	cze	pol	rus	man	kor	jap	✔
4 MFCC	10	17	34	19	24	15	25	17	22	30	25	5	30.34
4 MFCC Tartini	20	25	40	26	22	24	35	16	21	31	25	12	37.08
4 MFCC Tartini nPVI													
13 MFCC	29	42	45	28	31	32	31	40	14	39	31	41	50.31
13 MFCC Tartini	43	48	43	31	34	27	29	43	36	40	33	43	56.18
13 MFCC Tartini nPVI													
41 MFCC	58	61	63	53	58	48	58	59	60	56	65	64	87.77
41 MFCC Tartini	63	62	63	51	58	47	60	60	63	59	65	62	89.01
41 MFCC Tartini nPVI													
													58.45

Comparison Between All Languages

Naïve Bayes, Normalised Audio, Average Across File

All Languages	eng	ger	dut	fra	ita	spa	cze	pol	rus	man	kor	jap	✔
4 MFCC	0	2	1	1	2	3	6	0	0	0	3	2	16.67
4 MFCC Tartini	1	5	3	0	0	4	5	1	0	3	4	1	22.50
4 MFCC Tartini nPVI	1	6	3	1	0	4	5	3	0	4	2	0	24.17
13 MFCC	1	6	4	2	3	4	7	5	2	1	4	4	35.83
13 MFCC Tartini	1	8	5	0	1	5	8	4	1	4	5	2	36.67
13 MFCC Tartini nPVI	2	8	5	1	2	4	8	5	1	5	5	1	39.17
41 MFCC	2	5	3	2	5	3	6	3	2	2	3	4	33.33
41 MFCC Tartini	1	5	3	2	4	3	6	3	1	3	2	3	30.00
41 MFCC Tartini nPVI	1	5	3	2	5	3	6	4	2	4	2	4	34.17
													30.28

Naïve Bayes, Normalised Audio, 5s Segments

Instances ➤	68	66	67	67	68	66	65	67	68	66	68	66	
All Languages	eng	ger	dut	fra	ita	spa	cze	pol	rus	man	kor	jap	✔
4 MFCC	0	22	35	24	24	8	23	2	26	20	9	0	24.09
4 MFCC Tartini	8	26	48	16	7	10	30	7	26	27	15	3	27.84
4 MFCC Tartini nPVI													
13 MFCC	9	35	40	40	20	10	22	22	25	29	33	40	40.57
13 MFCC Tartini	5	36	44	40	16	12	22	24	25	35	33	26	39.70
13 MFCC Tartini nPVI													
41 MFCC	48	53	58	44	36	28	36	49	53	40	52	55	68.91
41 MFCC Tartini	45	55	55	46	37	30	38	53	54	43	51	47	69.16
41 MFCC Tartini nPVI													
													45.05