

# Determination of Bahasa Melayu Word List From Friday Sermon Transcripts Using PHP and MySQL

Mokhtar Harun<sup>a</sup>, Muhammad 'Aasim Asyafi'ie Ahmad<sup>b\*</sup>, Siti Zaleha Abdul Hamid<sup>a</sup>, Fareha Abdul Rahman<sup>a</sup>, Puspa Inayat Khalid<sup>a</sup>

<sup>a</sup>Fakulti Kejuruteraan Elektrik, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

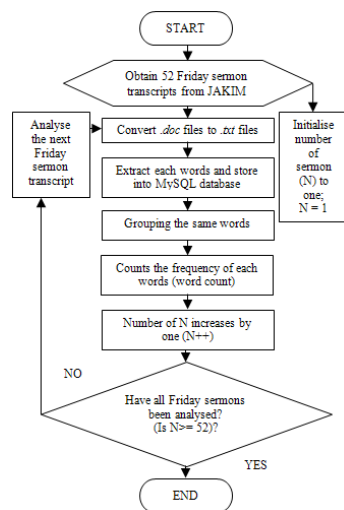
<sup>b</sup>UB2-1039, Kolej 10, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

\*Corresponding author: muhammad\_aasim64@yahoo.com

## Article history

Received :12 July 2012  
Received in revised form:  
4 April 2013  
Accepted :15 April 2013

## Graphical abstract



## Abstract

Almost all of the speech intelligibility (SI) assessors today are from the West. The words used in the development of these SI assessors are mainly English. However, recent finding has found that word components that contribute to intelligibility of speech is different from one language to another. The purpose of this study is to determine Bahasa Melayu (BM) word list from Friday sermon transcripts. The project was undertaken by deriving BM words spoken during 52 Friday sermons read in mosques in Wilayah Persekutuan Kuala Lumpur. PHP is the programming language used to compute the frequency of words appears in the Friday sermon transcripts. These words were then stored into a database. The database used is MySQL. It has been found that the highest number of repeated word is 2509. The list is then truncated so that the numbers of repeated words range from 52 to 2509 times. The frequency of Bahasa Melayu words used in Friday sermons is an initial and important inventory that can be used in the development of SI assessor using BM words.

**Keywords:** Word list; Malay; speech intelligibility; mosque, languages

## Abstrak

Hampir semua penilaian kejelasan ucapan (SI) berasal dari negara-negara Barat. Perkataan-perkataan yang digunakan untuk membuat penilaian SI selalunya berdasarkan bahasa Inggeris. Walaubagaimana pun, penemuan baru telah mendapati bahawa komponen perkataan yang menyumbang kepada kejelasan bersuara adalah berbeza dari satu bahasa kepada bahasa yang lain. Tujuan kajian ini adalah untuk menentukan perkataan-perkataan Bahasa Melayu (BM) daripada transkrip khutbah Jumaat. Projek ini telah dijalankan dengan memperoleh perkataan-perkataan yang dibaca semasa 52 khutbah-khutbah Jumaat dalam masjid-masjid di Wilayah Persekutuan Kuala Lumpur. Bilangan ulangan perkataan dalam khutbah telah dikira dengan menggunakan kod PHP dan disimpan dalam pangkalan data MySQL. Jumlah tertinggi ulangan perkataan adalah 2509. Senarai ini kemudian dikecilkan supaya bilangan perkataan berulang adalah antara 52-2509 kali. Kekerapan perkataan Bahasa Melayu yang digunakan dalam khutbah-khutbah Jumaat adalah inventori awal penting yang boleh digunakan dalam pembangunan penilai SI yang menggunakan perkataan BM.

**Kata kunci:** Senarai perkataan; bahasa Melayu; kejelasan percakapan; masjid; bahasa

© 2013 Penerbit UTM Press. All rights reserved.

## 1.0 INTRODUCTION

One of the predictor for word recognition is the frequency of the word under test. Specifically, word frequencies can be a predictor for the amount of time needed to recognise a certain word [1,2]. High frequency words, for example, can be perceived and produced faster [3]. Thus, it is of no surprise that researchers have invested time in investigating word frequency. The earliest word frequency list published was made by Thorndike and Lorge [4]. The authors compiled a list of American English words and their frequencies so that it can be used to determine which words are to be selected for teaching the students. Some of the more recent

word list used by language researchers is published in [5]. This new list used the emergence of new technology, specifically the internet, to collect more data and thus gain a more accurate result.

However, the word lists provided are not available in Malay. Some other languages that have their own word list database include Dutch [6] and French [7]. This fact need to be taken into account because speech signals contain different components that are important for intelligibility of speech. These components vary among different languages. For example, in English, vowels are more important for sentence intelligibility whereas consonants are slightly more important for word intelligibility [8]. According to [9], Danes understood spoken Swedish more than Swedes

understood Danish. One of the conclusions the authors draw is that Danish is less intelligible than Swedish. A comparison study between Mandarin and English had been conducted and it shows that in Mandarin, high-energy voice segment like vowels are important to intelligibility. However, in English, information from weak consonants and from vowel/consonant transition is more important to intelligibility [10]. This is further supported when [11] discovered that the spectral difference between vowels in Finnish and in English differs around 100Hz. For tonal languages such as Mandarin and Cantonese, the most important variable to be used to discern one tone to another is its fundamental frequency [10].

### 1.1 SPEECH INTELLIGIBILITY

In Malaysia, in order for a contractor to receive full payment from client for the listening room he or she built, the contractor is often required to pass speech intelligibility (SI) test [12]. But, these SI tests usually use English language as medium of test. This can prove problematic because non-native to the English language perceives the English language differently than the native, especially the phone /p/, /t/, /k/ and /g/ [13]. The authors from [14] also discovered that human phoneme recognition depends on speech variability. This problem intensifies when [11] reports that conversational and clear speech for both Finnish and English are different.

SI assessors developed by using non-English materials are growing. For example, the authors from [15] developed a logatom (nonsense words) word lists specifically to test the speech intelligibility in American Spanish speaking environment. These nonsense words is said to be able to assess the intelligibility of a room. Danish now also have a word list that Danes can use to assess speech intelligibility [16]. Japanese also have developed word list of their own which categorised the words according to their familiarity. Different Japanese words can bring different intelligibility, i.e. some words are less intelligible. Malay also has its own word list made by [17], but this word list is designed to assess the hearing ability of a person and so the specific words used are not that important. Swedish word list have also been developed by [18] which modified the method first introduced by [19] which shorten the time needed for diagnostics. To assess the quality of Turkish speech communication system, [20] have developed a Turkish word list which is based on earlier English intelligibility test, the Diagnostic Rhyme Test [21]. Although not specifically a word list, German also developed list of sentences to be used for intelligibility test. This sentence list can be useful to a variety of applications such as clinical audiology, hearing aid and communication systems evaluation - all in German language [22]. From this, it can be concluded that each language need to explore their own speech intelligibility tests to account for the uniqueness of each language.

Malay, which is spoken by 250 million people [23], has only about 500 monosyllables and half of it consists of loan words from other languages, especially English [24]. This effectively limits the use of diagnostic rhyme test since the test requires a sizeable portion of monosyllabic words even though it is considered to mimic everyday speech [20]. The most common form of words in Malay is of disyllabic type. The number of trisyllabic word structure is small in Malay. There are few words that contain four syllables or more but most of these words are loan words from Sanskrit, Arabic and recently English [25]. Due to these inherent qualities, it is important that further research is done for Malay language.

## 2.0 METHODOLOGY

The spoken Malay words were obtained from 52 Friday sermon transcripts that were spoken in mosques located at Wilayah Persekutuan Kuala Lumpur. Only the speech transcripts were used because the videos are too slow and the voices can be undecipherable at times. Also, 52 sermons were selected to get an average number of words spoken for one year. It is known that sermons' content may depends on the time of the year, so this step is taken to ensure homogeneity.

The transcripts were acquired from Jabatan Kebajikan Islam (JAKIM) official website. Many of which are available in .doc forms. Even though Friday sermons are usually delivered in two parts, many of the transcripts in the website only carry the first part. This is justified since the second part is the same in every sermon and bears little difference. Figure 1 shows the sample of the sermon.

Jabatan Kemajuan Islam Malaysia  
"AMALAN BERSEDEKAH MEMBAWA BERKAT"  
(8 APRIL 2011 / 4 JAMADILAWAL 1432H)

---

الْحَمْدُ لِلَّهِ الْعَاقِلِ :

مَثَلُ الَّذِينَ يُبْتِغُونَ أَمْوَالَهُمْ فِي سَبِيلِ اللَّهِ كَمَثَلِ حَبَّةٍ أَتَتْ سَنَابِلَ سَبْعِ سَنَابِلَ فِي كُلِّ سُبُكَّةٍ مِائَةٌ حَبَّةٌ وَاللَّهُ يُضَاعِفُ لِمَنْ يَشَاءُ وَاللَّهُ وَاسِعٌ عَلِيمٌ

أَشْهَدُ أَنْ لَا إِلَهَ إِلَّا اللَّهُ وَحْدَهُ لَا شَرِيكَ لَهُ، وَأَشْهَدُ أَنَّ مُحَمَّدًا عَبْدُهُ وَرَسُولُهُ، اللَّهُمَّ صَلِّ وَسَلِّمْ عَلَى سَيِّدِنَا مُحَمَّدٍ وَعَلَى آلِهِ وَأَصْحَابِهِ أَجْمَعِينَ.

أَمَّا بَعْدُ، فَيَا عِبَادَ اللَّهِ، اتَّقُوا اللَّهَ، أَوْصِيكُمْ وَإِيَّايَ بِتَقْوَى اللَّهِ، فَقَدْ فَازَ الْمُتَّقُونَ.

**Muslimin Yang di Rahmati Allah,**  
Di kesempatan berada pada hari Jumaat yang mulia ini, saya berpesan kepada diri saya dan kepada sidang Jumaat sekalian marilah sama-sama kita meningkatkan ketakwaan kita kepada Allah SWT dengan bersungguh-sungguh melaksanakan segala perintah-Nya dan meninggalkan segala larangan-Nya. Mudah-mudahan kita memperoleh keberkatan dan keredhaan serta perlindungan Allah SWT di dunia dan di akhirat. Mimbar pada hari ini akan menyampaikan khutbah bertajuk: "AMALAN BERSEDEKAH MEMBAWA BERKAT".

**Muslimin Yang di Rahmati Allah,**  
Sedekah adalah suatu sunnah dan amalannya para sahabat yang sangat mulia. Dalam tradisi Islam ia ada beberapa tahap seperti zakat, sedekah, derma dan hadiah. Zakat

Figure 1 A sample of one of the 52 Friday sermons

### 2.1 PHP and MySQL

To organise the data for this research, a database management system is needed. The database needs to be able to store a huge amount of data and manage those data according to their relevance. The database also needs to be reliable, easy to use and is able to retrieve the stored data quickly.

From the criteria stated, this research had chosen MySQL as the database to be used. The SQL in its name stands for Structured Query Language which is a programming language designed to manage data. MySQL is the most widely used open source database management system, which was first introduced in 1995. In addition, MySQL is free of charge and does not require any license for non-commercial use. But to manipulate the data inside MySQL, we need to use a certain programming language.

The programming language chosen for this research is Personal Home Page (PHP). Even though PHP is used mainly in developing web pages, it can also be applied to a wide variety of situation including manipulating MySQL data. In this paper, PHP is used to manipulate the Friday sermon transcripts so that it can be stored into MySQL database. After that, PHP is used to extract

the stored data, manipulate them to obtain new data and then store the new data back into MySQL database.

The Friday sermons transcripts must first be standardised before PHP can be used to manipulate the texts within. This is done by removing any images and Arabic letters. The numbers represented by Arabic numerals were replaced with appropriate words. Acronyms and initialism were also replaced with their corresponding words. This is because the main objective of this step is to find meaningful Malay words and store them into the database. After that, the files were exploded so that the words are separated by lines and this will make it easier to categorise by inserting the words into different arrays. These raw data which were stored in arrays were then stored into 52 tables in MySQL database. This step is repeated until all the 52 .txt files were covered.

After inputting all the words into MySQL database in 52 different tables, all these data were fetched and merged into one table. The words in this table were arranged alphabetically so that the same words can be grouped together. More PHP codes were implemented so that it groups together the same words into one array. These words were automatically inserted into a different table in MySQL database with the corresponding word.

At this step, the words were arranged alphabetically but have not been counted for their number of frequency. The data from the latest table was fetched and were run through more PHP codes. The same words were group together into one array and the number of the elements of said array corresponds to the frequency for the word. The numbers of elements were inserted into another table in MySQL database alongside with the corresponding words. This final table can then be extracted into Excel or Word format and can be analysed. The codes were written by using notepad and PHPMyAdmin was used to manage the database due to the ease of use. The process is summarised in Figure 2.

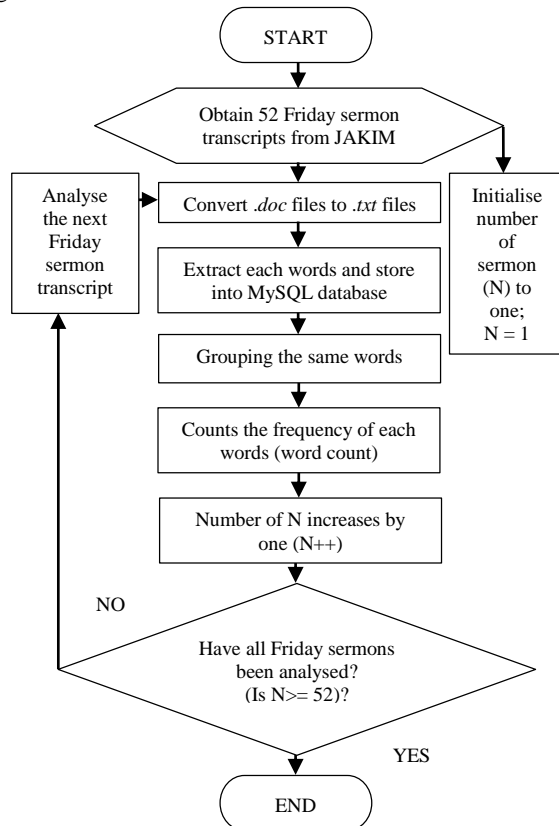


Figure 2 Flowchart for the process of obtaining frequency of Malay words from speech transcripts

### 3.0 RESULTS AND DISCUSSION

The Malay word list has been extracted from a total of 52 Friday sermons. There are total of 26244 words in the sermon transcripts and of those words, there are 5937 distinct words. The word with the highest frequency of word occurrence is 2509 times. Disyllabic words are the most commonly used in the transcripts. Table 1 lists spoken frequency of Malay words that is spoken in the range of 2509 to 102 times. The most spoken word is 'dan' (2509) which means that on the average; the word 'dan' was spoken about 50 times in each Friday sermon.

Table 1 The list of words obtained from 52 Friday sermons arranged based on the spoken frequency from 2509 to 102 times

| Words (Malay) | Words (English) | Frequency |
|---------------|-----------------|-----------|
| dan           | (and)           | 2509      |
| yang          | (which)         | 1969      |
| Allah         | (Allah)         | 1074      |
| di            | (at)            | 1056      |
| kita          | (we)            | 977       |
| dengan        | (with)          | 908       |
| ini           | (this)          | 718       |
| dalam         | (in)            | 706       |
| kepada        | (for)           | 701       |
| mereka        | (they)          | 496       |
| Islam         | (Islam)         | 485       |
| tidak         | (no)            | 485       |
| sama          | (same)          | 431       |
| akan          | (will)          | 428       |
| al            | (al)            | 408       |
| untuk         | (for)           | 394       |
| pada          | (at)            | 374       |
| telah         | (have)          | 357       |
| orang         | (people)        | 344       |
| itu           | (that)          | 336       |
| adalah        | (is)            | 290       |
| kamu          | (you)           | 288       |
| dari          | (from)          | 287       |
| umat          | (community)     | 273       |
| oleh          | (by)            | 262       |
| sebagai       | (as)            | 261       |
| negara        | (country)       | 251       |
| hari          | (day)           | 250       |
| bahawa        | (that)          | 248       |
| maka          | (so)            | 246       |
| anak          | (child)         | 219       |
| serta         | (with)          | 218       |
| mimbar        | (pulpit)        | 218       |
| ayat          | (word)          | 217       |
| manusia       | (people)        | 216       |
| diri          | (self)          | 212       |
| segala        | (all)           | 210       |
| juga          | (also)          | 209       |
| Rasulullah    | (Rasulullah)    | 202       |
| nya           | (him)           | 198       |
| Quran         | (Quran)         | 197       |

| Words (Malay) | Words (English) | Frequency |
|---------------|-----------------|-----------|
| atau          | (or)            | 194       |
| marilah       | (come)          | 193       |
| menjadi       | (become)        | 193       |
| daripada      | (from)          | 188       |
| rahmati       | (blessed)       | 179       |
| kerana        | (because)       | 174       |
| bagi          | (give)          | 171       |
| muslimin      | (muslimin)      | 170       |
| dapat         | (got)           | 161       |
| surah         | (surah)         | 155       |
| dunia         | (world)         | 154       |
| ada           | (have)          | 154       |
| maksudnya     | (meaning)       | 153       |
| ke            | (to)            | 152       |
| agama         | (religion)      | 151       |
| satu          | (one)           | 150       |
| lain          | (other)         | 147       |
| firman        | (decree)        | 140       |
| sesungguhnya  | (indeed)        | 138       |
| ia            | (it)            | 137       |
| lagi          | (again)         | 134       |
| Jumaat        | (Friday)        | 133       |
| hidup         | (live)          | 131       |
| akhirat       | (afterlife)     | 128       |
| seperti       | (such as)       | 127       |
| masa          | (time)          | 127       |
| perkara       | (matter)        | 124       |
| ialah         | (is)            | 123       |
| lebih         | (more)          | 123       |
| kehidupan     | (livelihood)    | 121       |
| bulan         | (month)         | 117       |
| sentiasa      | (always)        | 117       |
| masjid        | (mosque)        | 115       |
| antara        | (between)       | 114       |
| baik          | (good)          | 113       |
| setiap        | (every)         | 113       |
| apa           | (what)          | 109       |
| memberi       | (give)          | 108       |
| sidang        | (assembly)      | 107       |
| agar          | (so that)       | 106       |
| tentang       | (about)         | 105       |
| melakukan     | (perform)       | 103       |
| harta         | (wealth)        | 102       |

The two most spoken words ('*dan*' and '*yang*') are monosyllable and they are neither verb nor noun. The third most spoken word, '*Allah*', is spoken a lot mainly because the samples are from Friday sermons whose audience are all Muslims. The fifth most spoken word ('*kita*') is particularly interesting because it tells us that Friday sermons, at least those that are spoken in Kuala Lumpur, urges the congregates to see themselves as a Muslim collective rather than an individual. The word ('*negara*') that is spoken for 251 times implies that the Friday sermons in Malaysia are also used to unite the people in the country.

**Table 2** The list of words obtained from 52 Friday sermons arranged based on the spoken frequency from 98 to 61

| Words (Malay) | Words (English) | Frequency |
|---------------|-----------------|-----------|
| khutbah       | (sermon)        | 98        |
| saya          | (I)             | 97        |
| tinggi        | (height)        | 94        |
| jika          | (if)            | 94        |
| ketika        | (at)            | 93        |
| sekalian      | (every)         | 91        |
| boleh         | (able)          | 91        |
| masyarakat    | (community)     | 90        |
| amalan        | (practice)      | 89        |
| mengambil     | (to take)       | 87        |
| seorang       | (somebody)      | 87        |
| perlu         | (need)          | 85        |
| merupakan     | (constitute)    | 84        |
| secara        | (in)            | 84        |
| supaya        | (so that)       | 84        |
| hendaklah     | (need to)       | 84        |
| tetamu        | (guest)         | 84        |
| besar         | (great)         | 83        |
| melaksanakan  | (implement)     | 83        |
| syurga        | (heaven)        | 83        |
| menyeru       | (call)          | 82        |
| ibu           | (mother)        | 82        |
| atas          | (above)         | 81        |
| beriman       | (faith)         | 81        |
| bahasa        | (language)      | 80        |
| tetapi        | (but)           | 80        |
| alam          | (nature)        | 79        |
| tahun         | (year)          | 78        |
| kami          | (us)            | 77        |
| mudah         | (easy)          | 76        |
| baginda       | (sire)          | 75        |
| para          | (every)         | 74        |
| malam         | (night)         | 72        |
| iaitu         | (that is)       | 71        |
| rakyat        | (citizen)       | 69        |
| berada        | (at)            | 69        |
| mulia         | (holy)          | 69        |
| berlaku       | (happen)        | 68        |
| sisi          | (side)          | 68        |
| tersebut      | (mentioned)     | 67        |
| sesuatu       | (something)     | 67        |
| apabila       | (when)          | 66        |
| amat          | (very)          | 66        |
| nabi          | (prophet)       | 65        |
| semua         | (all)           | 65        |
| dia           | (he/she)        | 65        |
| bukan         | (not)           | 65        |
| riwayat       | (narration)     | 64        |
| mudahan       | (hopefully)     | 64        |
| perintah      | (order)         | 64        |
| sehingga      | (to)            | 64        |

| Words (Malay) | Words (English)      | Frequency |
|---------------|----------------------|-----------|
| benar         | ( <i>true</i> )      | 63        |
| bapa          | ( <i>father</i> )    | 63        |
| banyak        | ( <i>many</i> )      | 63        |
| ingin         | ( <i>want</i> )      | 63        |
| bermaksud     | ( <i>meaning</i> )   | 62        |
| pendidikan    | ( <i>education</i> ) | 61        |
| pihak         | ( <i>party</i> )     | 61        |
| menjaga       | ( <i>keep</i> )      | 61        |

Table 2 lists the words spoken in the range of 98 to 61 times. It seems that, the words are less monosyllabic and more disyllabic or trisyllabic as their spoken frequency become smaller. Table 3 shows the words that are spoken in the range of 60 to 52 times. It features more 4 or more syllabic words than the previous 2 tables.

**Table 3** The list of words obtained from 52 Friday sermons arranged based on the spoken frequency from 60 to 52

| Words (Malay) | Words (English)           | Frequency |
|---------------|---------------------------|-----------|
| sendiri       | ( <i>self</i> )           | 60        |
| tanggungjawab | ( <i>responsibility</i> ) | 60        |
| ketakwaanan   | ( <i>piety</i> )          | 59        |
| penuh         | ( <i>full</i> )           | 59        |
| bumi          | ( <i>earth</i> )          | 58        |
| mendapat      | ( <i>got</i> )            | 58        |
| usaha         | ( <i>effort</i> )         | 58        |
| terhadap      | ( <i>there</i> )          | 58        |
| tenaga        | ( <i>energy</i> )         | 57        |
| air           | ( <i>water</i> )          | 56        |
| pula          | ( <i>also</i> )           | 56        |
| pahala        | ( <i>reward</i> )         | 56        |
| seseorang     | ( <i>one</i> )            | 56        |
| ilmu          | ( <i>knowledge</i> )      | 56        |
| amanah        | ( <i>trust</i> )          | 56        |
| sebagaimana   | ( <i>as</i> )             | 55        |
| sahaja        | ( <i>only</i> )           | 55        |
| hanya         | ( <i>only</i> )           | 55        |
| ramadhan      | ( <i>ramadhan</i> )       | 55        |
| dirahmati     | ( <i>respected</i> )      | 55        |
| larangan      | ( <i>ban</i> )            | 54        |
| berpesan      | ( <i>instruct</i> )       | 54        |
| terdahulu     | ( <i>earlier</i> )        | 53        |
| dosa          | ( <i>sin</i> )            | 53        |
| melalui       | ( <i>through</i> )        | 53        |
| membawa       | ( <i>lead</i> )           | 53        |
| cara          | ( <i>way</i> )            | 52        |
| meninggalkan  | ( <i>leave</i> )          | 52        |
| tempat        | ( <i>place</i> )          | 52        |

From these 3 tables, it can be seen that Malay language mainly consists of disyllabic words and trisyllabic words. Only 5% words from the results are monosyllables. About 50% are disyllabic. Trisyllabic words consist of 36% and polysyllabic

words (words that contain 4 or more syllables) occupy 10 % of the result.

Even though the specific meanings of the words are not important in developing SI assessors, the result shows what the Malaysian government considers important in a Friday sermon. For example, the word '*anak*' which is spoken for 219 times; this implies that one of the main concerns of the government is to ensure that the younger generation respect their parents and thus strengthening family relation. The word '*akhirat*' is spoken for 128 times which is lower than '*anak*', this shows that our government is less inclined to remind the congregations of afterlife than reforming young people.

These words are then arranged according to the range of frequency it occupies. Table 4 shows the arrangement of these words.

**Table 4** The frequency of words arranged according to certain range and the number of words for each range

| Spoken frequency range | Number of words | Percentage (%) |
|------------------------|-----------------|----------------|
| 1001-2000              | 3               | 2              |
| 901-1000               | 2               | 1              |
| 701-900                | 3               | 2              |
| 501-700                | 0               | 0              |
| 401-500                | 6               | 4              |
| 301-400                | 5               | 3              |
| 201-300                | 19              | 11             |
| 101-200                | 45              | 26             |
| 91-100                 | 7               | 4              |
| 81-90                  | 17              | 10             |
| 71-80                  | 10              | 6              |
| 61-70                  | 25              | 15             |
| 52-60                  | 29              | 17             |
| <b>TOTAL</b>           | <b>171</b>      | <b>100</b>     |

From Table 4, it can be seen that the number of words whose frequency ranges from 52-100 is 88. The number of words in the range of 101-200 is 45; almost half of the number of words in the previous range. The number of words that have frequency in the range of 201-300 is 19, which half the number of the previous range.

**Table 5** The number of words based on the frequency range and the number of syllables

| Spoken frequency range | Syllables   | Number of Words |
|------------------------|-------------|-----------------|
| 1001-2000              | 1           | 3               |
|                        | 2           | 1               |
|                        | 3           | 0               |
|                        | 4 and above | 0               |
| 101-1000               | 1           | 2               |
|                        | 2           | 48              |
|                        | 3           | 23              |
|                        | 4 and above | 5               |
| 52-100                 | 1           | 0               |
|                        | 2           | 37              |
|                        | 3           | 38              |
|                        | 4 and above | 13              |

The distribution of words also depends on the frequency range as shown in Table 5. Disyllabic and trisyllabic words, for instance, only appear in the frequency range of 52-1000. And generally speaking, the least spoken word is, said words contain more syllables.

#### 4.0 FUTURE WORKS

Future works are still required to establish a speech intelligibility to be used solely for Malay language. The test will be used to study the effect of reverberant sound on spoken Malay words. Also, as pointed out by [17], adults are more influenced by word frequency compared to children. Thus, the data here can be used to determine what words are more appropriate to be used in a Friday sermon.

#### 5.0 CONCLUSION

The analysis of 52 Friday sermon transcripts has revealed that there are total of 26244 words in the sermon transcripts and of those words, there are 5937 distinct words. The highest spoken frequency of Bahasa Melayu words is 2509 whereas the second highest is the word 'yang' which is repeated 1969. As shown in Table 4, the highest spoken frequency of Bahasa Melayu words is in the range of 101-200 which constitutes 26% of the total 171 words. The frequency range between 301-2000 doesn't have many words which suggest that the sermon contains variety of words.

#### Acknowledgement

This research is supported by the Research University Grant (RUG) by Universiti Teknologi Malaysia (VOT no. 04J30).

#### References

- [1] Yap, J. M. and Balota, D. A. 2009. Visual Word Recognition of Multisyllabic Words. *Journal of Memory and Language*. 60: 502-5.
- [2] Balota, D. A., Sergent-Marshall, S. D., Cortese, M. J. and Spieler, D. H. 2004. Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*. 133(2): 283-316.
- [3] Jescheniak, J. D. and Levelt, W. J. M. 1994. Word-Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 20: 824-843.
- [4] Thorndike, E. L. and Lorge, I. 1944. *The Teacher's Word Book of 30,000 Words*. New York: Columbia University: Teachers College.
- [5] Brysbaert, M. and New, B. 2009. Moving Beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*. 41: 977-990.
- [6] Baayen, R. H., Piepenbrock, R. and Gulikers, L. 1995. The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium. University of Pennsylvania, Philadelphia.
- [7] New, B., Pallier C., Brysbaert, M. and Ferrand, L. 2004. Lexique 2: A New French Lexical Database. *Behavior Research Methods, Instruments and Computers*. 36(3): 516-524.
- [8] Fogerty, D. and Humes, L. E. 2012. The Role of Vowel and Consonant Fundamental Frequency, Envelope, and Temporal Fine Structure Cues to the Intelligibility of Words and Sentences. *Journal of the Acoustical Society of America*. 131(2): 1490-1501.
- [9] Gooskens, C., van Heuven, V. J., van Bezooijen, R. and Pacilly, J. J. A. 2010. Is Spoken Danish Less Intelligible than Swedish? *Speech Communication*. 52(11-12): 1022-1037.
- [10] Wong, L. L. N., Ho, A. H. S., Chua, E. W. W. and Soli, S. D. 2007. Development of the Cantonese speech intelligibility index. *Journal of the Acoustical Society of America*. 121(4): 2350-2361.
- [11] Granlund, S., Hazan, V. and Baker, R. 2012. An Acoustic-phonetic Comparison of the Clear Speaking Styles of Finnish-English Late Bilinguals. *Journal of Phonetics*. 40(3): 509-520.
- [12] Ahmad Khan Said, Mohamad Ngasri Dimon, and Mokhtar Harun. 2001. *Measurements of A-Weighted SPL, RASTI and Alcons of Prayer Halls in Masjid Wilayah Persekutuan Kuala Lumpur*. Technical Report. UT.CC-CONS-2.224 (58).
- [13] Broersma, M. and Scharenborg, O. 2010. Native and Non-native Listeners' Perception of English Consonants in Different Types of Noise. *Speech Communication*. 52: 980-995.
- [14] Meyer, B. T., Jurgens, T., Wesker, T., Brand, T., & Kollmeier, B. 2010. Human Phoneme Recognition Depending on Speech-intrinsic Variability. *The Journal of the Acoustical Society of America*. 128(5): 3126-3141.
- [15] Sommerhoff, J. and Rosas, C. 2012. Logatom Corpus for the Assessment of the Intelligibility in Spanish Speaking Environments and Its Relation with STI Measurements. *Applied Acoustics*. 73: 1190-1200.
- [16] Nielsen, J.B. and Dau, T. 2009. Development of a Danish Speech Intelligibility Test. *International Journal of Audiology*. 48: 729-741.
- [17] Tian, K. Q., Mukari, S. Z. M. S., Wahab, N. A. A., Razak, R. A., Omar, M. and Maamor, N. 2008. Malay Hearing in Noise Test. *International Journal of Audiology*. 47: 379-380.
- [18] Hällgren, M., Larsby, B. and Arlinger, S. 2006. A Swedish Version of the Hearing In Noise Test (HINT) for Measurement of Speech Recognition. *International Journal of Audiology*. 45(4): 227-37.
- [19] Nilsson, M., Soli, S.D. and Sullivan, J. A. 1994. Development of the Hearing In Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise. *The Journal of the Acoustical Society of America*. 95(2): 1085-1099.
- [20] Palaz, H., Bicil, Y., kanak, A. and Dogan, M. U. 2005. New Turkish Intelligibility Test for Assessing Speech Communication Systems. *Speech Communication*. 47(4): 411-423.
- [21] ANSI. 1989. *Method for Measuring the Intelligibility of Speech over Communication Systems*. United States, S3.2.
- [22] Kollmeier, B. and Wesselkamp, M. 1997. Development and Evaluation of a German Sentence Test. *The Journal of Acoustical Society of America*. 102(4): 2412-2421.
- [23] Tadmoor, U. 2009. *The World's Major Languages*. 2nd ed. B. Comrie, ed. Abingdon, Oxon: Routledge.
- [24] Karim, N. S. 1995. *Tatabahasa Dewan Edisi Baharu*. Kuala Lumpur, Dewan Bahasa dan Pustaka.
- [25] Teoh, B. S. 1994. *The Sound System of Malay Revisited*. Kuala Lumpur, Dewan Bahasa dan Pustaka.