# Evaluation of the Nato Spelling Alphabet[1]

Alistair D N Edwards

Department of Computer Science
University of York
York
YO10 5GH

alistair.edwards@york.ac.uk

It is important – sometimes imperative – that speech radio communication is unambiguous and clear. While modern technologies have reduced the problems of noise, it is still often the case that a message can be garbled. In some instances it is necessary to spell out words, but there are also many instances when the information to be communicated consists largely of letters, notably in callsigns. It is not safe to use the conventional letter names ('Ay', 'Bee', 'Sea' etc) because they are too easy to confuse; think of 'Bee' versus 'Pea' for instance. These sounds are too short and simple. Instead it is conventional to use a so-called *phonetic alphabet* to represent the letters being spoken, usually *Alfa, Bravo, Charlie etc.* We use words because they are more easily distinguished from each other; they are longer and have greater variation. A question is why we use that particular set of words and are they optimal in terms of being distinct and proof against ambiguity?

The alphabet used in telecommunications is generally known as the *Nato Phonetic Alphabet*[2], and is listed in Table 1. This is by no means the only such alphabet that has been devised over time, but it is the international standard. It was first published by the International Civil Aviation Organization (ICAO) in 1959 and adopted by Nato.

[1] An extended version of this article, *Testing the November Alfa Tango Oscar Spelling Alphabet*, is available online, at https://www-users.york.ac.uk/~ade1/research/Alphabets/

[2] Although this is the common name for this alphabet, strictly speaking it is *not* a phonetic alphabet. As discussed below, the International Phonetic Alphabet (IPA) is a phonetic alphabet: it consists of symbols to represent the phoneme sounds of languages. A better name is the Nato Spelling Alphabet, and that is what we refer to hereafter in this article.

| Letter | Nato code word | Nato pronunciation | IPA representation |
|---|---|---|---|
| A | Alfa | AL FAH | ælfə |
| B | Bravo | BRAH VOH | brɑvoʊ |
| C | Charlie | CHAR LEE | ʧɑrli |
| D | Delta | DELL TAH | dɛltə |
| E | Echo | ECK OH | ɛkoʊ |
| F | Foxtrot | FOKS TROT | fɑkstrɑt |
| G | Golf | GOLF | gɑlf |
| H | Hotel | HOH TELL | hoʊtɛl |
| I | India | IN DEE AH | ɪndiə |
| J | Juliett | JEW LEE ETT | dʒuliet |
| K | Kilo | KEY LOH | kɪloʊ |
| L | Lima | LEE MAH | laɪmə |
| M | Mike | MIKE | maɪk |
| N | November | NO VEM BER | noʊvɛmbər |
| O | Oscar | OSS CUR | ɔskər |
| P | Papa | PAH PAH | pɑpə |
| Q | Quebec | KEH BECK | kwəbɛk |
| R | Romeo | ROW ME OH | roʊmioʊ |
| S | Sierra | SEE AIR RAH | siɛrə |
| T | Tango | TANG GO | tæŋgoʊ |
| U | Uniform | YOU NEE FORM | junəfɔrm |
| V | Victor | VIK TAH | vɪktər |
| W | Whiskey | WISS KEY | wɪski |
| X | X-ray | ECKS RAY | ɛksreɪ |
| Y | Yankee | YANG KEY | jæŋki |
| Z | Zulu | ZOO LOO | zulu |

Table 1. The Nato spelling alphabet. The words to be used are listed in the second column. There is a prescribed pronunciation for each of them, listed next (the syllable to be accented being underlined). Finally the pronunciation is shown using the IPA phonetic symbols, as described below. Note that one allowed variation is that in countries where the drinking of alcohol is strongly discouraged an alternative W-word for *Whiskey* is allowed.

That all the words in the alphabet are common English words should not be surprising, since English is the official language of radio communication. At the same time, though, Nato is an international organization, many of the users of the alphabet will have other first languages, but the alphabet should be easily accessible to them. This also explains the unconventional spelling of some of the words. For instance, it is necessary to know that in English ‘ph’ is sounded as an *f* sound if *Alpha* is spelled in the conventional orthography

There are a number of desirable properties of such an alphabet. One of them is that the words should be distinct, and particularly not easily mistaken one for another. For instance,

an earlier version of the ICAO alphabet used the word ‘Coca’ to represent *C*, but it was found that *Coca* and *Echo* were frequently confused. So, how effective is this alphabet? How well are its elements distinguished from each other? That is the main focus of this article.

## Phonetics

Phonetics is the science of the *sounds* of language. While we use the 26 letters of the alphabet to spell all English words, there are rather more sounds in the language, English being a very unphonetic language. There are, in fact, about 44 distinct sounds – or *phonemes* – in English. We can nevertheless spell them all because the language uses letter combinations and conventions to represent the additional sounds, such as ‘sh’, ‘ee’ and ‘th’.

A simple definition of a phoneme is a unit of a word which it if were replaced by another phoneme the meaning of the word would be changed. For instance, if one replaces the 't' sound at the beginning of the word 'toffee' with a 'k' sound then the word would become 'coffee'.

There is a notation which uniquely represents all of the sounds of English (and more from other languages), which is the International Phonetic Alphabet, or IPA. Many of the symbols of the IPA resemble letters in conventional alphabets (e.g. a, b and c) – and variations thereon (such as upside-down letters ɐ, ə, ʌ). It is thus possible to take an English word and represent its sound uniquely as a set of IPA symbols. (There is more information about the IPA and the pronunciation of its symbols on https://www.dictionary.com/e/key-to-ipa-pronunciations/.)

The final column of Table 1 shows the IPA representation of each of the Nato code words. This translation has been derived from the website https://tophonetics.com/.

## Edit distances

A collection of characters (such as a word) can be referred to in the abstract as a ‘string’. It is possible to calculate the degree of difference between two strings, by measuring the *edit distance*. The edit distance between two strings is really a count of the number of single-character editing operations (insertions, deletions or substitutions) that would be required to transform one string into the other. Taking the earlier example, the edit distance between *toffee* and *coffee* is 1, since it takes just one substitution to make that transformation. The edit distance between *toffee* and *free* is 3: two deletions (*t* and *o*) and one substitution (*r* for *f*).

The edit distance between two strings is thus a measure of their similarity. A short edit distance implies that the two strings are similar (*toffee* and *coffee*), but a long edit difference means that the two are more different (*toffee* and *free*).

## A study

It is thus possible to measure the distinctiveness of the words in the Nato alphabet by obtaining their IPA ‘spellings’ and then measuring the distance between each pair of words. Any pairs with short edit distances are the most likely to be confused for each other, while long edit distances imply good distinguishability.

A simple, if quite crude, measure of the overall quality of the alphabet is the modal edit distance. In this case the median is 6. Those pairs with the shortest edit distances are the most likely to be mistaken for each other. These all have an edit distance of 3 and are listed in Table 2[3]. Note that the potential confusions are symmetrical. That is to say that it is equally likely that a listener would mis-hear 'Alfa' as 'Golf' or 'Golf' as 'Alfa' (at least according to the edit distance).

| | | | | |
|---|---|---|---|---|
| 1 | A | Alfa | P | Papa |
| 2 | A | Alfa | G | Golf |
| 3 | G | Golf | P | Papa |
| 4 | G | Golf | Z | Zulu |
| 5 | L | Lima | M | Mike |
| 6 | W | Whiskey | Y | Yankee |

Table 2. All the word pairs with the shortest edit distance (3). Note that the edit distance is symmetrical. Thus the edit distance for Alfa/Papa is the same as for Papa/Alfa, implying that they are equally likely to be confused for each other. (A complete listing of all the edit distances can be found in the extended, online version of this paper.)

Conversely, those with the longest edit distance (9) are least likely to be confused, and these are all that distinct from *November*. They are listed in Table 3.

[3] There is no need to examine the measurement of all of these edit distances here. It is probably quite evident that, for instance, the IPA representation of A or Alfa is *ælfə* while G-Golf is gɑlf, and these 'spellings' share just one symbol (f) and hence *three* substitutions would have to be made to transform one into the other. The edit distance used in this study is the Levenshtein Distance (Black & Paul, 2008).

| | | |
|---|---|---|
| 1 | A | Alfa |
| 2 | C | Charlie |
| 3 | D | Delta |
| 4 | F | Foxtrot |
| 5 | G | Golf |
| 6 | J | Juliett |
| 7 | L | Lima |
| 8 | M | Mike |
| 9 | O | Oscar |
| 10 | P | Papa |
| 11 | S | Sierra |
| 12 | V | Victor |
| 13 | W | Whiskey |
| 14 | X | X-ray |
| 15 | Y | Yankee |
| 16 | Z | Zulu |

Table 3. The words with the longest edit distance (9) from *November*.

*November* is evidently very distinct. This is interesting because in an earlier version of the ICAO alphabet *N* was represented by *Nectar*, but this was easily confused with *Victor* with an edit distance of 2. Statistically *November* is the most distinct of all the words, with a median edit distance of 9. In a sense this is not surprising. Looking at it acoustically, *November* has 3 syllables and so greater redundancy in its pronunciation. Looking at it in terms of edit distance, its IPA representation (noʊvɛmbər) consists of 10 symbols, more than any of the other words. Editing from one of the shorter words is going to involve at least several deletions.

A simple way of comparing different alphabets is to take the median average of all the edit differences. For the Nato alphabet this is 6, and that is at least one more than for nearly all of the other alphabets investigated[4]. Another useful statistic is the number of times the shortest edit distance occurs, in other words how many pairs are most likely to be confused. For the Nato alphabet the minimum distance is 3 and that occurs 12 times[5].

One comparison would be with the conventional pronunciations of the letters. If we did not bother with *Alfa, Bravo, Charlie,* but used *Ay, Bee, Sea etc* would there be more errors? This analysis suggests there would, as summarized in Table 4. The median for the conventional pronunciation is 2, which implies poor overall distinctiveness, and a minimum edit distance of 1 implies very similar-sounding words, and there are as many as 80 occurrences.

[4] The exception is an alphabet devised by the International Telecommunication Union (ITU), in 1926. This has a median edit distance of 8 and a minimum distance of 5. However, it was deemed unsuitable for adoption by the ICAO for other reasons. ICAO (1959, p.9) states, '[O]perating experience has indicated that the words were unsuitable because they were unusual in everyday language and because they lacked desirable phonetic qualities'. Furthermore, all the words were place names (*Amsterdam, Baltimore, Canada,…*) which might cause pronunciation problems for non-native English speakers.

[5] Note that occurrences of short lengths are effectively double counted. That is because the distance (for instance) between *Alfa* and *Golf* is the same as the distance between *Golf* and *Alfa*. This makes sense because if *Alfa* may be confused with *Golf*, then *Golf* can be mistaken for *Alfa*.

| Alphabet | Edit Distances | | |
|---|---|---|---|
| | Median | Minimum (occurrences) | Maximum (occurrences) |
| Nato | 6 | 3 (12) | 9 (32) |
| Conventional | 2 | 1 (80) | 7 (48) |

Table 4. A comparison of the edit distances for the Nato alphabet compared to the conventional pronunciation of letters.

## Discussion

According to this study, the established Nato Alphabet does seem to be quite robust. Would it be possible to devise a better one, one with fewer potential confusions (i.e. all with longer edit distances than 3)? Certainly it would be possible to find alternatives to the letters in Table 2, for instance, but there would be a number of consequences.

Firstly, the very fact that this alphabet is standardized is very powerful. All users know it. To introduce another one would be confusing. And having a limited vocabulary is useful: when a receiver is expecting to hear a letter they know that they can expect to hear only one of the 26 words in the vocabulary.

A memo on this point from Nato (Nato, 1955) suggested another hazard in trying to devise a better alphabet:

> *One of the firmest conclusions reached was that it was not practical to make an isolated change to clear confusion between one pair of letters. To change one word involves reconsideration of the whole alphabet to ensure that the change proposed to clear one confusion does not itself introduce others.*

Or, as ICAO (1959, pp.14-15) puts it, rather more graphically:

> *The problem is not unlike that of pushing a dent out of a child's celluloid ball – even a successful push leaves a small dent in another place.*

Finally, if one were to find more-distinct alternative words letters in Table 2 they would be likely to be longer, to have more syllables, which would then slow down the rate of communication.

To conclude we have demonstrated that one way and another the Nato alphabet is about as good as could be expected, and it is likely to remain in use.

## References

*Black, Paul E., ed. (14 August 2008), "Levenshtein distance", Dictionary of Algorithms and Data Structures [online], U.S. National Institute of Standards and Technology, retrieved 2 November 2016*

ICAO (1959) *The Evolution and Rationale of the ICAO (International Civil Aviation Organization) Word-Spelling Alphabet*, July (PDF). Archived (PDF) from the original on 2016-03-10. Retrieved November 1, 2017.

Nato (1955) *SGM-675-55: Phonetic Alphabet for NATO Use* Archived (PDF) from the original on 12 April 2018.

## Acknowledgement

Thanks to Leandro (G5LSI) for the encouragement and comments on early drafts of this article.