

# The Balancing of Vowel Inventories

## Introduction

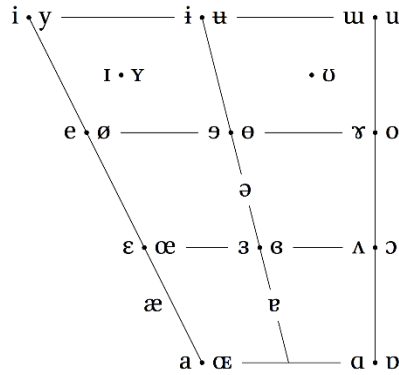


Figure 1 – The IPA Vowel Chart

A vowel is defined by Merriam-Webster as “one of a class of speech sounds in the articulation of which the oral part of the breath channel is not blocked and is not constricted enough to cause audible friction (Vowel).” There are over 7000 confirmed languages discovered so far on the planet, and every single one of them contains vowels (How Many Languages). All of the vowels of a language make up said language’s *vowel inventory*. A vowel inventory can range from two vowels, like in the language Cuvok, to fourteen, like in Standard German. While this is a relatively large range of sounds, the average number of vowels in a language is 6.64, or between six and seven, shown in Figure 2. Languages like Hebrew, Macedonian, Italian, and Nepali all fall within a deviation of  $\pm 1$  of this average.

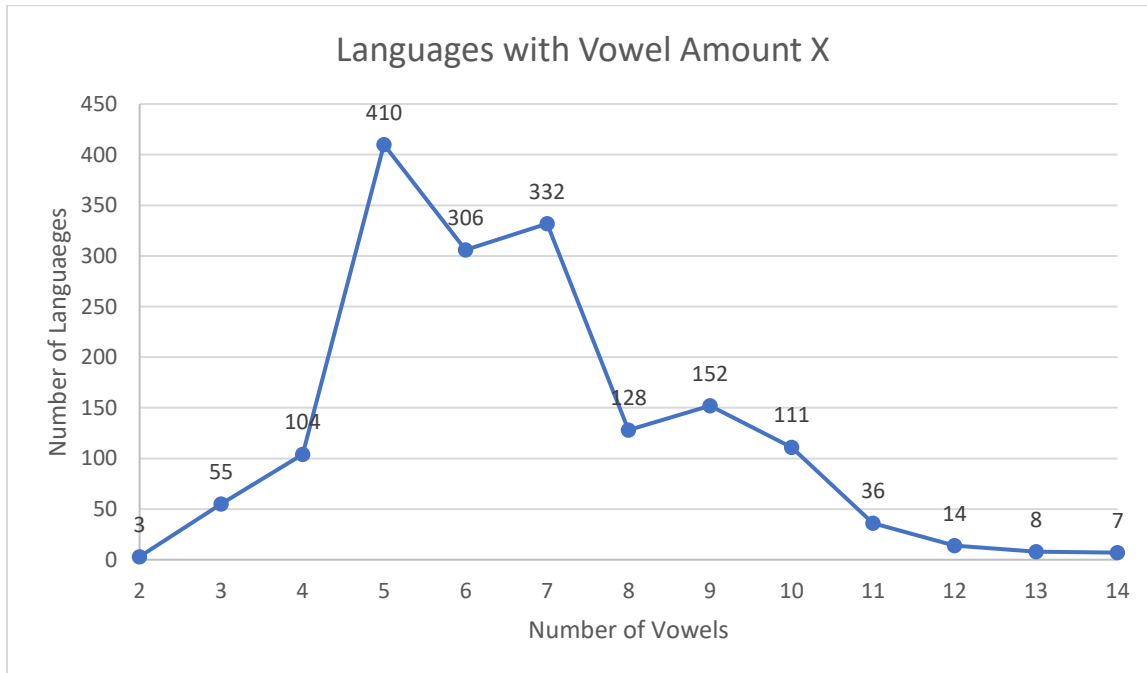


Figure 2 – Amount of Languages with Vowel Amount X

This paper will discuss possible well-balanced vowel inventories for every vowel amount from two to fourteen. There are, however, a few caveats. First, this paper does not take into account variations of a vowel. If a language has /u u: ù/, only /u/ will be recorded, as variations are not important for this paper. Also, if a language only has /ù/, it will still be recorded as its base vowel, /u/. Second, while there are 7000+ documented languages, this paper only covers the 1672 languages found on PHOIBLE, which is a self-described “repository of cross-linguistic phonological inventory data, which have been extracted from source documents and tertiary databases and compiled into a single searchable convenience sample. (PHOIBLE).”

Before the analysis, we will first look at what a well-balanced vowel inventory looks like. As the term implies, a balanced inventory has the same or almost the same number of vowels on either side of a central line, which represents central vowels and the center of the mouth. Figure 3 shows a balanced and unbalanced inventory of three vowels (3V). The inventory on the left is balanced because, if one were to draw a line vertically down the middle, there would be the same number of vowels on either side. On the contrary, if the same line were drawn on the inventory on the right, there would be more vowels on the left than on the right side of the line. Something to keep in mind, though, is that not all inventories are balanced. Alawa, a language native to Australia, has no back vowels, as shown in Figure 4. Figure 5 shows the balanced inventory of Sinhala, a language of Sri Lanka. Sinhala has ten basic vowels and is balanced.

i		u
	a	

i		u
e		

Figure 3 – 3V

	ɪ	
e		
	a	

Figure 4 – Alawa

i		ɔ u
ɪ		
e		o
	ə	
æ	a	ɑ

Figure 5 – Sinhala

Another concept to focus on when balancing is that of roundness. Almost all of the documented vowels have a rounded counterpart, with exceptions being /ɔ æ ə ɐ/. Excluding these four vowels, there are still twelve vowels on the IPA chart with rounded counterparts. It is much more common for a language to have both rounded and unrounded vowels than only one form of rounding. It is, however, almost never the case that there is an even amount of rounded and unrounded vowels in the same language. Referring back to Figure 3, the inventory on the left, while balanced, is not roundness balanced (RB). This particular inventory contains two unrounded vowels and only one rounded vowel, which is perfectly okay and natural.

For this paper, there are several abbreviations to be aware of. Using the example *A10a*, *A* stands for *average*, or *analyzed*, which is just the most common inventory among all of the languages in the dataset with *x* number of vowels, which in this example is represented by the *10*. Instead of *A*, there may also be a *P*, which stands for *perfectly balanced*, and is used for inventories the writer as altered from the *A-inventory* to have the same amount of front and back vowels. Lastly, *a* represents what iteration of the inventory is being represented in a figure. In the case of an *A-inventory*, every letter represents another data-derived inventory. In the case of a *P-inventory*, every letter represents another possible option for a balanced inventory of *x* number of vowels.

**Non-standard terms used**

Outskirts – the non-central vowels (i a u o e ε ɔ α) and their opposite-rounding counterparts

Primary vowel – one of the five most common vowels (i a u o e)

**14 Vowels**

The analysis will start with the largest vowel inventory amount, fourteen, and work its way down, as there is less variation in the larger inventories. Of the 1600+ languages used in the study, only seven of them, or 0.42% of them, contain fourteen distinct vowels. As mentioned earlier that the average is between six and seven vowels, it is understandable that there would be so few with this many.

Upon comparison and analysis of these seven languages, the following inventory is the most common, balanced inventory for fourteen vowels.

i y				u
	ɪ ʏ		ʊ	
e ø				o
ε œ		ə		ɔ
		a		

*Figure 6 – A14*

As mentioned previously, it is not uncommon for a language to have more unrounded than unrounded vowels or vice versa. For a perfectly balanced 14V inventory, one can remove /y ø/ or /y œ/ or and replace them with /u ʌ/ or /u ɤ/ respectively. Figures 7 and 8 show these two perfectly balanced 14V inventories.

i y				ʉ u
	ɪ		ʊ	
e				o
ε œ		ə		ʌ ɔ
		a		

*Figure 7 – P14a*

i y				ɯ u
	ɪ		ʊ	
e ø				ɤ o
ɛ		ə		ɔ
		a		

Figure 8 – P14b

Notice how in Figures 7 and 8 there are an even number of vowels on both sides of the center.

### 13 Vowels

Having only one more language than the 14V collection, eight of the 1600+ languages, or 0.48%, contain thirteen distinct vowels. The wedge (/ʌ/) is technically a mid back unrounded lax vowel, the unrounded equivalent of /ɔ/. However, in some languages the wedge is treated as a stressed version of schwa (/ə/). Therefore, the wedge is more centralized than the rounded counterpart.

ɪ		ɪ		u
	ɪ		ʊ	
e				o
ɛ		ə ʌ		ɔ
a				ɑ

Figure 9 – P13a

The biggest difference between a 13V and 14V inventory is that a 13V inventory contains both a front and back low vowel, whereas 14Vs only have one of them.

ɪ		ɪ		u
	ɪ		ʊ	
e				o
ɛ œ				ʌ ɔ
a				ɑ

Figure 10 – P13b

Should the wedge not be centralized, there would be more back vowels than front vowels. Therefore, schwa can be replaced with /œ/, like in Figure 10, to restore balance.

ɪ		ɪ		u
	ɪ		ʊ	
e				o
ɛ œ		ə		ʌ ɔ
		a		

Figure 11 – P13c

If there is a want for whatever reason for a centralized vowel, the two low vowels can be replaced with a single, low central vowel, and the central vowel can be kept, as in Figure 11. There are more variants of 13Va and 13Vb. The vowels /œ ʌ/ can be lowered to /æ ɒ/, or raised to /ø ɤ/ or /y ɯ/.

## 12 Vowels

Doubling the amount of 14V languages, there are fourteen 12V languages, or 0.84% of the languages in the dataset.

i		ɨ		u
	ɪ		ʊ	
e				o
ɛ		ə ʌ		ɔ
		a		

*Figure 12 – A12*

The only difference between the standard 13V and the standard 12V is that a 12V language only has one low vowel instead of two, which is similar to the formation of 13Vb. Again, as mentioned in the 13V section, the wedge (/ʌ/) is technically a mid back unrounded lax vowel, the unrounded equivalent of /ɔ/, but some languages the wedge is treated as a stressed version of schwa (/ə/). And again, the central vowels /ə ʌ/ can move apart and become the opposite-rounded counterparts of the non-central vowels.

## 11 Vowels

At 2.16%, 11V inventories are much more common than larger inventories. Of the thirty-six languages from the dataset, Figure 13 shows the average 11V inventory.

i				u
	ɪ		ʊ	
e				o
ɛ		ə		ɔ
a				ɑ

*Figure 13 – A11a*

Contrary to the three previous inventories, 11V languages will on average not have rounded/unrounded counterparts of the same sound; only one sound appears per box. It is interesting to note that, while Figure 13 is the most common 11V inventory, Figure 14 shows an inventory used by only slightly less languages.

i		ɨ		u
	ɪ		ʊ	
e				o
ɛ				ɔ
a				ɑ

*Figure 14 – A11b*

Unlike vowels on the outskirts of the vowel chart, central vowels are much more volatile and likely to change.

i		i		u
	ɪ		ʊ	
e				o
ɛ		ə		ɔ
		a		

Figure 15 – P11

If one does not like to differentiate between low vowels, Figure 15 has an alternative formation. This figure has both a high and a mid central vowel, and also a low central vowel instead of a front and back coupling.

### 10 Vowels

A 10V inventory is very similar to that of an 11V. The only difference is that, while the standard 11V contains two low vowels, the standard 10V contains only one, as seen in Figure 16. While this is a minor difference, there exists 111 languages, or 6.66% of the total, with ten vowels, completely overshadowing the thirty-six with eleven.

i				u
	ɪ		ʊ	
e				o
ɛ		ə		ɔ
		a		

Figure 16 – A10a

i				u
	ɪ		ʊ	
e				o
ɛ		ə		ɔ
				ɑ

Figure 17 – A10b

Given that 10V languages don't, on average, contain both a front and back low vowel, there is more freedom to alternate between the two positions. Figures 16 and 17 show the same 10V inventory, but with alternating low vowels. The latter figure is not perfectly balanced, but as mentioned earlier, languages are rarely perfectly balanced.

i		i		u
	ɪ		ʊ	
e				o
ɛ				ɔ
		a		

Figure 18 – A10c

i		i		u
	I		U	
e				o
ɛ				ɔ
				ɑ

Figure 19 – A10d

Figures 18 and 19 follow the same pattern as Figures 16 and 17, but with a high central vowel instead of a mid central vowel.

### 9 Vowels

At 9.12%, the 152 languages with nine vowels represent almost one tenth of the total languages in the study.

i				u
	I		U	
e				o
ɛ				ɔ
		a		

Figure 20 – A9

Not a particularly interesting vowel inventory, but there are still some variations that can happen.

i				u
e ø				ɤ o
ɛ				ɔ
		a		

Figure 21 – P9a

Figure 21 shows a 9V inventory with rounded tense mid vowels instead of high lax vowels.

i				u
e				o
ɛ œ				ʌ ɔ
		a		

Figure 22 – P9b

Figure 22 shows a 9V inventory with rounded mid lax vowels instead of high lax vowels.

i		i		u
e				o
ɛ		ə		ɔ
		a		

Figure 23 – P9c

Figure 23 loses the high lax vowels in lieu of the mid and high central vowels.

## 8 Vowels

Appearing less often than 9V inventories, the 128 languages with eight vowels make up 7.68% of the languages. Languages with eight vowels include Malayalam, Portuguese, and Slovenian.

i				u
e				o
ɛ		ə		ɔ
		a		

*Figure 24 – A8*

As in P9c, P8 contains six non-central vowels, but only contains two central vowels.

i		i		u
e				o
ɛ				ɔ
		a		

*Figure 25 – P8a*

As seen in variations of other vowel inventories in this paper, the mid central vowel can be swapped out for the high central vowel instead.

i				u
e				o
ɛ				ɔ
a				ɑ

*Figure 26 – P8b*

Figure 26 is an interesting inventory. While it is perfectly balanced, there are no central vowels. It is very rare for a language to have both front and back vowels but no central vowels. It is recommended to avoid this type of inventory construction. If the two low vowels are desired, the most likely change would be the mid lax vowels centralizing into the mid lax vowel and the wedge, shown in Figure 27.

i				u
e				o
		ə ʌ		
a				ɑ

*Figure 27 – P8c*



## 7 Vowels

Having the third most common number of vowels in a vowel inventory are those languages with seven vowels. These 332 languages, making up 19.93% of the dataset, include Bengali, Italian, and Nepali.

i				u
e				o
ɛ				ɔ
		a		

*Figure 28 – A7*

Figure 28 is a very stable, very well balanced vowel inventory, and is the average vowel inventory for a 7V language. Again, if there is a desire for two low vowels, the mid lax vowels can be replaced with a central lax vowel, as in Figure 29.

i				u
e				o
		ə		
a				ɑ

*Figure 29 – A7a*

i		i		u
e				o
a				ɑ

*Figure 30 – A7b*

Figure 30, called the “upside down ‘u’”, replaces the mid central vowel from Figure 29 with a high central vowel.

## 6 Vowels

With 306 languages, 6V inventories are only slightly less abundant than 7V inventories. In the average 6V inventory, there are two vowels in the three main variations of backness. There are, however, two different standard 6V inventories. While they do not differ in backness of vowels, they do differ in height.

i		i		u
e				o
		a		

*Figure 31 – A6a*

Figure 31 shows the most common 6V inventory, equipped with the five primary vowels and the high central vowel.

i				u
e				o
		ə		
		a		

*Figure 32 – A6b*

Figure 32 has another, albeit less common, average 6V inventory. In A6b, the high central vowel is replaced with the mid central vowel.

	ɪ		ʊ	
e				o
		ə		
		a		

*Figure 33 – P6a*

Figures 33-36 show several other balanced, but not necessarily naturalistic 6V inventories. Figure 34 replaces the tense high vowels with high lax vowels.

		ɨ		
	ɪ		ʊ	
e				o
		a		

*Figure 34 – P6b*

Figure 34 is the same as Figure 33, except the mid central vowel has been replaced with a high central vowel.

i				u
	ɪ		ʊ	
		ə		
		a		

*Figure 35 – P6c*

Figure 35, referred to by the author as the “flux capacitor”, has replaced the tense mid vowels with the high lax vowels.

i		i		u
	ɪ		ʊ	
		a		

Figure 36 – P6d

Figure 36 is the same as Figure 35, except with the mid central vowel instead of a high central vowel.

## 5 Vowels

5V inventories are by far the most common type of inventory. With 410 languages, 5V inventories make up 24.61% of all of the languages used in the dataset. And since it is the most common number of vowels, there are many more standard inventory constructions than other languages.

i				u
e				o
		a		

Figure 37 – A5a

Figure 37 shows by far the most standard vowel inventory. Over 95% of languages will have at least one of these five vowels.

i				u
ɛ				ɔ
		a		

Figure 38 – A5b

Another of the most common 5V inventories is seen in Figure 38. This inventory has replaced the tense mid vowels with the mid lax vowels.

i				u
e				
				ɔ
		a		

Figure 39 – A5c

Figures 39-44 show some more common, albeit not as common, variations. Figure 39 has combined features of Figures 37 and Figure 38. While changing the back mid tense vowel to the back mix lax vowel, it has retained the front mid tense vowel.

i				u
				o
ɛ				
		a		

*Figure 40 – A5d*

Figure 40, similar to Figure 39, has changed only one of the mid vowels, but in this case, it changed the back vowel instead.

i				u
e		ə		
		a		

*Figure 41 – A5e*

i				u
		ə		o
		a		

*Figure 42 – A5f*

Figures 41 and 42 have replaced either the back mid tense vowel or the front mid tense vowel respectively with the mid central vowel.

	ɪ		ʊ	
ɛ				ɔ
		a		

*Figure 43 – A5g*

Figure 43 has one of the more unique 5V inventories. The only tense vowel in the inventory is the low vowel. All of the others are lax.

		i		
	ɪ		ʊ	
		ə		
		a		

*Figure 44 – P5*

Figure 44, dubbed “the cross”, is not a known inventory, but is unique in that it only has one of the primary vowels, with the other four being lax and more central.

## 4 Vowels

104, or 6.24% of the languages contain only four vowels. The two most common 4V inventories will have /i a u/ and one of the other two primary vowels, as seen in Figures 45 and 46

i				u
e				
		a		

*Figure 45 – A4a*

i				u
				o
		a		

*Figure 46 – A4b*

i				
e				o
		a		

*Figure 47 – A4c*

Another 4V inventory, shown in Figure 47, disregards the high back vowel completely, keeping the two tense mid vowels instead.

i		i		u
		a		

*Figure 48 – A4d*

The fourth most common of the 4V inventories is shown in Figure 48. It is interesting that the languages would add another high vowel, seeing as there are already two and there are also no mid vowels. A much more infrequent version of Figure 48 exists, but with the mid central vowel instead of a high central vowel, as in Figure 49.

i				u
		ə		
		a		

*Figure 49 – A4e*

### 3 Vowels

The triangle is the most stable shape in nature (Math), and this concept passes along to language inventories too. Although there are only fifty-five languages in the dataset, making up a mere 3.3% of the total, 3V inventories are the least likely to change. There are several standard 3V inventories, as shown in Figures 50-54.

i				u
		a		

*Figure 50 – A3a*

This inventory is the most common of the 3Vs. While not containing any mid vowels, like Figure 50, neither the high vowel nor the low vowel spaces are cluttered.

	ɪ		ʊ	
		a		

*Figure 51 – A3b*

Figure 51 is a halfway point between Figures 50 and 52. It contains high lax vowels instead of tense high vowels.

e				o
		a		

*Figure 52 – A3c*

Not common, per se, Figure 52 is found in several languages and lacks high vowels completely. In place of said high vowels, there exist two mid vowels.

i				
				o
		a		

*Figure 53 – A3d*

Lastly, there is the inventory in Figure 53. It is not perfectly balanced, but it contains everything a vowel inventory needs. It has high, mid, and low vowels. It has front, central, and back vowels. It has rounded and unrounded vowels.

		i		
		ə		
		a		

Figure 54 – A3e

Another type of vowel system is that of the vertical vowel system. As seen in Figure 54, there are only central vowels at varying heights. In languages with this type of structure, front and back vowels are treated as allophones of these central vowels.

## 2 Vowels

The general consensus among linguists is that a language must have at least three vowels, which is why finding a language with less so unique. There are only three languages, making up 0.18% of the dataset, that contain less than three vowels. This is in fact such a rare occurrence, that all of the 2V languages consist of a low vowel and a mid central vowel, as in Figure 55.

	ə	
	a	

Figure 55 – A2a

There is, however, a theory that Proto-Indo-European, the mother language of most of the languages of Europe and western Asia, contained the vowel system in Figure 56.

e		o

Figure 56 – A2b

## 1 Vowel

While not found in any known language, there is still a way to have a balanced 1V inventory. The most believable vowel in this case would be /ə/, or possibly /a/.

## Conclusion

There are hundreds of vowel inventories found in the 7000+ recorded languages. Of these 7000+, approximately one quarter, or 23.8%, have been analyzed for this paper. While this does not cover the majority of languages, it covers an amount large enough to be significant. This paper has covered all vowel inventories between one and fourteen vowels, describing their characteristics and other ways they could be structured. While this paper focused only on balanced inventories, balanced does not always mean correct. Many languages do not have a perfectly balanced inventory, and they have been doing just fine.

## Sources

"How Many Languages Are There in the World?" *Ethnologue*. N.p., 28 Feb. 2017. Web. 17 Mar. 2017.

"Inventories." *PHOIBLE Online*. N.p., n.d. Web. 17 Mar. 2017.

"Math and Mystery of the Triangle." *All Articles RSS*. N.p., n.d. Web. 20 Mar. 2017.

*PHOIBLE Online*. N.p., n.d. Web. 17 Mar. 2017.

"Vowel." *Merriam-Webster*. Merriam-Webster, n.d. Web. 17 Mar. 2017.

<https://i.stack.imgur.com/5lh3J.png> (IPA vowel chart)