A NEW ACOUSTIC REFERENCE FRAME FOR VOWELS

H. F. V. Boshoff
University of Pretoria

ABSTRACT

We propose a new reference frame for vowels which is very simply defined in terms of acoustic measurements of the first three formant frequencies. A scaling discretization of formant ratios (SDF) is used to create a regular partition of formant space, depending on only one free parameter: the scaling constant *R*. It is shown that a one-to-one mapping exists between a subset of this partition and either the 'rounded' or the 'unrounded' cardinal vowels as recorded by their originator, Daniel Jones. For some purposes it may therefore be possible to replace the cardinal vowels with the midpoints of elements in the partition, which we call *acoustic reference vowels*. The value of *R* is experimentally determined to be 1.61±0.03. Two considerations taken from a communication engineering perspective are formalized, confirming the empirically determined value of *R*. In measurement terms, the SDF has the properties of a ratio scale.

1. INTRODUCTION

During the 19th Century, vowels were categorized according to variations on one of two schemes: The 'German' triangle introduced by C.F. Hellwag in 1791 and the 'English' square rediscovered by A.M. Bell in 1867 [1]. Towards the end of that century, Passy and Trautmann made efforts to reconcile the two competing viewpoints, influencing Daniel Jones and the IPA [2]. The current (1993) IPA vowel system is still based on this somewhat unsatisfactory compromise, consisting of peripheral 'cardinal' vowels augmented by a number of intermediate qualities.

One of the (many) difficulties with Jones's cardinal vowel system, is that it is essentially an oral tradition, and therefore difficult to transmit and standardize; despite the fact that everyone uses the same diagram [3].

We approach the problem of vowel classification from the general perspective of measurement scales. If it were possible to take the speech signal and measure objectively the vowel quality, obtaining the same result as perceived by an expert phonetician, the only problem left would concern the delimitation of categories.

However, despite many attempts, such a measurement has not been feasible. Leaving aside the question of agreement among phonetic experts, there remain problems of static vs. dynamic vowels, the quality of a vowel relative to the vowel space of the speaker, and the influence on perception of the vowel's phonetic context [3].

These issues are of a psychological nature, and are similar to the ones found e.g. in color perception, tone judgment and perception of pain, warmth and heaviness.

An illuminating exposition of four possible scales of measurement can be found in Stevens [4]: nominal, ordinal, interval and ratio. The simple assignment of a vowel to one of a finite number of classes constitutes a *nominal* measurement. Arranging the vowels in a definite sequence may be considered the result of *ordinal* measurements. Interval measurements demand in addition the creation of equal intervals, and ratio scales presuppose also an absolute zero point.

Stevens [4] wrote that "[m]ost psychological measurement aspires to create interval scales, and it sometimes succeeds... *Ratio scales* are those most commonly encountered in physics." It is worth noting that in his choice of 'cardinal' vowels, Jones [5] placed them at "approximately equal intervals," at least for the 'front' and 'back' vowels separately. The diagrams were drawn accordingly, and placement of a vowel on it may be described as an attempt to perform two different *interval* measurements on the vowel. The 'cardinal' vowel scheme thus aspires to be an interval scale.

Another approach may be taken however, as exemplified by temperature measurement. Although initially based on the perception of heat and cold, the measurement of temperature gradually developed to become independent of the human sense of warmth. Nobody would consider the measurement of temperature 'psychological' today. Yet the human sense still exists, and a figure of 'perceivable temperature' may be quoted, taking into account wind speed, humidity and other factors like time of year along with the temperature. So while 'warmth' had to be simplified to 'temperature' to allow the development of a physical ratio scale (the kelvin scale), the circle is completed when we find temperature invaluable in predicting comfort levels.

I propose an analogous simplification for the realm of vowel quality. It turns out that the definition of an acoustic *ratio* scale for vowels is possible, although the absolute zero of it (similar to that of temperature) is unattainable (at least articulatorily).

We find in fact something even more surprising: the size of our 'degree' will be suggested by the theory to be developed. If this excursion should take us away from the quality judgements of expert phoneticians, we may still entertain the hope of later finding our new scale invaluable when predicting perceived vowel quality.

2. A SCALING DISCRETIZATION OF FORMANT RATIOS

The first simplifying assumption is that a *scaling* discretization of formant space (SDF) is very well suited to be a reference scheme. One advantage of this may be illustrated by mapping the logarithm of frequency to a circle, or, to avoid overlap, to a spiral. Equal formant frequency ratios will then correspond to equal angles. Vocal tracts at different scales only affect the absolute orientation. Might vowel quality judgements be related to angular spacing of formant resonance positions on the cochlea? We need to define a rigorous framework that will allow the investigation of this and other possibilities.

2.1. Definitions

The SDF consists of *ratio regions,* delimited by the expression

$$R^{k_n} \le \frac{F_{n+1}}{F_n} < R^{k_n+1}$$

where F_n is the frequency of formant n in Hz, n=1,2, k_n=0,1,2,… and R is the *scaling constant*; the only free parameter of the SDF. Conventionally, $F_1<F_2<F_3$, so the k_n can't be negative. Figure 1 shows two projections of the SDF, one (P1) contains the regions for n=1, and the other (P2) those for n=2. The ratio regions are labeled A, B, C, etc, corresponding to the numbers k_n=0,1,2…

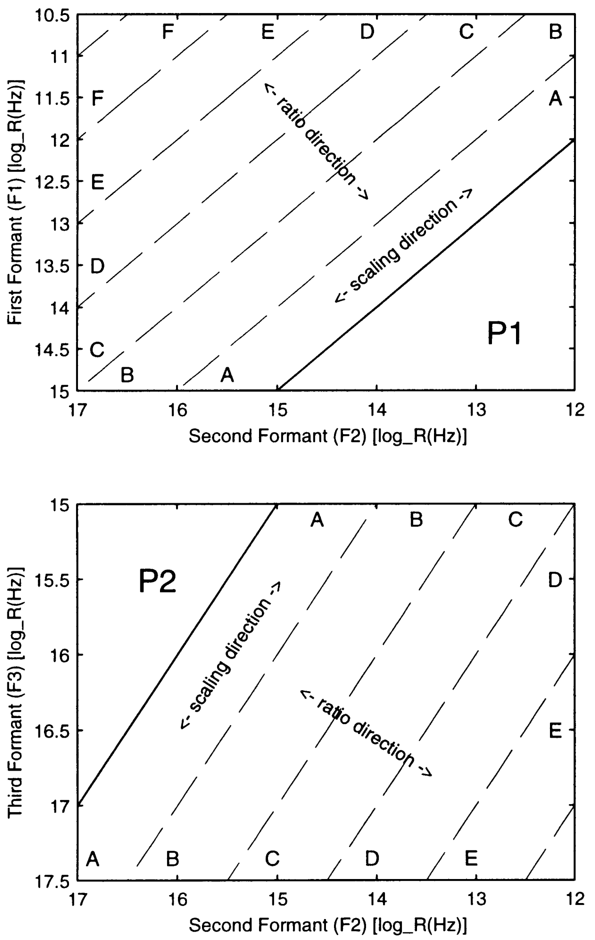

Figure 1. Two projections, P1 (top) and P2 (bottom), of the scaling discretization of formant space (SDF) by powers of R. Diagonal lines connect points of constant frequency ratio. The triangles with the labels P1 & P2 are by definition inaccessible.

The direction along the diagonal lines separating the ratio regions is called the *scaling* direction, because the formants scale at constant ratio in this direction. The one perpendicular to it, in which the ratio changes fastest, is called the *ratio* direction.

Any ratio region in P1 may be combined with one in P2 to form a three dimensional *ratio cell,* denoted by an ordered pair of labels $\{S_1,S_2\}$, where S_i is an element of {A,B,C…}. This notation is usually abbreviated to S_1S_2, e.g. EA or BD.

The term *acoustic reference vowel* (ARV) is used both in a wide and a narrow sense [6]. In the narrow sense, the ARVs are defined as the semi-discrete set of vowels for which the formant ratios conform to the equation

$$\frac{F_{n+1}}{F_n} = R^{k_n+\frac{1}{2}} \qquad \text{for } k_n = 0,1,2\ldots \text{ and } n = 1,2$$

that describes the center lines of the ratio regions. Every pair of integers (k_1,k_2) yields a unique reference vowel, which may be denoted by $S_{k1}S_{k2}$, like the ratio cell of which it forms the center. Table 1 shows 24 cells, where k_1 is limited to 6 and k_2 to 4.

Ratio region in P2

Ratio region in P1				
F	FA	FB	FC	FD
E	EA	EB	EC	ED
D	DA	DB	DC	DD
C	CA	CB	CC	CD
B	BA	BB	BC	BD
A	AA	AB	AC	AD
	A	**B**	**C**	**D**

Table 1. Ratio cells containing the acoustic reference vowels

The definition of the acoustic reference vowels discretizes only two of the three formant ratio dimensions. The exact positions on the two ratio region center lines are not specified. This degree of freedom in the scaling direction is what makes it possible for people at different scales (notably men, women and children) to pronounce the 'same' acoustic reference vowel.

In the wide sense, ARVs may refer to elements of the partition of vowel space which coincides with the ratio cells. All the vowels in the same ratio cell together form one element of the partition. This element is *dominated* by the narrow sense ARV contained in its center.

3. MAPPING THE CARDINAL VOWELS TO THE ACOUSTIC REFERENCE VOWELS

The SDF is an extremely simple acoustic scheme, which is an advantage; but it would have only marginal value for phonetic purposes if the ARVs derived from it were unrelated to any of the traditional vowel systems. We therefore investigated the mapping to the ARV system of Daniel Jones's cardinal vowels [7].

3.1. Data

An audio cassette copy of the original Jones recordings [8] of the cardinal vowels was obtained from the Linguaphone Institute. It was played back on a high quality commercial tape player, and the signal sampled at 16 kHz with 16 bit resolution using the commercial CSL filtering and sampling hardware and software. The full data set consists of 266 vowel utterances; 198 of them represent a single cardinal vowel quality spoken on a fixed pitch. About 100 of these were selected for analysis.

Such vowels of unchanging quality do not occur in ordinary, continuous speech, and they may be regarded as (important) laboratory artifacts, perhaps in the same way as pure chemical elements. This is exactly what is required of a general phonetic reference frame, and they may be compared to the acoustic reference vowels in a straightforward manner.

3.2 Acoustic analysis method

A semi-interactive pitch-synchronous formant analysis was performed on each vowel segment, using functions and interfaces developed in Matlab by the author [6]. This yielded frequency values for the first three formants, which may be viewed as trajectories in three dimensions. The trajectories approach *points* to the extent that the vowel quality was actually kept constant.

Preliminary experiments with the SDF and the dataset of Peterson and Barney [9], indicated that a value of R=1.62 yielded good separation between their categories [6].

Every pitch cycle of every utterance was assigned to a ratio cell of the RDF by calculating formant ratios and using the preliminary value of R. Taking the cell with the majority of cycles for every cardinal vowel as its target, the parameter R was varied to determine the value for which the largest number of cycles overall lay within their respective target cells.

3.3 Results

The optimization of the scaling constant R yielded a maximum at R =1.613. Taking a decrease of 1% in the success criterion to estimate the uncertainty, we found a range of ±0.03. Table 2 gives the mapping of the cardinal vowels to the SDF [7]. On average, 85% of the pitch cycles of every cardinal vowel is associated with a single ratio cell, and in many cases this number is 100% or close to it. Most cells contain both a 'rounded' and an 'unrounded' vowel. With the exception of cardinals 17 and 18, a one-to-one relation is found between either the 'rounded' or the 'unrounded' cardinal vowels, and a subset of the SDF cells.

Mapping of the Jones Cardinal Vowels to the SDF								
Cardinal vowel			SDF ratio cells				Rounding	
No	V	N	% Pitch cycles measured in:	Chosen	%	Adj %	Jones	M
1	i	757	EA (73.7) FA (26.3)	EA	73.7	100.0	U	U
2	e	547	DA (99.8) EA (0.2)	DA	99.8	99.8	U	U
3	ɛ	499	CA (100)	CA	100.0	100.0	U	U
4	a	472	BA (100)	BA	100.0	100.0	U	U
5	ɑ	682	AC (50.6) AB (49.4)	AC	50.6	100.0	U	U
6	ɔ	490	BC (67.3) BB (28.8) AC (3.9)	BC	67.3	67.3	R	U
7	o	539	BC (88.3) BD (11.7)	BC	88.3	88.3	R	R
8	u	730	BD (78.6) BC (11.7) CC (10.3)	BD	78.6	88.9	R	R
9	y	549	EA (95.3) FA (4.7)	EA	95.3	100.0	R	R
10	ø	468	DA (100)	DA	100.0	100.0	R	R
11	œ	452	CA (92.7) BA (7.3)	CA	92.7	92.7	R	R
12	ɶ	377	BA (84.6) AA (8.0)	BA	84.6	93.7	R	R
13	ɒ	358	AC (88.8) BC (11.2)	AC	88.8	88.8	R	R
14	ʌ	394	BB (100)	BB	100.0	100.0	U	U?
15	ɤ	372	BB (69.6) CB (30.4)	BB	69.6	69.6	U	R?
16	ɯ	821	CB (61.6) DB (38.4)	CB	61.6	100.0	U	R?
17	ɨ	727	EA (90.5) DA (9.5)	EA	90.5	90.5	U	U?
18	ʉ	597	DA (85.4) EA (14.6)	DA	85.4	85.4	R	R?
			Average		84.8	92.5		

Table 2. Analysis of the Jones cardinal vowels, with the number of pitch cycles (column N) from his pronunciation of each, for which the formants were measured. (Due to smoothing in time, these measurements are not all independent.) The percentages of pitch cycles measured in each ratio cell, appear in the next column. The cell containing the highest percentage was chosen, shown with its value. If the contents of cells not allocated to any vowel would be included in the counts of adjoining majority cells, a higher value would result, as indicated in the column marked 'Adj %.' Rounding according to Jones and by interpretation of the measurements (M) are shown in the last two columns, with '?' indicating uncertainty. From [7].

Using the given allocations, the first 16 cardinal vowels may be placed as indicated in Table 3. The vowel [ɑ] is the only one splitting nearly equally between two cells, and it is provisionally put in both. The cells occupied by the cardinal vowels form a *compact* set in both the rounded and the unrounded cases.

All and only the 'front' vowels map to cells labeled with XA. Most cardinal 'back' vowels map to XB and XC, with only [u] going to BD. It seems that the traditional phonetic labels 'front,' 'central' and 'back' correspond to values of k_2. This would imply that the 'front-back' distinction (at least in the static cardinal sense) is best made using the formant ratio F_3/F_2, and not in the F1-F2 plane as is usually done.

The placement of symbols in Table 3 is unsatisfactory for a variety of reasons [7]. The proposed use of the IPA vowel symbols for the acoustic reference vowels is shown in Table 4.

Ratio region in P2

<table>
<tr><td rowspan="6">Ratio region in P1</td><td>E</td><td>i|y</td><td>-|-</td><td>-|-</td><td>-|-</td></tr>
<tr><td>D</td><td>e|ø</td><td>-|-</td><td>-|-</td><td>-|-</td></tr>
<tr><td>C</td><td>ɛ|œ</td><td>-|ɯ</td><td>-|-</td><td>-|-</td></tr>
<tr><td>B</td><td>a|ɶ</td><td>ʌ|ɤ</td><td>ɔ|o</td><td>-|u</td></tr>
<tr><td>A</td><td>-|-</td><td>ɑ|-</td><td>ɑ|ɒ</td><td>-|-</td></tr>
<tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td></tr>
</table>

Table 3. The first 16 cardinal vowels placed in the periodic table, based on the acoustic analysis. Vowels measured as ‘unrounded’ are on the left, ‘rounded’ ones on the right.

Ratio region in P2

<table>
<tr><td rowspan="6">Ratio region in P1</td><td>E</td><td>i|y</td><td>-|-</td><td>-|-</td></tr>
<tr><td>D</td><td>e|ø</td><td>ɨ|ʉ</td><td>-|-</td></tr>
<tr><td>C</td><td>ɛ|œ</td><td>ɘ|ɵ</td><td>ɯ|u</td></tr>
<tr><td>B</td><td>æ|ɶ</td><td>ɜ|ɞ</td><td>ɤ|o</td></tr>
<tr><td>A</td><td>-|-</td><td>a|ɒ</td><td>ɑ|ɔ</td></tr>
<tr><td></td><td>A</td><td>B</td><td>C</td></tr>
</table>

Table 4. Proposed use of the IPA vowel symbols for the acoustic reference vowels

One last modification to the representation is to rotate the table through 45º . The resultant scheme is shown in Figure 2. The main advantage of this change is that [i] and [u] appear on the same level, as traditionally assumed when using the phonetic feature ‘height’ [10]. Unused cells are left undrawn.

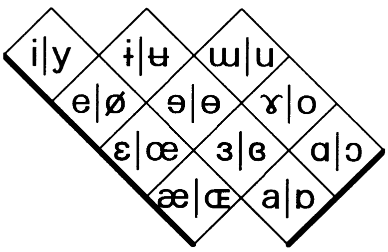

Figure 2. A rotation of the table of acoustic reference vowels

4. DERIVATION OF THE SCALING CONSTANT *R*

Finding a single parameter acoustic system accounting for all the cardinal vowel qualities, forces the question: What is so special about the value that allows this? An explanation may be sought in terms of desirable properties of vowels as discrete elements in an auditory communication code.

The first assumption already mentioned, is that of a *scaling* scheme. The code is shared by all speakers and listeners irrespective of size of their vocal tracts and cochleae.

Another defensible assumption is that of *robustness*, especially against the effect of non-linearities in the generation, transmission and reception media. Non-linear filtering of the vowel signal, which contains a number of discrete frequencies, may lead to the creation of sum and difference frequencies via an effect known as ‘mixing’ in modulation theory.

A minimally scaling combination of three formants is a *borderline* vowel, formed with equal upper and lower intervals:

$$F_3/F_2 = F_2/F_1 = R.$$

If this signal is to be robust to first order against non-linearities, we may require that $F_3 - F_2 = F_1$. Combining these equations, yields as one of two solutions

$$R = (1+5^{1/2})/2 = G.$$

G is the ‘golden ratio,’ which results inter alia as the limit of the ratios of consecutive terms in the Fibonacci sequence. Generalizing this approach leads to other borderline vowels, which together reconstitute the SDF [7].

It is thus possible to derive a value for the scaling constant R from two quite general principles, which is compatible with the value found by analyzing Jones's cardinal vowels.

5. CONCLUSION

The scaling discretization of vowel space constitutes a ratio scale for ‘objectified’ vowel quality. Intervals are measured in log-frequency differences between formant positions, and the absolute zero is the point where $F_1 = F_2 = F_3$. In a physical system, eigenvalues never coincide exactly, implying that this point is not attainable as resonant frequencies of any acoustic tube. Two dimensions of vowel quality also have zero points separately, namely $F_1 = F_2$ and $F_2 = F_3$.

The value of R obtained empirically and theoretically forms an obvious step size to use when measuring vowel quality. It is remarkable how well it accounts for the impressionistic vowel scales of the phonetic tradition.

ACKNOWLEDGMENTS

The work reported here was completed while enjoying a post doctoral fellowship with Prof. E.C. Botha at the University of Pretoria. It was started as part of my PhD research at the University of Stellenbosch, under the guidance of Dr. M.W. Coetzer and Dr. J.A. du Preez. The contributions of the two Institutions and the three scholars are gratefully acknowledged.

REFERENCES

[1] Michaelis, G. 1881. Über die Anordnung der Vokale. *Herrigs Archiv*, Braunschweig, George Westermann.
[2] Viëtor, W. 1923. *Elemente der Phonetik*. 7[te] Auflage. Leipzig, O. R. Reisland.
[3] Ladefoged, P. 1967. *Three areas of Experimental Phonetics*. London, Oxford University Press.
[4] Stevens, S. S. 1946. On the theory of scales of measurement. *Science*, 103, 677- 680.
[5] Jones, D. 1964. *An Outline of English Phonetics*, 9[th] Ed. Cambridge, Heffer.
[6] Boshoff, H.F.V. 1997. *The acoustic discretization of vowel space*, Ph D Thesis, University of Stellenbosch. In Afrikaans.
[7] Boshoff, H.F.V. and Botha, E.C. 1998. An acoustical analysis of the cardinal vowels as spoken by Daniel Jones. *Proceedings of the South African Symposium on Communications and Signal Processing*, 73-78. Piscataway, IEEE.
[8] Jones, D. 1956. *Cardinal vowels spoken by Daniel Jones* (Text of records with explanatory notes). London, Linguaphone Institute.
[9] Peterson, G.E. and Barney, H.L. 1952. Control methods used in a study of vowels, J. *of the Acoustical Society of America*, 24, 175-184.
[10] Boshoff, H.F.V. and Botha, E.C. 1998. On the structure of vowel space: A genealogy of general phonetic concepts. *Proceedings of the International Conference on Spoken Language Processing*. Sydney.