On the acoustic and perceptual characterization of reference vowels in a cross-language perspective

Jacqueline Vaissière

ON THE ACOUSTIC AND PERCEPTUAL CHARACTERIZATION OF REFERENCE VOWELS IN A CROSS-LANGUAGE PERSPECTIVE

Jacqueline Vaissière

Laboratoire de Phonétique et de Phonologie, UMR/CNRS 7018 Paris, France
Jacqueline.vaissiere@univ-paris3.fr

ABSTRACT

Due to the difficulty of a clear specification in the articulatory or the acoustic space, the same IPA symbol is often used to transcribe phonetically different vowels across different languages. On the basis of the acoustic theory of speech production, this paper aims to propose a set of focal vowels characterized by an almost complete merging of two adjacent formants: F1 and F2, F2 and F3, and F3 and F4 (sometimes F4 and F5 for some speakers). These reference vowels constitute a subset of Jones's Cardinal Vowels (CVs); they are the only vowels that can be called "quantal" in Stevens' sense. Formant merging creates a vowel-specific sharp concentration of spectral energy in a narrow region of the frequency scale. This acoustic result results from very specific articulatory configurations and entails special perceptual characteristics.

Keywords: IPA, vowels: focal, quantal, cardinal

1. INTRODUCTION

This paper defines a set of reference vowels to serve as a basis for case studies of vowel systems as well as for cross-language comparisons. These vowels are defined on the basis of the acoustic theory of speech production [11]. This proposal draws on various models of the vowel space: Quantal Theory (QT) [34] Dispersion theory (DT) [25]and Dispersion-Focalization Theory (DFT) [32]Extensive use is made of Maeda's articulatory program [27], and of findings concerning spectral integration and Center of Gravity effects [8] The reference vowels thus defined turn out to constitute a subset of Daniel Jones's cardinal vowels (CVs) [19].

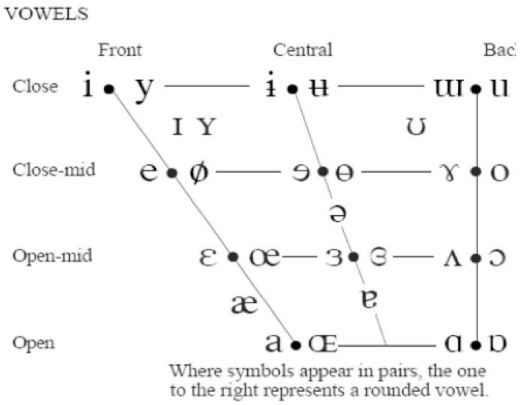
First, I review the principles of Jones's CVs and the IPA vowel chart. Then I set out the model that links articulation, acoustics and perception. Finally, I provide a description of the subset of CVs that are *focal* ("quantal") in acoustic terms.

2. IPA CHART AND THE CARDINAL VOWELS

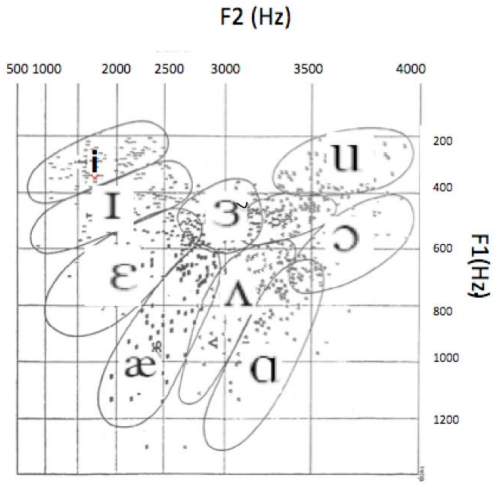
2.1. The IPA vowel chart, and other proposals

The first International Phonetic Alphabet (IPA) was proposed in 1886 by a group of European language teachers led by Paul Passy. Since then, the IPA has been revised several times. Its aim is to provide a universal standard for transcribing all speech sounds [18]. It has been widely used for over a century by linguists, language teachers, and speech therapists.

Figure 1: (a): IPA vowel chart. (b): Peterson and Barney's formant plot for the English vowels (male, female and children speakers) [29].



(a)



(b)

The vowels in the IPA (Fig. 1a) are described essentially using three articulatory dimensions: (i) frontness-backness of tongue position (horizontal axis); (ii) height of the tongue (vertical axis); and (iii) rounding/spreading of the lips, encoded through the use of distinct symbols (e.g. [i] *vs.* [y]). Other dimensions are added, when necessary, by means of diacritics, such as velum state (nasalization), phonation type (breathiness, creakiness), tongue root advancement/retraction, and secondary narrowing along the VT (palatalization, velarization, pharyngealization). All the parameters are articulatory.

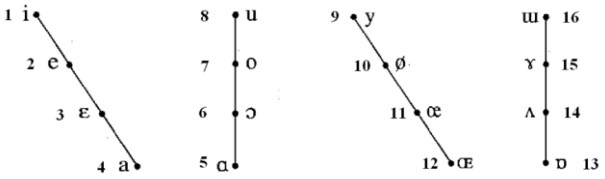
The IPA was originally designed for transcribing phonemic oppositions. The articulatory characterization that it provides is not precise enough to pinpoint a specific vowel quality. The articulatory description of vowels is much more complex that that of consonants: the constriction is less strong, and several articulatory configurations are often available to produce the same percept, as can be easily demonstrated through the use of articulatory models [27]. In this light, it does not actually come as a surprise that the same symbol occasionally receives contradictory characterizations. For example, [a] is considered as a *front* open vowel (IPA, Bloch and Trager's system [5]); it is an issue where the boundary falls between this front open vowel and [æ]. American usage does not clearly distinguish [a] from [ɑ], and uses [a] for a low *back* unrounded vowel [30]: see e.g. Chomsky and Halle's system [9], discussed in [2].

Such inadequacies encourage a loose use of IPA vowel symbols in language descriptions: the choice of symbols is guided by structural arguments (e.g. vowel alternations), rather than by considerations of phonetic accuracy. Clearly, we need more precise tools.

2.2. Jones's proposal

Jones's cardinal chart aims to characterize the *phonetic* quality of the vowels. In his *Outline of English Phonetics* [19], Jones claims that "a good ear can distinguish well over fifty vowels, exclusive of nasalized vowels, vowels pronounced with retroflex modification, etc." The Cardinal Vowels (CVs) (eight primary CVs and eight secondary CVs, see Fig. 2) aim to provide reference points to specify the quality of the vowels in a cross-language perspective: any vowel quality, from any language, can be described by interpolating between the reference points. The CVs are widely employed to this day.

Figure 2: Jones's CVs. Left: primary; right: secondary.



Jones gave an *articulatory* definition for the three first primary CVs, [i], [ɑ] and [u]. [i] is the highest and most fronted vowel that a human vocal tract (VT) can produce, with spread lips. [u] is realized with the tongue as "back" and as high as possible in the mouth, with pursed lips. [ɑ] is uttered with the tongue as "low" and "back" as possible in the mouth. The other five primary CVs, [e ɛ a o ɔ], are defined by Jones as '*auditorily* equidistant' between these three 'corner vowels': [e], [ɛ] and [a] are auditorily at an equal distance from each other between [i] and [ɑ]; likewise for [o] and [ɔ], between [u] and [ɑ]. The auditory distance was judged to be directly related to tongue height. Choosing the opposite lip configuration yields the 8 secondary CVs, Ladefoged pointed out the need for a new basis for defining CVs: the description in terms of *highest point of the tongue* does not reflect actual tongue position [21]. Moreover, according to Jones, the CVs can only be learnt through oral instruction from a teacher who knows them.

- *Acoustic characteristics of the CVs*

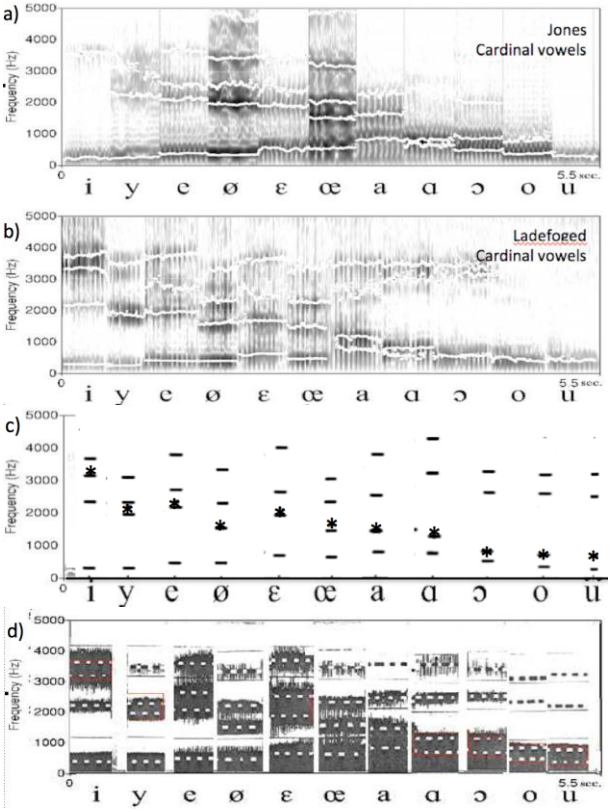
The CVs have not been explicitly related to their acoustic characteristics (see however [26]).

Figure 3 illustrates the rendition of eleven CVs uttered by Daniel Jones (DJ) himself, and by Peter Ladefoged (PL), along with the French oral vowels by a male speaker (FR). The renditions of the CVs by DJ and CV are available on the Web [33]. Note that in Parisian French, all vowels are monophthongs; each oral vowel can be fully specified by a single spectrum. The French vowels are referred to by Jones as good examples of CVs.

Let us observe the 33 vowels in Fig. 2. The 5 primary CVs, [i u ɔ o ɑ], and the secondary CV [y], exhibit a clear concentration of energy due to the merging of two formants: F3 and F4 for [i], creating a concentration of energy above 3200 Hz (hence my notation: $F3F4^{3200Hz}$), F2 and F3 for [y] ($F2F3^{1900Hz}$) and F1 and F2 for [ɑ] ($F1F2^{1000Hz}$), [ɔ]

(F1F2^{800Hz}), [o] (F1F2^{600Hz}), [u] (F1F2^{400Hz}). The renditions of these vowels are very similar to the French vowels. The notable discrepancies are the following: DJ's [y] does not sound as [y] (similar remark for DJ's [ø]); observe that there is a lack of F2 and F3 merging in DJ's [y]. The grouping of F2 and F3 around 1900 Hz (for a male speaker) is a defining acoustic characteristic for a vowel to be perceived as close to cardinal [y], as will be discussed below. PL's [a] has a concentration of energy around 1000 Hz due the grouping of F1 and F2, while for DJ and FR, the F2 of [a] is at a mid distance between F1 and F3.

Figure 3: From top to bottom: (a) eleven CVs as spoken by Jones and (b) Ladefoged; (c) the values of the formants used in [4]: F2' (marked by a cross) is indicated; (d) the French oral vowels as spoken by a male speaker. Note that when two formants are very close, a single peak is detected.



3. MODELING THE LINK BETWEEN ARTICULATION AND ACOUSTICS

3.1. F1, F2 and F2'

- *Articulatory chart and formant frequencies*

There is a well-known correspondence between the articulatory vowel space as described by the IPA chart and the acoustic vowel space where F1 is plotted against F2 (or the distance between F1 and F2). A typical formant plot is represented under the IPA chart in Figure 1. The vertical F1 corresponds to vowel "height": the "lower" the vowel, the higher the F1. The horizontal F2 axis corresponds to tongue advancement: the more "back" the vowel, the lower the F2 frequency value [23].

- *F1 and F2 and the specification of the vowels' phonetic value*

The F1/F2 plot offers a fairly good visual separation of the vowels. But the three articulatory dimensions of the IPA chart (or of Stevens and Fant's models, shown in Fig. 4) are reduced to two dimensions, raising the issue of whether two dimensions, such as the two first formant frequency values, can provide an adequate acoustic representation of vowels. The answer depends on the location of the concentration of energy on the frequency scale [10]. When the energy is concentrated in the low frequencies (say, under 1000 Hz), the first two formants, F1 and F2, are sufficient for creating the quality of the back CVs. By the law of acoustics, the upper formants of these back vowels are of weak intensity, and therefore carry little perceptual weight (if any). The first two grouped formants are even perceptually equivalent to a single formant, so that back vowels can be synthesized using a single formant [10]. When the energy is not concentrated in the low frequency range, however, several formants above F1 are of comparable strength, and have a perceptual weight. In languages that contrast front *vs.* mid, round *vs.* unrounded vowels, F3 plays a critical role. In French, for example, F1 and F2 for /i/ and /y/ can be similar for some speakers (see Figure 7 for an example). F1 and F2 have proved insufficient for imitating the phonetic values of the non-back vowels. In short, F1 and F2 alone are not adequate to represent the acoustic characteristics of the whole set of CVs.

- *F2': getting at a perceptually relevant aggregate value for the formants above F1*

F2' (F2 prime) is an aggregate computed from F2 and higher formants. The F2' frequency substitutes a single peak to all formants above F1, aiming to mirror their perceptual integration [4], [6]. F2' can be determined by experiments in which a subject is asked to adjust a second formant in a 2-formant vowel to match an original multi-formant stimulus: F1 is fixed, and is equal to the original F1 frequency of the vowel; F2' is variable at will. F2'

is called a *perceptually relevant formant value*. Figure 3c illustrates the formant values proposed to the listeners for the CVs and the resulting perceived F2' [4]. There are quite a few different formulas to estimate F2' and their predictions differ. The vowels can be divided into three groups nonetheless, depending on the relationship between F2 and F2'. Generally, when F2 is above 2000 Hz (as in [i e]), F2' is higher than F2; it is close to F4 (or even higher) for [i] in languages like Swedish and French where the vowel is characterized by the grouping of F3 and F4 (like the cardinal [i]: see Fig. 3a, b, c and d). It lies in-between F2 and F3 for [y], for which F2 and F3 are grouped. When F2 is below 1000 Hz, F1 and F2 are bunched together, and F2' is close to F2 (sometimes close to F1 for [u]).

F2' therefore serves a dimension-reduction function, from four formants to just two. Vowel mapping based on F1 *vs.* F2' is more successful than F1 *vs.* F2 in separating the front high and mid rounded and unrounded vowels [12]. Synthesis based on F1 and F2' is not very natural for front vowels, however [15]. F2' corresponds to the best approximation of the upper formants by a single value, but it is not really perceptually equivalent to the original. To conclude, F1 and F2' do not provide a complete acoustic representation of the vowels. On the other hand, the first four formants reproduce the quality of the vowels with very high accuracy.

3.2. Studying the relationships through modeling

The relationships between the articulatory space (the VT profiles), the acoustic space (the formant frequencies), and the perceptual space are complex and not linear [35]. Modeling allows for a detailed investigation into these relationships. Specifically, it yields insights into the gestures that result in the clustering of two or more formants.

Modeling is based on the source-filter theory, i.e. the principle of the independence between the (voice) source at the glottis (phonation) and the filtering by supraglottal cavities (articulation) [7], [11], [35]. The relationship between vowel articulation and vowel spectra mainly lies in the fact that a constriction near a pressure node lowers the formant frequency, whereas a constriction near a pressure antinode raises it.

Any mid-sagittal profile (obtained from X-ray or MRI data) can be converted into a cross-sectional area function where the VT is represented by a series of cylindrical sections of averaging area along a straight axis from the glottis to the lips. The area function preserves the resonance characteristics of the VT [11]; the area function is transformed into an acoustic spectral transfer function and the resulting sound is generated and can be heard.

If desired, the area function is able to reproduce the details seen on mid-sagittal profiles; the sagittal profiles may be simplified by the concatenation of simple tubes, for example two connected tubes for [i], [y], [ɑ] or [a], and four connected tubes for [o], [ɔ] and [u] [12], [35]. To estimate the sensitivity of each formant to a small or large articulatory change [14], each section of the area function can be slightly perturbed (constricted or expanded), and the transfer function calculated. The acoustic characteristics of the resulting signal can be then compared with those of the original sound (if available). The synthesized signal can be used as stimulus for perception tests (for details see [27]).

3.3. Fant's nomograms

A nomogram is a very useful way to display the acoustic consequences of modifying constriction position, constriction size, and degree of lip opening, as illustrated in Fig. 4. Fant [11] has shown that the vocal tract transfer function estimated from X-ray data corresponding to vowels can be quite accurately calculated from a four-tube, three-parameter model [36].

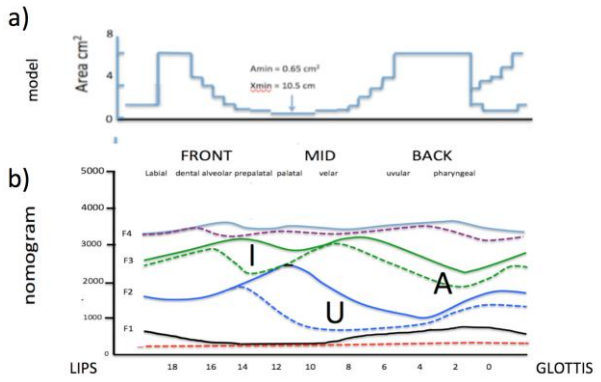
Note that the three parameters used to specify the vowels are not the same as those three parameters used in the IPA chart (Fig. 1). The first parameter is the distance from the glottis to the center of the constriction, the second is its area, and the third is the length-to-opening ratio of the lip tube area. [i], [u] and [ɑ] correspond respectively to a constriction on the front (palatal), mid (velar) and back (pharyngeal) parts of the VT.

In Figure 4, the constriction size is fixed and the two varying parameters are (i) the location of the constriction, from glottis (on the right) to lips (on the left) and (ii) the lip configuration with two states, constricted and opened. Human speakers can only produce vowels over a range that is less than half that the range represented in Fant's nomograms [24]. Nevertheless, nomograms represent well the essential resonance characteristics of the VT.

There are basically three regions. When the

constriction is near the front end of the VT, the distance between F1 and F2 is much larger than the distance between F2 and F3. This region corresponds to i- and e-like sounds (zone "I" in Fig. 4). When it is close to the glottis, F1 is high and the region corresponds to open sounds (zone "A"). When the lips are rounded, there is a region where the distance between F1 and F2 is much smaller than the distance between F2 and F3, and F1 and F2 are low in frequency (zone "U"). The regions where F2 is high and therefore close to F3 (zone I) or F1 and F2 converge (zone A and zone U) correspond to quantal regions, as described by Stevens [34].

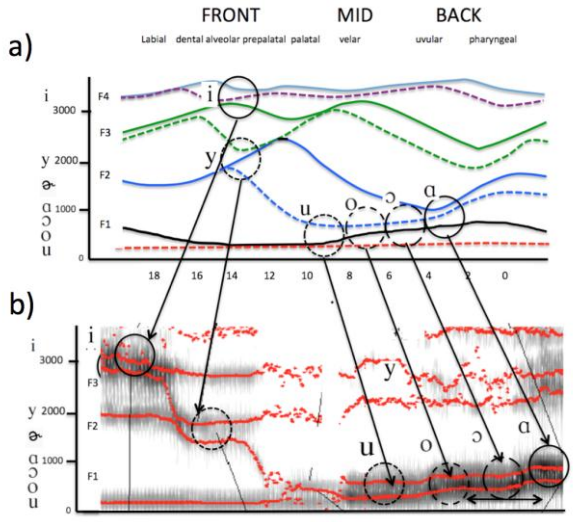
Figure 4: (a): Area function corresponding to Fant's second model. The vertical arrow represents the location of the maximum tongue constriction in the VT that is varied from glottis to lips. The minimal cross-section area at the constriction is fixed here to 0.65 cm². (b): the corresponding nomogram. Straight and dotted lines correspond respectively to open (in the solid lines) and rounded/protruded lips (in the dashed lines). Black, blue and green colors refer respectively to F1, F2 and F3.



4. ACOUSTIC DEFINITIONS FOR REFERENCE VOWELS

Now I propose an acoustic definition of the reference vowels which are a subset of CVs manifesting formant clustering. Figure 5 represents the same nomogram as in Figure 4, but the points where two formants converge are singled out by circles. As clearly stated in Stevens's Quantal Theory (QT), when the frequency of a formant is maximally high or low, it usually goes hand in hand with formant convergence. By the law of VT acoustics, when two formants converge, their amplitude increases by 6 dB per halving their distance [11], creating a sharp spectral salience in a well-defined frequency range. Formant merging may be favored because it corresponds to articulatory stability, as stated by the QT, or for auditory reasons.

Figure 5: Top: same nomogram as in Fig. 3. The points of formant clustering are circled, and the corresponding CVs are indicated. Bottom: six vowels uttered by a native of French.



CV No. 1: C1[i] = prepalatal (⇑F3F4)$_{3200Hz}$

When the constriction is very fronted, i.e. in the prepalatal region, F3 reaches a maximum (transcribed as ⇑). F3 and F4 converge at about 3,200 Hz (for a male speaker). The lips are spread (the solid lines correspond to a spread configuration of the lips in Fig. 6). F3 is affiliated to the front cavity (indicated by underlining in our notation), which is made as short as possible. Articulatory modeling shows that the tongue has to be placed parallel to the palate to create a half-wave-length resonance, the type of resonance which creates the highest frequency. F2 is not maximal. The vowel fits well to the CVs uttered by PL and DJ, and to the /i/ of French [37] and Swedish [12].

Gendrot, et al. [17] compared the four first formant frequencies of /i/ in continuous speech in English, German, French, Spanish, Portuguese, Arabic and Mandarin. Their results indicate that French /i/ has the lowest F1, the highest F3 and the highest F4, as well as the smallest distance between F3 and F4 (see Table 1).

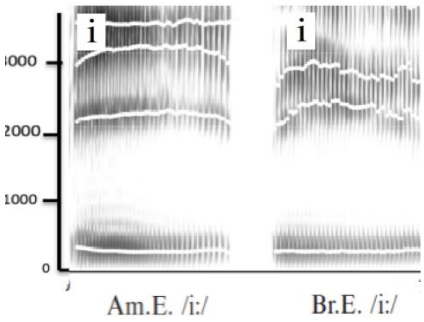
This reference vowel does not seem very common, maybe because it requires a high articulatory precision. The next figure illustrates two types of /i/, as pronounced by Ladefoged (the sounds are available on the Internet). The /i/

represented on the left sounds much "sharper" that the second one.

Table 1: Mean F1, F2, F3 and F4 frequencies values and the distance between F3 and F4 [17].

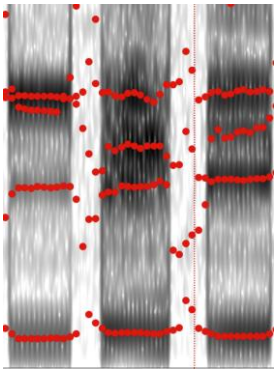
	F1	F2	F3	F4	F4 – F3
German	319 (70)	1991 (222)	2610 (239)	3621 (248)	1012 (269)
English	352 (61)	2044 (186)	2503 (199)	3442 (225)	939 (244)
Arabic	398 (130)	2102 (169)	2678 (141)	3364 (295)	686 (258)
Spanish	375 (57)	2126 (155)	2784 (149)	3634 (126)	851 (226)
French	**302 (87)**	**2024 (158)**	**2848 (228)**	**3494 (258)**	**646 (230)**
Italian	347 (61)	2065 (231)	2693 (236)	3589(400)	895 (301)
Mandarin	360 (109)	2132 (358)	2836 (290)	3644 (265)	809 (304)
Portuguese	344 (67)	1906 (185)	2503 (277)	3576 (277)	1075 (329)

Figure 6: The two types of [i], as spoken by Ladefoged [33].



Note that focal vowels seem to be as sensitive to coarticulation as non-focal vowels [31] [36]. Figure 7 illustrates the spectrograms corresponding to the central portion of the vowel [i], in isolation, and in uvular context. When the vowel is surrounded by [ʁ], the length of the front cavity increases, and the front cavity resonance (here: F3) tends to decrease in frequency. F1 tends to increase and /i/ sounds close to [e].

Figure 7: Spectrogram corresponding to the central part of [i], spoken in isolation (left), in the uvular context [ʁiʁ] (mid), and to the vowel /e/ (right). Spoken by the author.



	/i/	/i/ (rir)	/e/
F1	401	468	418
F2	2461	2452	2560
F3	3500	3014	3213
F4	3700	3766	3783

Cardinal C9 [y]= (F2F3) ^{1900Hz}

(F2F3)^{1900Hz} corresponds to the narrowest passage in the prepalatal region (the second highest circle in Fig. 4), where F3 is most sensitive to rounding, and the lips are rounded. In the transition from [i] to [y], F2 becomes a resonance of the front cavity. Languages contrasting [i] and [y] seem to prefer a prepalatal position for both [41]. (F2F3)^{1900Hz} does not correspond to Jones's /y/, nor to Swedish or German, but clearly corresponds to the rendition of cardinal vowel /y/ by PL and to French /y/.

Cardinal C8 [u]: (↓F1⇓F2)^{400Hz}

F1F2 clustering corresponds to the lowest possible concentration of energy. F1 and F2 correspond to two Helmholtz resonances, the type of resonances that produces the lowest resonance frequency. It requires two strong constrictions, at the lip and at the middle of the mouth. It represents the lowest concentration of energy that a human VT can produce. The vowel corresponds to DJ's and PL's CV [u].

Cardinal C5 [ɑ]: ⇑ (F1F2)^{1000Hz}

It corresponds to the highest possible clustering of the two first formants. A constriction at the root of the tongue leads to an even higher F1 and an /æ/-like sound [13], with a separation of F1 and F2 (see Fig. 1).

Creating Cardinal C6 [ɔ] and Cardinal C7 [o]

As stated by DJ, the two other back vowels may be created as equidistant from Cardinal C8 [u] and Cardinal C5 [ɑ]. For keeping F1F2 clustered, the tongue constriction has to move back from C8 to C5 synchronously with a delabialization gesture.

Mid vowel [ɚ]= (F2⇓F3)^{1500Hz}

[ɚ]= (F2⇓F3)^{1500Hz} is not among the CVs. Nonetheless, it represents an extreme in terms of a low F3, which gets as low as 1,500 Hz. The production of (F2⇓F3)^{1500Hz} is achieved by a constriction in the pharyngeal region, plus lip rounding and a bunching of the tongue. Palatal retroflexion is one gesture that lowers F3 (alveolar retroflexion lowers F4) [11]. The three necessary constrictions correspond to the three points along

the vocal tract where the volume velocity nodes of F3 are located [7].

The three other front primary vowels C2, C3 and C4 (see Fig. 2) do not correspond to a less constricted VT [13]. These vowels are more difficult to define in acoustic terms. They have in common two peaks of equal strength above F1 and no focalization.

5. CONCLUDING REMARKS

When the IPA was created, an acoustic analysis of the vowels could not be performed: acoustic phonetics really began with the invention of the sound spectrograph in the 1940s, and it developed from the early 1950s onwards. Technical progress in articulatory synthesis, real-time spectrographic displays, and progress in the acoustic theory of speech production now make it possible to study the characteristics of vowels.

The target for a vowel seems to be much easier to describe in acoustic rather that in articulatory terms. A phonemically defined contrast involves even more than two gestures. For example, Wood's data [41] showed that contrasting [i] and [y] involves a whole package of maneuvers (rounding of the lips, tongue retraction and larynx lowering); all the gestures enhance the contrast between the two vowels in acoustic terms.

A traditional phonological feature (such as *round* or *back*) is generally described by a defining gesture: lip rounding for the feature *round*, or tongue retraction for the feature *back*. Both lip rounding and tongue retraction lead to the lengthening of the front cavity, thus to the lowering of the formants associated with that cavity. The two gestures have to work in strong synchrony, for the manipulation of F3 for the front vowels (spreading and fronting), or to keep F1 and F2 clustered for the back vowels (rounding and backing): the more back the vowels, the less rounded the lips (4 different phonetic degrees of rounding for the back vowels to keep F1F2 clustered).

The finding that a small number of vowels are acoustically focal opens numerous perspectives for future research, such as: Are they any easier to recognize than other vowels? Are their coarticulatory properties any different? What is their distribution among the world's languages? Schwarz and coworkers [32] found that focalization led to more stable patterns in discrimination tasks, but more work has to be done in this direction. As for the distribution across languages, focal ("quantal") vowels do not appear to be particularly common. According to Ladefoged [22], the Ngwe language of West Africa has 8 vowels which are rather similar to the 8 primary cardinal vowels. French is often cited as having vowels close to the CVs (note, however, that younger generations have lost the opposition between [ɑ] and [a]; the opposition between [œ] and [ø] is currently weakening).

As a final perspective, the use of an articulatory model with more parameters (and more constraints) allows for a more realistic study than the former three-parameter model. In Maeda's models, the parameters are statistically derived from real X-ray data [37], [38], [39].

6. ACKNOWLEDGEMENTS

We would like to thank Shinji Maeda and Alexis Michaud for their helpful comments.

7. REFERENCES

[1] Badin, P., Perrier, P., Boë, L.J., Abry, C. 1990. Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *J. Acoust. Soc. Am.* 87, 1290-1300.

[2] Barry, W., Trouvain, J. 2008. Do we need a symbol for a central open vowel? *J. Int. Phonetic Association* 38(3), 349-357.

[3] Beddor, P., Hawkins, S., 1990. The influence of spectral prominence on perceived vowel quality. *J. Acoust. Soc. Am.* 6, 2684-2704.

[4] Bladon, R, Fant, G. 1978. A two-formant model and the cardinal vowels, speech transmission laboratory. *Quarterly Progress Status Report,* Royal Institute of Technology, Stockholm, 1, 1-8.

[5] Bloch, B., Trager, G. 1942. *Outline of Linguistic Analysis*. Baltimore: Linguistic Society of America.

[6] Carlson, R., Granström, B., Fant, G. 1970. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress Status Report (STL-QPSR)* 2/3, 19-35.

[7] Chiba T, Kajiyama M. *The Vowel: Its Nature and Structure.* Tokyo: Tokyo-Kaiseikan.

[8] Chistovich, L., Sheikin, R., Lublinskaya, V., 1979. "Centers of gravity" and the spectral peaks as the determinants of vowel quality. In Lindblom, B., Ohman, S. (eds.), *Frontiers of Speech Communication Research.* London: Academic Press.

[9] Chomsky, N., Morris, H. 1968. *The Sound Pattern of English.* New York: Harper and Row.

[10] Delattre, P., Liberman, A., Cooper, F., Gerstman, L. 1952. An experimental study of the acoustic determinants of vowel color. *Word* 8, 195-210.

[11] Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

[12] Fant, G. 1973. *Speech Sounds and Features.* Cambridge. MA and London, UK: MIT Press.

[13] Fant, G., Båvegård, M. 1997. Parametric model of VT area functions: Vowels and consonants. *Speech, Music and Hearing – Quarterly Progress Status Report: Stockholm* 38(1), 1-20.

[14] Fant, G., Pauli, S. 1974. Spatial characteristics of vocal tract resonance modes. *Proc. Speech Comm. Sem.*, Stockholm, Sweden, 74, 121-132.

[15] Fant, G., Risberg, A. 1963. Auditory matching of vowels with two formant synthetic sound, *STL- Quarterly Progress Status Report: Stockholm* Stockholm, 4, 7-11.

[16] For comparing British and American English vowels as spoken by Ladefoged: *http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/vowels/chapter3/american.aiff* and BBC English vowels: *http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/vowels/chapter3/bbcenglish.html*

[17] Gendrot, C., Adda-Decker, M., Vaissière, J. 2008. Les voyelles /i/ et /y/ du français: aspects quantiques et variations formantiques. *Proc. Journées d'Etude de la Parole* Avignon.

[18] International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press.

[19] Jones, D. 1956. *An Outline of English Phonetics.* Cambridge, UK: Cambridge University Press.

[20] Jones, D. 1956. *Cardinal Vowels* (8th ed.). Cambridge: W. Heffer & Sons.

[21] Ladefoged, P. 1962. *Elements of Acoustic Phonetics*. Chicago: University of Chicago Press.

[22] Ladefoged, P. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

[23] Ladefoged, P. 1993. *A Course in Phonetics*. (3rd edition). Fortworth, TX: Harcourt Brace Jovanovich College Publishers.

[24] Ladefoged, P., Bladon, A. 1982. Attempts by human speakers to reproduce Fant's nomograms. *Speech Communication* 9, 231-298.

[25] Liljencrants, J., Lindblom, B., 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast, *Language* 48, 839-862.

[26] Lindblom, B., Sundberg, J., 1969. A quantitative theory of cardinal vowels and the teaching of pronunciation. *Quarterly Progress Status Report*, Royal Institute of Technology, Stockholm 2-3, 19-25.

[27] Maeda, S. 1990. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer Academic, 131-149.

[28] Michaud, A., Vaissière, J. 2009. Perceptual transcription and acoustic data: the example of /i/ in Yongning Na (Tibeto-Burman). *Chinese J. Phon.* 2, 10-17.

[29] Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.

[30] Pullum, G., Ladusaw, W. 1986. *Phonetic Symbol Guide.* Chicago and London: The University of Chicago Press.

[31] Recasens, D. 1984. V-to-V coarticulation in Catalan VCV sequences. *J. Acoust. Soc. Am.* 87, 1624-1635.

[32] Schwartz, J.-L., Boë, L.-J., Vallée, N., Abry, C. 1997. The dispersion-focalization theory of vowel systems. *J. Phon.* 25(3), 255-286.

[33] Sounds corresponding to the cardinal vowels are available online from: *http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html*

[34] Stevens, K.N., 1989. On the quantal nature of speech. *J. Phon.* 17, 3-45.

[35] Stevens, K.N., 1998. *Acoustic Phonetics.* Cambridge, MA: MIT Press.

[36] Stevens, K.N., House, A.S. 1963. Perturbation of vowels articulations by consonantal context, *J. Speech Hearing Res.* 6, 111-128.

[37] Vaissière, J. 2007. Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of sounds across languages. In Sole, M.-J., Beddor, P.S., Ohala, J. (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 54-71.

[38] Vaissière, J. 2009. Articulatory modeling and the definition of acoustic-perceptual targets for reference vowels. *Chinese J. Phon.* 2, 22-33.

[39] Vaissière, J. In press. Proposals for a representation of sounds based on their main acousticoperceptual properties. In Goldsmith, J., Hume, B., Wetzels, L. (eds.), *Tones and Features.* Mouton de Gruyter.

[40] Wood, S. 1982. X-ray and model studies of vowel articulation. *Working Paper* 23. Lund, Sweden: Lund University, Department of Linguistics.

[41] Wood, S. 1986. The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *J. Acoust. Soc. Am.* 80, 391-401.

[42] Yu, S.Y., Chen, Y.D., Wu, J.R. 2010. Spectral integration and perception of Chinese back vowel /ɤ/. *Sci China Inf Sci.* 53, 2300-2309.