On acoustic salience of vowels and consonants predicted from articulatory models

Jacqueline Vaissière

Laboratoire de Phonétique et Phonologie (LPP), CNRS/Univ. Paris 3
jacqueline.vaissiere@univ-paris3.fr

Abstract

This presentation deals on acoustic salience of vowels and consonants predicted from articulatory modeling (AM). An articulatory model describes vocal tract (VT) profiles with a small number of parameters that are interpretable in phonetic terms. Such a modeling allows a better understanding of the relationship between articulatory configurations and acoustic patterns. The goal of this paper is twofold. The first part is rather an introduction to the issue and aims to illustrate AM. The advantages and limits of studying speech with different types of modeling, such as simple models of VT area function and more elaborated AM, are very briefly reviewed, and articulatory models are introduced. In anthropological Maeda's model, the area function is reconstructed with seven articulatory parameters: one for the jaw position, three for the tongue, two for the lips and one for larynx height. Such an AM makes it possible to evaluate the separate contribution of individual parameters to the vowel formants, and their interaction. In the second part, I propose some hints to complement the traditional way of transcribing the sounds and coarticulatory effects by acoustic and articulatory notation. It is based on the notion of the F-pattern defined by Fant (1960) and the use of Maeda's model to explore the maximal articulatory and F-pattern spaces. Such exploration reveals the similarities of the F-pattern among vowels, semi-vowels, and consonants that share a similar tongue position and shape and lip configuration, but differing in the constriction size and shape. Exploring the whole F-pattern space helps to define prototypical vowels specified mainly by specific acoustic characteristics, and which may serve as references. This approach makes it possible to describe subtle but audible and consistent differences between the realizations of a similar phoneme of different languages traditionally represented by the same IPA symbols. Moreover, it incorporates coarticulatory and positional influences therefore thus uncovering the seeds for perceptual confusions among sounds due to their immediate phonetic context, which in turn can cause sound changes.

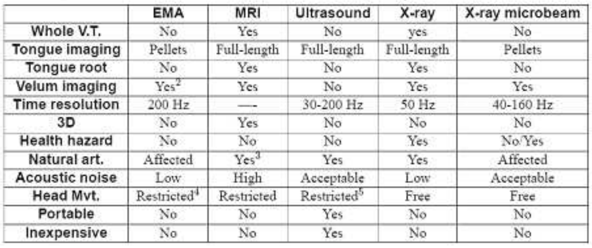
1. THE ADVANTAGES AND LIMITS OF DIFFERENT TYPES OF MODELING

1.1. Modeling of speech production

Modeling is a useful process for apprehending complex phenomena such as speech. It consists of successive approximations of the production system in terms of space and time by setting the smallest possible number of parameters, as independent as possible. It also has to be realistic enough as to produce natural sounds by specifying the values of parameters which can be phonetically, i.e. articulatory, acoustically and perceptually, interpreted. Modeling of the speech production process is based on the source-filter theory (Chiba and Kajiyama 1941; Fant 1960; Stevens 1998), where the generation of the source (phonation) and the filtering by the supraglottal cavities (articulation) are assumed to be independent. This presentation focuses on the filter model.

1.2. Data for the construction of articulatory models

An articulatory model can be constructed by statistical analysis of mid-sagittal profiles. How can one obtain the sagittal profiles necessary for articulatory modeling? *High speed X-ray* films give us the best representation of the time-varying VT. Its use is limited by the safe exposure time to X-ray radiations, however. It gives no information about the cross-sectional areas that govern the acoustics. Recent technical progresses have provided many new ways to visualize the VT (see Table 1). *MRI* is not invasive and it has the great advantage over all other methods of allowing a three-dimensional imaging of the whole VT, therefore to estimate the cross-sectional areas. The subject, however, has to sustain phonation and articulation for a rather long time, lying on his/her back. *Ultrasound echograph*, the least expensive system, is also non invasive, it can be portable. The whole VT is however not visible, the images are often noisy, and the tongue tip is often masked by the shadow of the lower jaw. *EMA* (ElectroMagnetic Articulography) and X-ray micro-beam system tract only movements of a limited number of sensors glued on the surface of the articulators, and data are not so adequate for the construction of comprehensive AM (see for a review, see the description of the European ASPI program).



	EMA	MRI	Ultrasound	X-ray	X-ray microbeam
Whole V.T.	No	Yes	No	yes	No
Tongue imaging	Pellets	Full-length	Full-length	Full-length	Pellets
Tongue root	No	Yes	No	Yes	No
Velum imaging	Yes[2]	Yes	No	Yes	Yes
Time resolution	200 Hz	—	30-200 Hz	50 Hz	40-160 Hz
3D	No	Yes	No	No	No
Health hazard	No	No	No	Yes	No/Yes
Natural art.	Affected	Yes[3]	Yes	Yes	Affected
Acoustic noise	Low	High	Acceptable	Low	Acceptable
Head Mvt.	Restricted[4]	Restricted	Restricted[5]	Free	Free
Portable	No	No	Yes	No	No
Inexpensive	No	No	Yes	No	No

Table 1: Comparison of VT shape recording techniques (From Laprie & al, 2006)

1.3. Area function and articulatory modeling

For pedagogical purpose, it is useful to represent the VT shape corresponding to different vowels by idealized tubes, such as a single tube, two tubes or a four-tube/three-parameter model. The complexity of the representation depends on the vowel.

Figure 1 illustrates mid-sagittal profiles as derived from X-ray data, the corresponding area functions estimated from the sagittal profiles, and the transfer functions corresponding to the six Russian vowels uttered by a male speaker (Fant 1960). (A *transfer function* is a compact description of the input-output relation for a linear system).

Three of these six area functions, i.e. the ones concerning /i/, /u/ and /ɑ/ , can be simulated with a limited number of acoustic tubes, as seen just below.

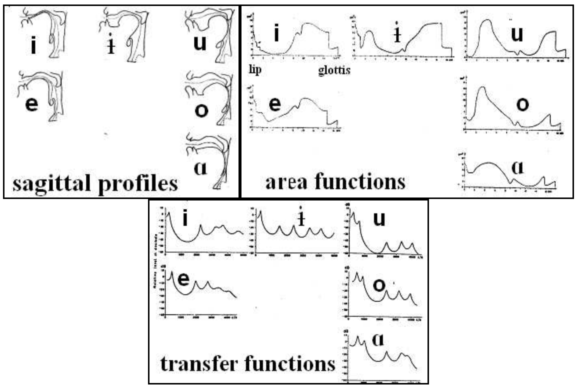


Figure 1: Sagittal profiles, area functions and transfer functions of six Russian vowels (from Fant 1960).

a) Modeling the neutral vowel

In Figure 2a, an articulatorily neutral vowel is modelled by a simple tube of constant cross-sectional area which is closed at one end (at the glottis) and open at the other (at the lips). As expected, the resonances tend to be equidistant, around 500 Hz, 1500 Hz, 2500 Hz for a tube of 17,5 cm, and higher when the tube is shorter: the resonance frequencies increase as the length of the tube decreases (for an excellent introduction to acoustic phonetics, see Johnson 1993). Note that the sounds corresponding to the various figures representing simulations are available on our web side, http://www.personnels.univ-paris3.fr/users/vaissier/pub/China2008.

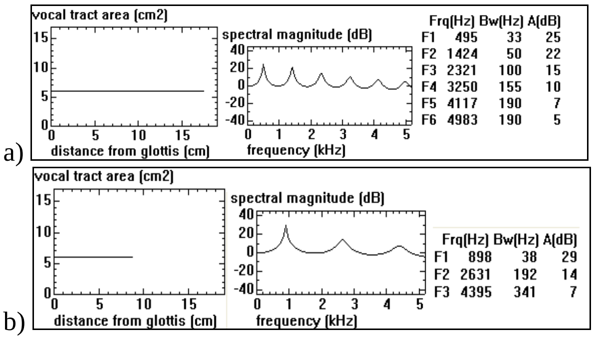


Figure 2: Representation of the neutral vowel by a simple tube 17,5 cm (a) and 8.8 cm (b) long. (S 1 and S 2). [S stands for the available sound on our web site].

According to the acoustic theory of speech production, the resonant frequencies, and thus formants, change with the position of the constriction relative to the maximum and minimum pressure points in the VT. Figure 3 (from Chiba and Kajiyama 1941) represents the position of the pressure nodes in a neutral VT: one for F1, two for F2, three for F3 and four for F4.

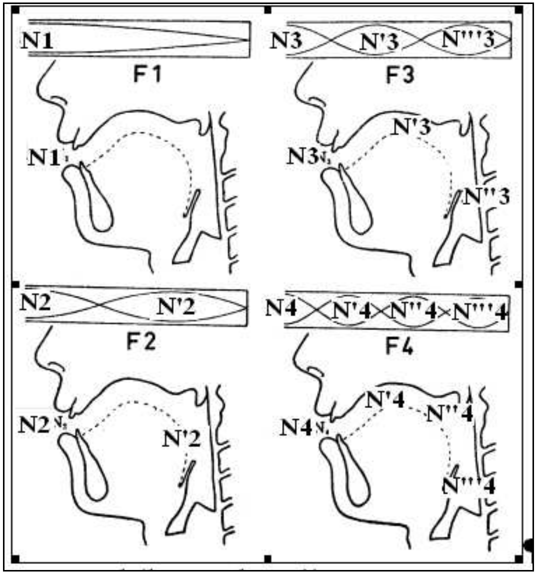
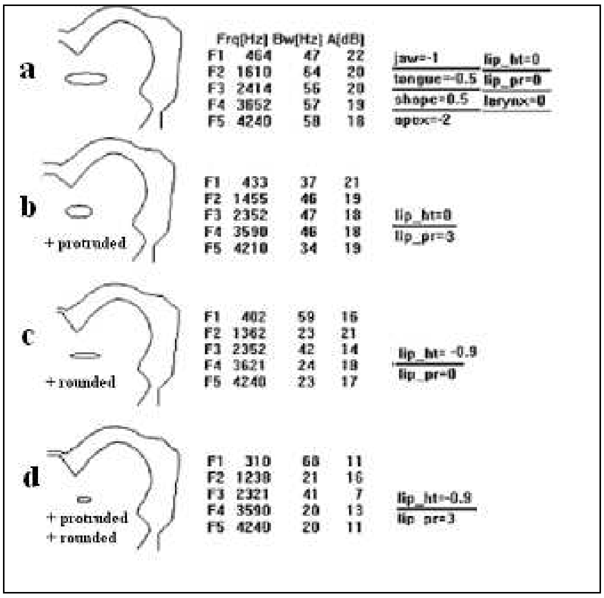


Figure 3: From Chiba and Kajiyama 1941. See text.

A constriction near a pressure node lowers the formant frequency, whereas a constriction near a pressure antinode raises it. A constriction at the point designated by Nx in Figure 3 will lead to a decrease in frequency of the corresponding formant (x refers to the assigned formant).

Let us give a first example. Figure 4 illustrates the simulation of protrusion and rounding in a neutral tube (*Figure 4*a) (points N1, N2, N3 and N4 in Figure 3). Rounding (*Figure 4*b) is twice as effective as protrusion

(*Figure 4*c) in lowering F1 (31 versus 64 Hz). But when combined (*Figure 4*d), the effect is enhanced: F1 is lowered by 154 Hz, because the ratio of area over length plays the essential acoustic role. As expected, all formants lower. Each articulatory maneuver has an effect on the different formants. In most cases, it is not possible to manipulate one formant independently of the other formants, but in most cases, a constriction has more effect on one formant than the others.



a	Frq[Hz]	Bw[Hz]	A[dB]
F1	464	47	22
F2	1610	64	20
F3	2414	56	20
F4	3652	57	19
F5	4240	58	18

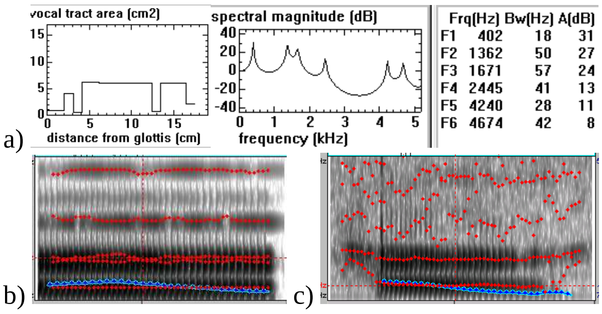
b			
F1	433	37	21
F2	1455	46	19
F3	2352	47	18
F4	3590	46	18
F5	4210	34	19

c			
F1	402	59	16
F2	1362	23	21
F3	2352	42	14
F4	3621	24	18
F5	4240	23	17

d			
F1	310	68	11
F2	1238	21	16
F3	2321	41	7
F4	3590	20	13
F5	4240	20	11

Figure 4: Neutral tube (top) and simulation of protrusion, rounding, and the combination of both protrusion and rounding. The four sounds (S 3, S 4, S 5, S 6) have a different color.

A second example: as exemplified in Figure 3 (bottom left), F2 can be lowered by two constrictions (N2, N'2) along the VT, one at the lips (N2) and one at the back (N'2). F3 can be lowered by three constrictions (N3, N'3 and N"3), and F4 by four constrictions (N4, N'4; N"4 and N'''4). Figure 5a and Figure 5b represent the results of the simulation of three constrictions along the neutral tract, at the places which lead to decreasing all formants. F3 is maximally lowered and drop from 2321 Hz to 1671 Hz, and clusters with F2, creating a concentration of spectral energy around 1500 Hz. It corresponds to an r-colored vowel, i.e. a vowel whose main distinctive acoustic cues is a low third formant. The sound has the same color than the English phoneme ɚ, in rhotic varieties of English as spoken by Peter Ladefoged (compare the two spectrograms in Figure 5b and c). There is no single articulatory way to produce a sound with well-defined acoustic characteristics, because of large compensatory possibility (as illustrated in paragraph 2.6). ɚ can be produced by a compound of labialization, bunching and backing of the tongue, corresponding respectively to N3, N'3 and N'''3 in Figure 3. The fourth constriction, at the glottis end, represents the laryngeal cavity.



	Frq(Hz)	Bw(Hz)	A(dB)
F1	402	18	31
F2	1362	50	27
F3	1671	57	24
F4	2445	41	13
F5	4240	28	11
F6	4674	42	8

Figure 5: a) Neutral vowel with a laryngeal tube and three constrictions, its transfer function and formant values. b) and c): Spectrograms of the resulting synthetic sound (S 7) to be compared with Peter Ladefoged's same vowel (S 8). Note that in the case of the natural sound, Praat detected only one formant for Ladefoged's converging F2/F3.

b) Modeling the VT by two tubes for /a, /i/, /y/ or by three tubes for /u/

All phonemes (except the neutral vowel) are created by characteristic perturbations of the cross-sectional area of the VT. The two vowels /ɑ/ and /i/ can be adequately modelled by two tubes (Figure 6a and b). For real sagittal profiles of /ɑ/ and /i/ (and /u/), see Figure 1.

*Vowel /a/ (*Figure 6a)*:* the retraction of the tongue body in the production of /ɑ/ narrows the pharyngeal region and widens the front of the mouth, which can be simulated (or modelled) by two tubes. Such a change from the neutral configuration leads concomitantly to raising F1 (↗F1) and lowering (↘F2), and therefore to the convergence of F1 and F2 around 1000 Hz.

*Vowel /i/ (*Figure 6b)*:* the narrowing of the palatal region by the raising of the tongue in the case of the production of /i/ can be represented by a narrow front tube. Such a palatal constriction leads to lower F1 (↘F1) and raise F2 (↗F2), and to increasing the distance between these first two formants.

*Vowel /y/ (*Figure 6c): Let us remark that the protrusion/rounding have a lowering effect on all formants. The magnitude of the effect of each gesture is however dependent on the global shape of the VT. When the tongue has no constriction (see Figure 4d and e), or has an /u/-like shape (see Figure 6), rounding and protrusion affect both F1 and F2, but no so much F3 (as explained in 2.3). Rounding is the key gesture to lower F2 (and F1) when the constriction is in the middle of the VT. When the tongue adopts an /i/-like, F3 is particularly sensitive to lip configuration. Rounding and protrusion will change /i/ into /y/ in French (Figure 6c).

*Vowel /u/ (*Figure 6d)*:* Four tubes are necessary to model the vowel /u/ (as the Cardinal Vowel number 8 [u], as in French, but not English or Japanese), which has a very low F2. The extreme lowering of F2 needs two constrictions one in the middle of the VT, and one at the lips (see Figure 3). If there is no lip constriction (compare Figure 6d with Figure 6e), F1 and F2 will rise and the vowel would sound like a central vowel (at least to a French listener). /u/ corresponds to the most “grave” vowel.

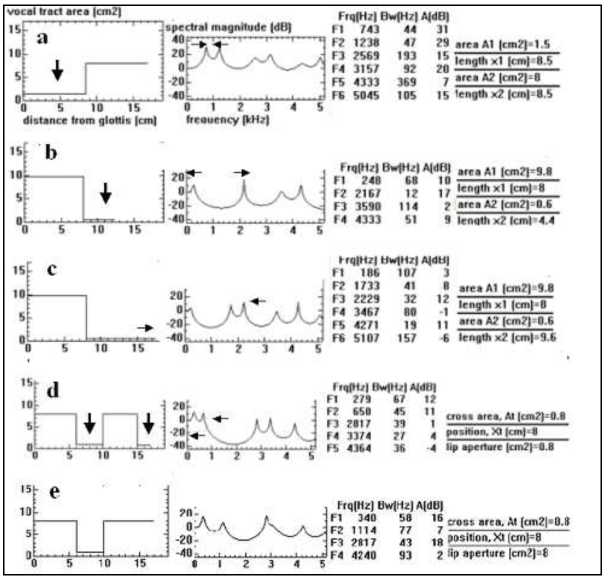


Figure 6: Representation of a vowel with posterior constriction (/a/-type: a), with anterior constriction (/i/-type: b) and lengthening of the front tube (/y/-type: c), and with central constriction with lip rounding /(/u/-type: d) and without rounding (e). S 9, S 10, S 11, S 12, S 13.

c) Modeling all oral vowels by a three-parameter model and four tubes

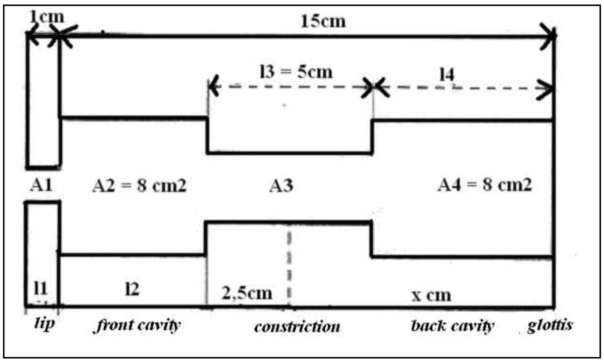


Figure 7: Fant’s first 3-parameter model that consists of four tubes. The three variables are A1, A3 and l4. See text.

The whole VT can be represented by a small number of parameters, which are more or less interpretable in articulatory terms. According to Fant (1960), the VT transfer function for the vowels can be accurately predicted from four-tubes specified by three parameters. These three parameters are the distance from the glottis to the major constriction, the constriction area, and the length over area of the mouth opening, see Figure 7 (See also Stevens and House 1955; Lindblom and Sundberg 1971). A1 specifies the lip configuration with a constant length of 1 cm, and the other parameters include the tongue position (l4 in Figure 7) and degree of constriction (A3). With this 3-parameter model, Fant has provided for the first time the famous “nomograms” that specify the mapping between the VT configuration and formants (nomograms are exemplified in Figure 11, Figure 15 and Figure 16).

d) Geometrical models

In geometrical models, the VT is represented on a more realistic way than with straight tubes. It is determined by the position of the articulators (jaw, tongue, lips, velum, larynx). In the first attempts, the tongue was represented by a circle (Coker and Fujimura 1966; Mermelstein 1973).

e) Data based models

In Harshman et al’s model (1971), the observed tongue profiles of 10 English vowels are well described by the weighted sum of only two major components (determined by using a procedure PARAFAC based on the principal component analysis): one factor for a forward movement of the back of the tongue concomitant with an upward fronting movement of the tongue blade and one for a forward of the tongue root associated with the upward backing movement of the tongue body. Neither the jaw nor the lips are taken into account, so that model should be considered somewhat incomplete.

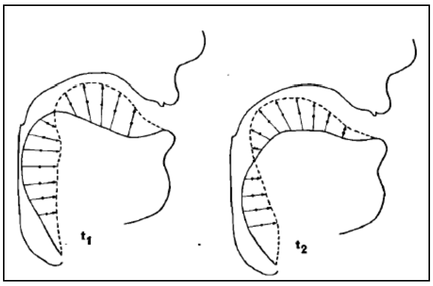


Figure 8: A graphical representation of the two factors that combine to form VT profiles of 10 English vowels (Harshman et al 1977).

All the preceding models, including Harshman et al’s, do not take jaw height into account, while it is known that the jaw is generally continuously moving in running speech, and it largely determines tongue height and lip aperture (note that it is however possible to speak with a clenched jaw).

1.4. Anthropologic Maeda’s articulatory model

From sagittal profiles to the definition of articulatory parameters

The representation by simple tubes and the nomograms are very useful tools for understanding basic acoustic laws, and the affiliation between cavities and formants, but they can generate patterns beyond human capability. For example, human speakers can produce vowels only over a range that is less than half that represented in Fant's nomograms (Ladefoged and Bladon 1982). The main advantage of AM is to generate area functions that could be generated by a human VT.

Maeda's anthropologic articulatory model is based on the statistical analysis of lateral X-ray images and labiofilms from running speech (Maeda 1979; Maeda 1988; Maeda 1989). Roughly 1000 frames of cineradiographic and labiofilm data on the VT corresponding to 10 French sentences uttered by two speakers have been digitized using a grid (Figure 9) and analyzed statistically.

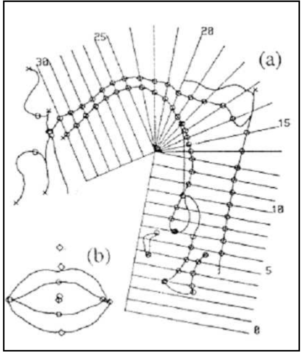


Figure 9: The grid used by Maeda for measuring the shapes of the tongue.

Maeda used a guided PCA to extract first the effect of the jaw position. The seven parameters found by the statistical analysis are interpretable in traditional articulatory terms, which make it possible to link between sagittal profiles and more abstract distinctive features, such as front/back or +anterior and - anterior. The seven parameters are the following: one parameter for the jaw position (1), three parameters for the tongue: tongue dorsum position (2) and tongue shape (3) and tongue apex position (4), two parameters for the lip: aperture (5) and protrusion (6), and one for larynx height (7). The parameters correspond to the four speech organs (i.e. the jaw, tongue, larynx and lips), which manipulate the resonance characteristics of the supraglottal cavity to articulate the sounds (Maeda's model includes a nasal tract, which will be not discussed in the present paper; Maeda 1993).

Figure 10 illustrates the articulatory synthesis with Maeda's model of six French vowels, for which a representation by simple tubes for /ɑ/, /i/, /y/ and /u/ has been discussed before (paragraph 1.3). Note the similarities in the spectral patterns. It is fruitful to compare the transfer functions obtained by Fant (Figure 1), and those obtained by simple tubes (Figure 6): compare for example /i/ in Figure 10 with Figure 6b, /y/ with Figure 6c, /ɑ/ with Figure 6a, /oe/ with Figure 2a and /u/ with Figure 6d.

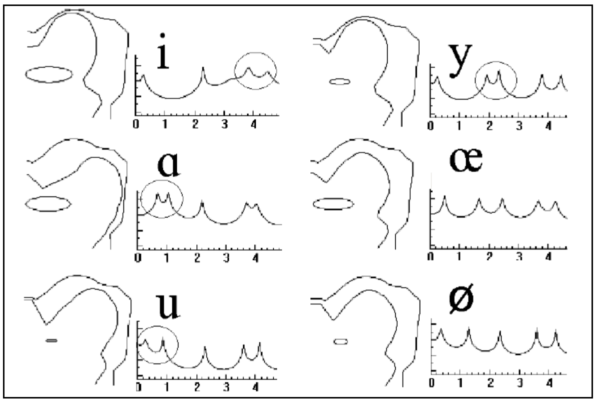
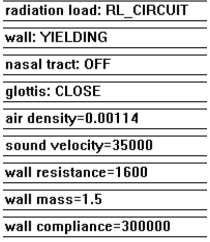


Figure 10: Reconstruction of six French vowels, using Maeda's model (from Vaissière, 2007a).S 14 S 15 S 16 S 17 S 18 S 19.

Note that Maeda's program (as most of the programs) makes it possible to control a number of other acoustic and biomechanical parameters concerning well-known effects: the effect of the *radiation load* (the bandwidth of the higher formants increases as formant frequency increases, Fant 1960), the effect of the *walls* at very low frequencies (the bandwidth of the lowest formant -under 500 Hz- and its frequency increase as the degree of constriction increases), the state of the *glottis* (open/close: when the glottis opens, the bandwidth of the formants associated with the back cavity increases), the opening of the nasal tract (connected/not connected), the *air density* and the *sound velocity*. Table 2 indicates the default parameters, which have been used for constructing the examples in the present paper.



radiation load: RL_CIRCUIT
wall: YIELDING
nasal tract: OFF
glottis: CLOSE
air density=0.00114
sound velocity=35000
wall resistance=1600
wall mass=1.5
wall compliance=300000

Table 2: Default acoustic and biomechanical parameters used in Maeda's model.

Three remarks:

a) The model is speaker-dependant. It corresponds to the single VT of a single subject. Anatomical differences of VT dimension (or size) between speakers are far from being negligible. But the same acoustic laws apply to all speakers and the articulated shapes of the VT are about the same when constricted in certain regions. Therefore a single model makes it possible to explore the capability of a human to create sounds. Badin and al (1990) compared (i) the nomograms produced by Fant's four-tube model, (ii) Ladefoged and

Bladon's attempt to reproduce the nomograms obtained by Fant, and (iii) nomograms created by Maeda's model. The general tendencies are similar. The nomogram produced by the articulatory model is closer to Ladefoged and Bladon's attempt than the four-tube model (for details, see Badin et al's paper): no lowering of F2 when the constriction is at the front end of the VT is observed in both cases.

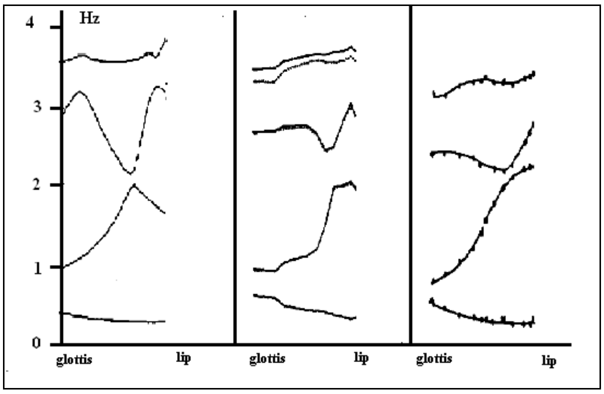


Figure 11: Nomogram produced by Fant's four-tube mode (left) and by Maeda's model (mid) when the constriction location is varied from 5 to 13 cm from the glottis. Right: attempt to imitate Fant's nomograms by A. Bladon (adapted from Badin et al 1990).

b) The model is language-dependant. It appears most likely that other parameters may have turned out to be relevant is the original data were collected from a language other than French. French has a rather large number of vowels (n=16 including the 4 nasals), it used labiality and nasality for contrasting the vowels, but it does not use, for example, retroflexion or the feature [ATR] (Advanced Tongue Root).

c) The cross-sectional areas are predicted from mid-sagittal dimensions using the most commonly adopted classic method proposed by Heinz and Stevens (1965), which is based on the assumption that the tongue has a flat surface, and that the opposite VT wall has a fixed, parabolic shape. A number of works have actually been carried out, using newly available 3D MRI data to obtain more detailed sagittal-to-area transformation. A given mid-sagittal distance does not necessarily have a unique corresponding area, but multiple possible areas depending on the identity of the articulation (see figure below, from Ericsdotter, 2007).

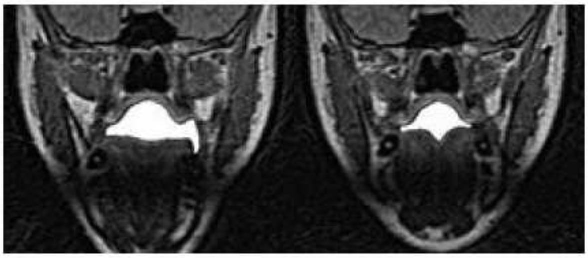


Figure 12: Example of articulations where the mid-sagittal distance is the same, but the cross-sectional areas differ in shape and size (by 2.5 cm^2. Male subject, velar region, vowel œ (left) and ʉ̟ (right) (from Ericsdotter, 2007).

1.5. From the 7 parameters to area function, then to transfer function

For synthesizing an oral vowel in Maeda's model, the 7 articulatory parameters are given as input to generate a mid-sagittal profile, from which a cross-section area function is derived; the transfer function is then calculated and the resulting sound is generated (*Figure 13*).

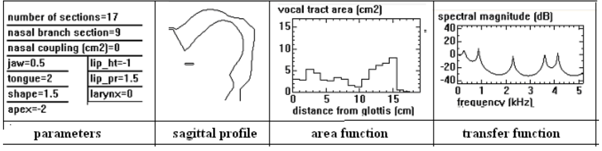

number of sections=17
nasal branch section=9
nasal coupling (cm2)=0
jaw=0.5
lip_ht=-1
tongue=2
lip_pr=1.5
shape=1.5
larynx=0
apex=-2

Figure 13: The 7 articulatory parameters used (left) to generate a sagittal profile, which is transformed into an area function, and then to a transfer function (left).S 20.

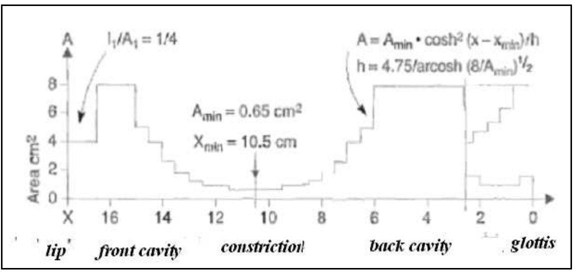
2. F-PATTERN, CONSTRICTION SIZE AND POSITION

2.1. The notion of F- pattern

What does F-pattern means? The filter function of the VT without branching cavities contains only resonances (i.e. no zeros), labeled F1, F2, F3, F4 and so on. The term F-pattern (Fant 1960:209) is a compound for a specification of these frequencies. It can be calculated for any known VT shape.

Figure 14 represents the second more elaborated three-parameter VT model. Figure 15 and Figure 16 are derived nomograms, illustrating the separate effect of the three parameters. Figure 15 illustrates the evolution of the F-pattern when the degree of constriction (the ration (l1/A1) is changed; Figure 16 represents the evolution of the F-pattern when the degree of constriction is fixed but the constriction position, and/or the front cavity opening are modified (Fant 1960).

a)



X (cm)→

Figure 14: The second more elaborated three-parameter VT model (Fant 1960).

2.2. The effect of the cross-sectional constriction area

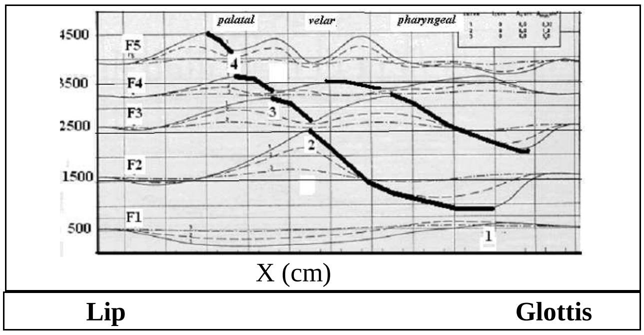


Figure 15: Nomogram derived using the second three-parameter VT model. X is the constriction coordinate in centimeters from glottis. Effect of the tongue constriction size: 0,32cm2 (plain line), 1.3 cm2 and 5 cm (dotted lines). See text.

The magnitude of the formant shift varies with the constriction size.

When the tongue passage becomes narrower (Figure 15), the *deviation* of the F-pattern from that of the natural values of a uniform tube (500 HZ, 1500 HZ, 2500 Hz, etc.) become larger and the acoustic dependency between the cavity in front of the constriction and that behind it increases. When the constriction becomes very narrow (fricative), or complete (stops), mainly formants due to the front cavity will determine the acoustic characteristics of theses consonants. But very importantly, the effect of varying the place of constriction on the F-pattern is *qualitatively* the same.

2.3. The effect of the constriction position and lip rounding

Figure 16 illustrates some remarkable points.

1) Converging formants: When one formant is maximally high of low, it tends to merge with another formant, the amplitude of the two formants is mutually enhanced; the spectrum is dominated by a concentration of spectral energy in the region where the two formants converge, creating a spectral dominance in that region.

2) *Rounding:* The points of convergence represent place where less articulatory precision is required, and formants frequencies are insensitive to small changes in the constriction location, according to the Quantal Theory (Stevens1989). The latter statement is only valid if lip configuration is not taken into account: F3 is very sensitive to lip configuration in the palatal region, and F2 in the velar region, a sensibility which is widely used in the languages of the world.

As for the effect on F3 in palatal vowels, lip rounding is used on many languages to contrast between round palatal vowels, such as /y/, and unrounded (spread) palatal vowels, as /i/, such as in German, French or Swedish. According to Wood (1982), languages, such as Swedish and French, prefer the prepalatal position for both [i] and [y], a position which is the most sensitive to lip configuration.

As for the effect on F2 in mid-constriction (such as /u/) and back-constriction vowels (such as /A/), rounding, position and size of the constriction are often adjusted in languages to keep the two first formants clustered (in Stevens's terms, rounding will enhance the back feature), with a spectral integration within the 3.5-Bark critical distance.

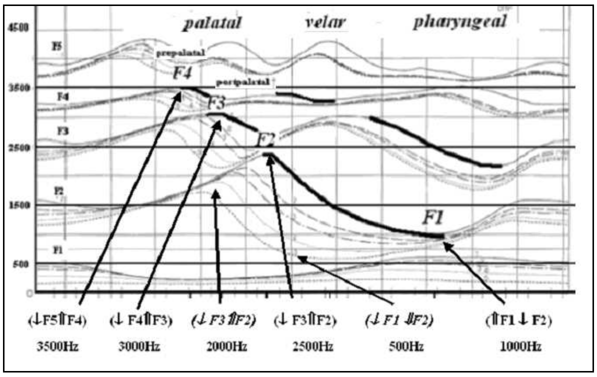


Figure 16: Nomogram that has been derived using the second three-parameter VT model represented in Figure 14. The tongue constriction size is fixed at 0.65 cm2. Effect of lip configuration: no lip rounding (plain lines), and decreasing lip area: 8, 4, 2, 0,65 and 0,16 cm2 (dotted lines). F1, F2, F3 and F4 indicate points where F1, F2, F3 and F4 are maximally high, respectively. Thick line represents the resonances due to the front cavity. (FnFn+1) indicates convergences between two formants. See text.

3) There are 5 remarkable constrictions.

a) pharyngeal */**a**/-region* : As the constriction is near the glottis, there will be first a convergence between the two lowest formants (F1F2) around 1000 Hz. F1 frequency is maximal (⇑F1) and so is the ⇑(F1F2) cluster. It corresponds to the traditional most "compact" vowel. (See Figure 6a and Figure 16, point F1).

*b) velar /**u**/-region:* When the constriction is in the velar region and the lips are rounded, another remarkable point is created: F1 is low and F2 are low, creating now the lowest possible concentration of energy. (See Figure 6d).

*c) mid-palatal /**i**/-region*: When the constriction is at the palatal region and the lips are spread, F2 and F3 converge in the higher frequencies, around 2500 Hz, at a place where F2 is maximal (⇑F2). (See Figure 6b and Figure 16, point F2).

*d) prepalatal /**i**/-region:* The fourth remarkable convergence is (F3F4), when F3 is maximally high. It happens above 3000 Hz. Let us call this region the creating ⇑(F3F4) with the highest F3 as possible. (See Figure 6b and Figure 16, point F3).

e) */y/-region:* Very interestingly, in this region where F2 and F3 are stable, F2 is highly sensitive to lip rounding. Lip rounding allows to transform ⇑(F3F4) into an (F2F3) convergence by lowering F3. (See Figure 6b and Figure 16, point F3, with rounding).

Table 3 summarizes the observations done on the position of the constriction, the lip configuration and the shape of the tongue.

Location of the constriction			
Lowest possible formant		**Highest possible formant**	
⇓F1	anterior part of the VT	⇑F1	posterior part of the VT
⇓F2	velar region + *lip rounding*	⇑F2	mid-palatal region + glottal region
⇓F3	pharyngeal region + *bunching of the tongue, retroflexion* + *lip rounding and lip protrusion*	⇑F3	apical and prepalatal regions + lip spreading + glottal region *(larynx lowering)*

Table 3: Tongue constriction position, lip configuration, tongue shape and F-pattern values: a summary.

The nomograms in Figure 15 and Figure 16 illustrate the effect of only two constrictions. Retroflexion, as a third constriction, allows to counteract the raising effect of palatalization on F2 and F3, by lowering F3.

Because the effect of constriction size is quantitative rather than qualitative, we recognize (at least) five acoustic classes, for a given place of articulation for which the modification of F-pattern due to a change in the size is similar across a vowel having a wide constriction area, a semi-vowel, a fricative with a narrow constriction, and a stop with complete closure.

The 5 acoustic classes correspond to five articulatory classes defined in terms of tongue constriction locations: labial (with constriction at the lip), neutral (without major constriction), palatal (the tongue is in a palatal, fronted position); velar (constriction at the middle of the VT, /u/-like shape), and pharyngeal, back, /ɑ/-like shape) (see Figure 16). The whole VT has to be taken in to account to calculate the F-pattern and very large compensatory manoeuvres can take place (for compensation between lip configuration and tongue parameters, see Maeda1989, 1990).

2.4. Acoustic similarities between consonants, semi-vowels and vowels

All phonemes sharing a similar sagittal profile would have a similar F-pattern.

We will first show similar F-pattern due to the phoneme identity, second, the variations due to the context on a given phoneme due to the surrounding phonetic context, and third the effect of phenomena such as palatalization.

a) Figure 17 illustrates the similarity between the mid-sagittal profiles corresponding to three phonemes with a constriction in the palatal region. Figure 17a illustrates the palatal vowel /i/ and the palatal semi-consonant /j/. Figure 17b illustrated the contextually palatalized /g/.

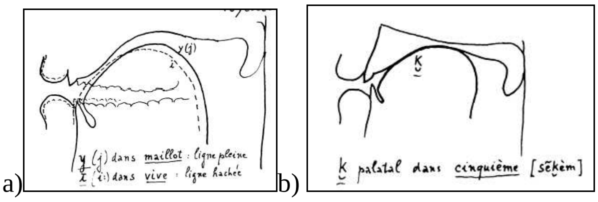


Figure 17: Sagittal profiles corresponding to (/i/) –in the solid line) and /j/ (in dashed line) in French (a) and of contextually palatalized /k/ (b) (from Straka 1965).

Figure 18 illustrates the similarity in terms of F-pattern. Between the palatal semi-vowel /j/ in two different vocalic contexts (/u/ and /a/ and the two contextually palatalized consonants, /g/ and /l/. They all display concentration of spectral energy in the F3-F4 region.

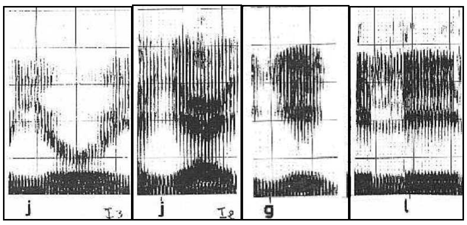


Figure 18: Spectrograms of /juj/, /jaj/ , /gi/ and /li/.

b) The F- pattern of any phoneme is partly determined by the F-pattern of the surrounding phonemes. The effect of a vowel on a consonant is illustrated in Figure 18. /u/ has an (↘F1↘F2↘F3↘F4) effect, and /a/ mainly an (↗F1) effect. As expected, the cluster (F3F4) of /j/ is lowered in the /u/ context and F1/j/ is raised in the /a/ context. It also illustrates the effect of a consonant on a vowel. /j/ an (↘F1↗F2↗F3↗F4) effect on /u/ and /a/. F1/a/ is lowered, F2/u/ and F2/a/ are raised. The magnitude of the effect on the surrounding phonemes depends on the language, and on the prosodic position of the phoneme, but the direction of the shift is always the same.

c) As well known, secondary constriction such as palatalization, velarization, and pharyngealization have respectively the same effect that the vowels /i/, /u/ and /ɑ/.

Figure 19 illustrates the similarities between the area functions of the Russian vowel /i/ and two phonologically palatalized consonants /fʲ/ and /sʲ/ (see X-ray data in Fant 1960:219).

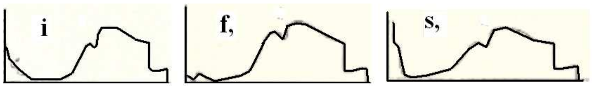


Figure 19: Sagittal profiles of /i/, and palatalized /fʲ/ and /sʲ/ (Russian sounds, from Fant 1960).

The three area functions correspond to an F-pattern with an high F2 (around 2000 Hz). Since the F-pattern during the consonant is the main determinant of the value of the formants at the vowel onset just after the release of the fricative (or stop) consonants, the same F2 frequency at vowel onset (around 1900-2000 Hz, Fant 1960: 221) is predicted and observed, independently of the place of the closure. The transitions into the vowels, therefore, do not carry information about the place of main constriction for the palatalized consonants (for the cues necessary to differentiate between the different palatalized consonant, see 3.2).

2.5. From fricatives to sonorants

As well known, in weak position, the glottis tends to be less open, and the tongue or lip constriction less tight. By changing the relative size of the glottis and constriction area, AM makes it possible to simulate sound changes and sound reduction as it happens in continuous speech.

Depending on the relative size of the glottis and the constriction area, the same F-pattern would correspond to a fricative, to a sonorant or to a vowel.

Figure 20 illustrates the modeling of an uvular consonant with three tubes: a back tube, a tube for the constriction and a front tube. A supraglottal noise is created downstream from the constriction, when the supraglottal constriction is small enough relatively to the glottis opening. In the first case, the opening at the glottis is smaller than the supraglottal constriction (0 versus 0.2), the sound is voiced, the source is at the glottis; the whole F-pattern (F1 to F4) is excited; the sound looks like a sonorant. In the second case, the opening at the glottis is larger than the supraglottal constriction (0.5 versus 0.25), the sound is unvoiced, the source is at the supraglottal constriction; only the resonances of the cavity in front of the constriction are excited, i.e. F2 and F4; the sound looks like a fricative (Yeou and Maeda 1995). Both sounds are allophones in French and in Arabic, they correspond to two different phonemes.

If the constriction tube is suppressed, the sonorant is transformed into a /ɑ/-like sound.

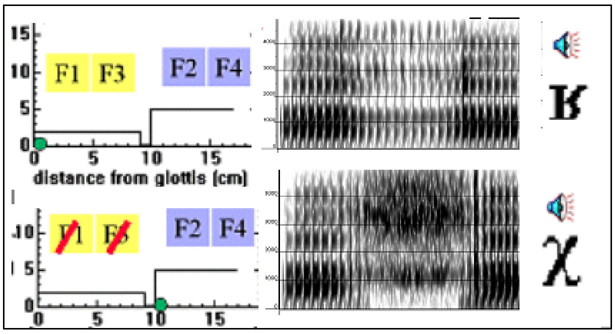
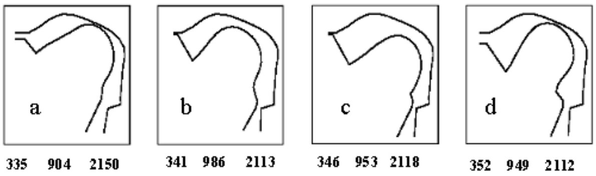


Figure 20: The constriction size has been set to 0.2 and 0.5 cm², respectively, and the glottis to 0 and 0,25 cm² (Yeou 1993). The affiliation between formants and cavities are indicated; crossed formants indicate that the formants are not excited. (S 21, S 22)

2.6. Compensatory and enhancement

An identical F-pattern can be the outcome of more than one articulatory state, allowing compensatory and reinforcing gestures between articulators.

Figure 21 illustrates VT configurations giving rise to similar F1, F2 and F3 values, as indicated, using a codebook based on Maeda's model.



Frequency of the first three resonances (in Hz)

Figure 21: Left: VT configurations giving rise to similar F1, F2 and F3 values, as indicated, using a codebook based on Maeda's model (from Ouni et al 2001).

Jaw and tongue dorsum positions compensate each other for palatal unrounded vowels; lip aperture and the jaw for the back rounded series (Maeda 1990).

Such a flexibility allows a number of strategies to perform a different gestures for the realization of a phoneme, depending on the articulatory requirements for the surrounding contexts, and still obtain an adequate F-pattern for that phoneme.

3. THE CONSONANTS

3.1. Calculating the F-pattern when the VT is close

The F-pattern is invisible during the closure portion. The F-pattern as a whole becomes visible when the closure is not complete, as often happens in continuous speech (see the consonant /g/ in Figure 18). It can be calculated for any known shape of the VT, including when there is no source, such as during the closure portion of stop consonants.

As described by Fant (1960), the F-pattern during a labial consonant is largely influenced by the general tongue shape required by the production of the surrounding vowels. It is thus expected that the F-pattern for the labial consonants will be dependant on the following vowel. Figure 22 displays the evolution of the sagittal profile in the sequence /ebu/: the tongue gets more and more retracted and protruded during /b/ closure.

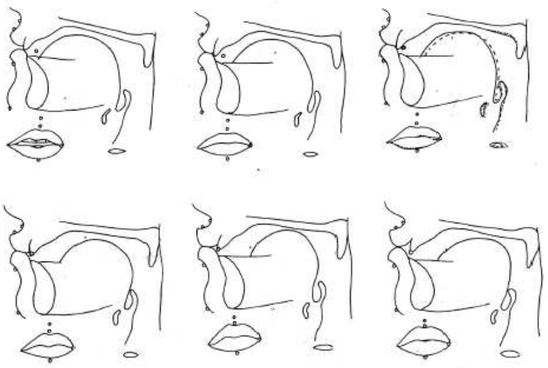


Figure 22:Moving tongue during the production of /b/ in /ebu/ (from Bothorel et al 1986).

AM makes it possible to simulate a [b]. Figure 23 illustrates such a simulation: the tongue is assumed to take exactly the shape required for the following vowel, /a/ (top), and /i/, bottom), during the labial closure, such as in the sequences [aba] or [ibi]. Lip closing is simulated by an extreme rounding of the lips (as suggested by Fant), which is compared with an open lip. As it can be extrapolated from the simulation, all formants will be raising at the labial release. F3 will be extremely affected during the /i/ case/: the transition will be equivalent, but much more rapid that the transitions in the /y/-/i/ transitions. The values found for the formant value for F2 when the opening is compared very well with Fant's results (Fant 1970; 128).

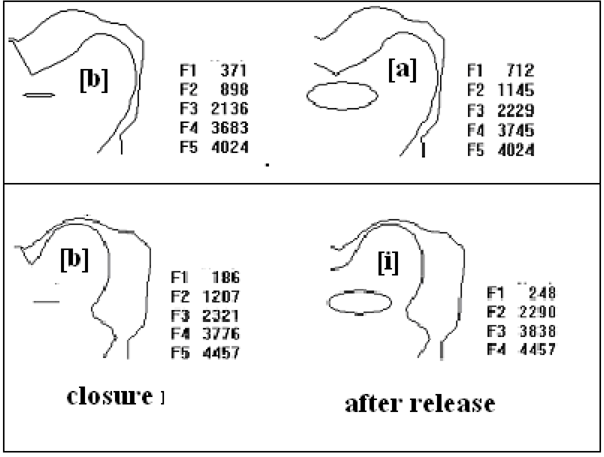


Figure 23: Simulation of the F-pattern during the quasi-closure and just after release of the consonant [b] in [ba] (top) and [bi] (bottom). S 23 S 24 S 25 S 26.

3.2. The F–pattern during the noise portion of stops and fricatives

A stop is composed of a closure portion, a release, a friction part, an aspiration portion (if aspirated), and transitions to the vowel. During the release and the friction parts of the consonant, only the resonances due to the front cavity become visible. Such resonances are represented by the thick lines in Figure 16. Then the formants (from F2 up to F5) manifest themselves due to the excitation by the aspiration at the glottis. All formants are finally visible during the transitions into the following vowel.

During voiceless fricatives, (only) the front cavities's resonances are mainly excited by noise created at the vicinity of the constriction and will be therefore visible on a spectrogram.

During voiced fricatives, the resonances due to the back cavities are also excited by the glottal voice source, and therefore the complete F-pattern becomes visible on the spectrogram

The thick lines in Figure 15 display the resonances due to the front cavity, which are excited during the frictional part of the stops , and during the fricatives. As the constriction moves along the VT, from the palatal region to the pharyngeal region, the F5, then F4, F3, and F2 are excited, creating a range of fricative noise with lower and lower frequency limit. For example, /s/ is a (>= F4) consonant, in the sense that F4 determines the lower frequency limit of the noise spectral band. In the same vein, /ʃ/ an F3 consonant and the consonants with a constriction in the back of the tongue would be F2-consonants, such as the uvular consonants in Figure 20. Very importantly, the F-pattern evolves continuously across vowels and consonants. and the formant which are due to the front cavity display a quasi-continuity with the following vowel.

The palatalized consonants (Figure 19) can be only distinguished in Russian by the noise characteristics, since the transitions into the vowel depend on palatalization, and not on the location of the primary constriction.

3.3. Shape of the tongue

Not only the constriction location, but the shape of the tongue also plays a relevant role in shaping the sounds. Figure 24 shows sagittal profiles corresponding to a typical apico-alvealor /t/ and a lamino-dental /t/: the change in the shape of the tongue is often concomitant to a change in place: apical tend to be alveolar and laminal tend to be dental; secondarily, apicality leads to a backing of the tongue (so it has an ↗F1). There is no clear different between the acoustic effect of retroflexion and alveo-apicality. The retroflexion effect are somehow orthogonal to the general fronting of the tongue: F2 and F3 are lowered. Depending on the exact of location, it will mainly have a lowering effect on F2 or F3.

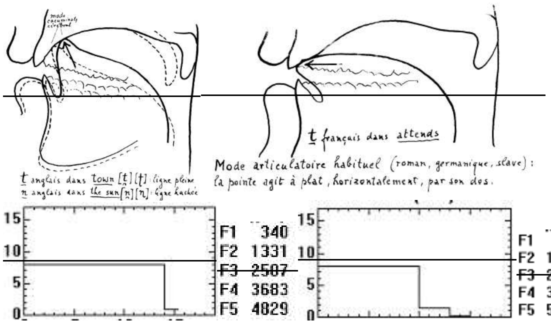


Figure 24: Top: typical /t/ in English (in "town) and in French (in "attends") (from Straka 1965). Bottom: simulation of a different shape of the tongue.

4. THE VOWELS

4.1. Reference vowels

The cardinal vowels have been proposed as a set of reference vowels for transcribing the vocalic sounds of the languages. They are described in articulatory terms: (1) close/ mid-close/ mid-open/open, (2) front/ central/ back, and (3) rounded/ unrounded.

Figure 25 represents the spectrograms to the cardinal vowels as pronounced by Daniel Jones himself, and by Peter Ladefoged, and typical French vowels. The French vowels are often proposed as adequate representative of the cardinal vowels, as they were by Daniel Jones himself. All vowels (except the contrast between /a/ and /ɑ/, the latter sounding nasalized) uttered by Ladefoged sound perfectly "French". Unlike Jones's /y/ (which does not sound French), Ladefoged's /y/ displays convergent F2 and F3 around 1900 Hz (the formant detection program could not separate the two clustered formants). Unlike Jones's front vowels other than /i/ and /y/, Ladefoged's ones display F3 at mid-distance between F2 and F4, which are typical of French.

Six of the vowels represented in Figure 25 display a grouping of two formants for six vowels. In the /i/ case, ($\underline{\text{F3}}$**F4**)], there is a concentration of spectral energy in the high frequencies; around 3000 Hz, by grouping F3 and F4 for /i/. In the French /y/, [($\underline{\textbf{F2}}$F3)], around 1900 Hz by grouping F2 and F3 (and F3 becomes a resonance of the back cavity). In the case of the four back vowels, the concentration of spectral energy in the low frequencies is created by the grouping of F1 and F2: around 1000 Hz for /ɑ/, the lowest possible for /u/, and intermediate for /ɔ/ and /o/.

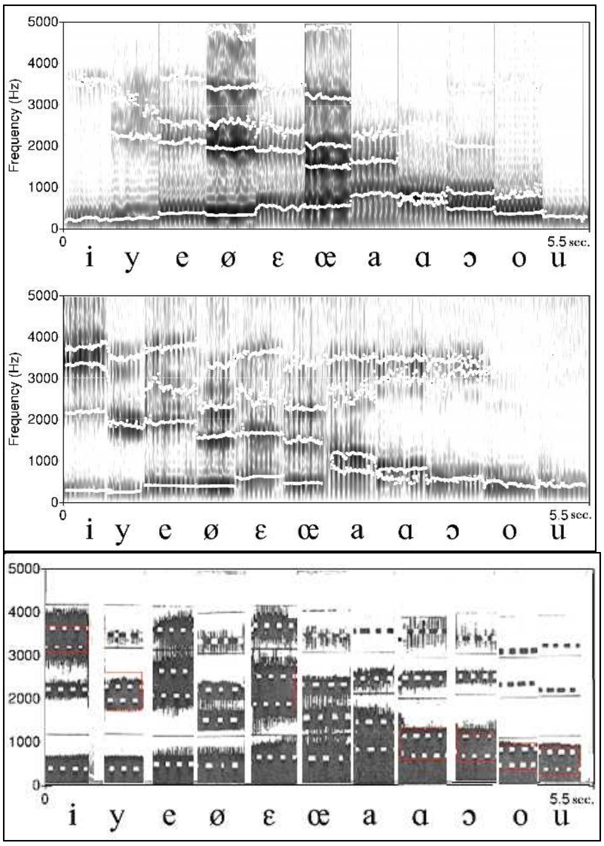
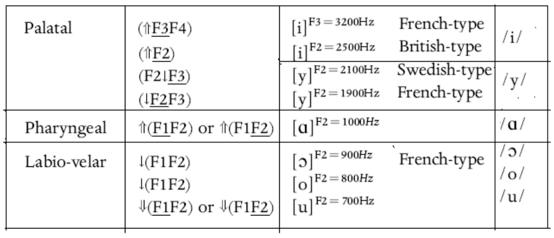


Figure 25: the cardinal vowels as proposed by Daniel Jones (top) and Peter Ladegofed (mid) and typical French vowels (bottom). All male speakers. 'f' designs the focal vowels.

The spectral dominance region corresponds approximately to the so-called *effective formant* (or F'2): around F4 for /i/ (for Swedish listeners), F3 for /y/ and F2 for the back vowel (Carlson et al 1970). It also corresponds to the "*main pitch*" of the vowels, as often mentioned by Fant (1960), which in turn correspond to the spectral peak in the burst of /k/ and /g/ located before the vowels).



Palatal	(⇑$\underline{\text{F3}}$F4) (⇑$\underline{\text{F2}}$)	[i]$^{F3=3200Hz}$ [i]$^{F2=2500Hz}$	French-type British-type	/i/
	(F2↓$\underline{\text{F3}}$) (↓$\underline{\text{F2}}$F3)	[y]$^{F2=2100Hz}$ [y]$^{F2=1900Hz}$	Swedish-type French-type	/y/
Pharyngeal	⇑($\underline{\text{F1}}$F2) or ⇑(F1$\underline{\text{F2}}$)	[ɑ]$^{F2=1000Hz}$		/ɑ/
Labio-velar	↓(F1F2) ↓(F1F2) ⇓($\underline{\text{F1}}$F2) or ⇓(F1$\underline{\text{F2}}$)	[ɔ]$^{F2=900Hz}$ [o]$^{F2=800Hz}$ [u]$^{F2=700Hz}$	French-type	/ɔ/ /o/ /u/

Table 4: Representation of some "focal" vowels.

Table 4 represents those vowels where two formants converge. The bracket indicates the convergence of the two formants. Underlined formants depend on the front cavity.

4.2. **Example: the vowel /**i/

Theoretically there are at least three types of focal /i/ (with two convergent formants): [(⇑F4F5)], [(⇑F3F4)], or [(⇑F2F3)].

Prototypical /i/ in French is a [palatal (⇑F3F4)], with a clear convergence of the formant F3 and F4, creating a concentration of spectral energy above 3000 Hz (for some speakers, [(⇑F4F5)] has also been attested). Jones and Ladefoged's /i/ display the same [(⇑F3F4)] characteristics. Simulation with AM leads to the conclusion that French /i/ is the most acute voiced, noise-free sound that a VT could generate (Vaissière, 2007b).

Gendrot et al (submitted) calculated the formant frequencies in eight languages, and confirm the particularity of the French /i/: it has, on the average, the highest F3 and the smallest (F4-F3) distance.

	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 (55)	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)

Table 5: Mean values of F1, F2, F3 and F4 for the vowel /i/ in eight languages. The standard deviation are indicated in parentheses. In addition, the distances between F3 et and F4 are indicated.

Two types of vowels /i/ are presented. The first represents the cardinal vowel /i/, [⇑F3F4)], with a clustering of F3 and F4. It corresponds marked as F3 in Figure 16.

According to Fant, a prepalatal constriction leads to the highest F3 (with lip spreading), and an eventual clustering with F4, and a mid-palatal to the highest F2 (and clustering with F3). Prepalatal /i/ is sharper than mid-palatal /i/. Figure 26 illustrates British English /i/ and American English /i/.

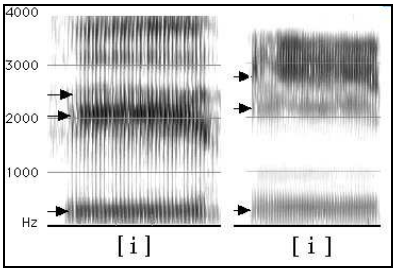


Figure 26: Spectrograms of British English /i/ and American English, proposed as typical by Ladefoged. Both vowels are described with reference to cardinal vowel 1, [i], which is the cardinal vowel closest to it.

The apparent trend of American English /i/ to be close to cardinal /i/, ([⇑F3F4)]), is however not conformed statistically (Table 5).

The sharpness of /i/ is accented by the effect of the laryngeal cavity. Simulation allows to test the contribution of anatomical detail. The laryngeal cavity, in particular, enhances the spectrum in high frequencies, and the difference is audible.

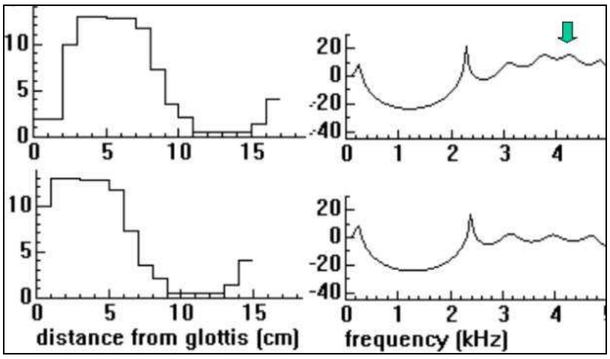
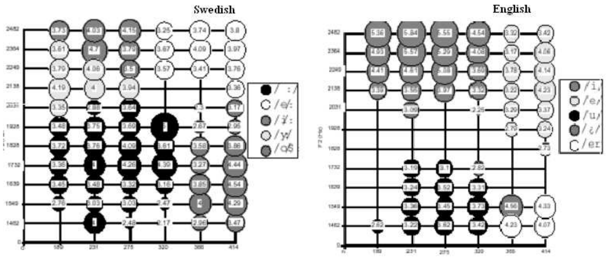


Figure 27: simulation of the acoustic effect of the presence and absence of the "laryngeal cavity" /i/ with and without laryngeal cavity. The sound on the right sounds shaper (S 27, S 28).

Swedish /i/ seems to be on the French type, but the eight languages represented in Table 5 do not include Swedish, unfortunately. As shown by Willerman and Kuhl 1996 (see Figure 28), Swedish and English listeners show different identification and goodness ratings when exposed to the same stimuli. The results suggests that the Swedish listeners prefer a higher F'2 than the English listeners: perceptual data correspond to the preceding observations.

The different /i/ sounds therefore corresponds to different sagittal profiles, to different acoustic characteristics, and to different cognitive representation in the listener's mind. AM help to give the different realisations in a different representation.


Swedish

English

Figure 28: Swedish listeners' identification and goodness ratings. (Willerman and Kuhl 1996).

5. CONCLUSIONS AND PERSPECTIVES

This paper has not deal with nasal vowels. Nasal vowels can be defined by reference to the F-pattern. The F-pattern of the vowels allows to predict the formant levels and spectrum envelopes Fant 1956). Nasal (and breathy vowels) are vowels for which the formant levels and spectrum envelopes cannot be predicted from F-pattern.

This paper is reminiscent of the Quantal Theory (Stevens 1989) and of the Dispersion-Focalization Theory (Schwartz et al 1997). The latter theory attempts to predict vowel systems based on the minimization of an energy function summing two perceptual components: global dispersion, which is based on inter-vowel distances; and local focalization, which is based on intra-vowel spectral salience related to the proximity of formants.

I have hopefully demonstrated how articulatory modeling is useful for studying the acoustics of speech in a rather enjoyable manner. It allows the students of speech to gain a rather deep knowledge of the link between sagittal profile, F-pattern and the resulting sounds. It is very helpful to illustrate vowels, semi-vowels, and consonants, compensatory and enhancing constrictions, and coarticulatory influences using the same notation and the same AM.

The same model has been used in France and abroad for a very large number of applications: teaching acoustic phonetics, interpreting MRI data concerning the singer's formant, simulating VT modifications in clinical phonetics, simulating the capabilities of the Neanderthals and sound contrasts could be produced by humans from birth to adulthood, creating codebooks for inversion programs, adapting the VT shape to the speaker (see Vaissière, 2007, for references).

AM has limits. It cannot simulate laterals, and the supraglottal frication sources are located in an ad hoc manner. In case of strong constrictions, more knowledge on the laws of aerodynamics has to be introduced (Ohala 1997).

AM demonstrate how useful it is to have articulatory data.

ACKNOWLEDGEMENTS

We would like to thank Shinji Maeda for his helpful comments, Takeki Kamiyama and Patricia Basset for their careful reading of the final version. Our mistakes are mine.

REFERENCES

Badin, P., Perrier, P., Boë, L-J., Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences, 1990, *J. Acoust. Soc. Am*. 87(3) : 1290-1300.

Bladon, A., and G. Fant., A two-formant model and the cardinal vowels. *Speech Transmission Laboratory Quarterly Progress Status Report* STL-QPRS, 1978, 1: 1-8.

Bothorel, A., Simon, P., Wioland, F., Zerling, J-P., *Cinéradiographie des voyelles et consonnes du français*, 1986, Strasbourg: Publications de l'Institut de Phonétique de Strasbourg.

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

Chiba, T. and M. Kajiyama, *The Vowel: Its Nature and Structure*, 1941,Tokyo, Tokyo-Kaiseikan Publishing Co.

Coker, C. H. and O. Fujimura, A model for specification of vocal tract area function, 1966, *J. Acoust. Soc. Am. 40*: 1271.

Delattre, P., Comparing the vocalic features of English, German, Spanish and French, *International Review of Applied Linguistics*, 1965, 2: 71-97.

Ericsdotter C., Detail in Vowel Area Functions, 2007, *Proceedings of the XVIth ICPhS*, Saarbrücken, 6-10 August 2007: 513-516.

Fant, G., On the predictability of formant levels and spectrum envelopes from formant frequencies. *For Roman Jakobson*, Mouton and Co, Gravenhage, 1956: 109-120.

Fant, G., *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*, 1960, The Hague/Paris: Mouton.

Fant, G., *Speech sounds and features*, 1973,Cambridge, Massachusetts: MIT Press.

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

Harshman, R., Ladefoged, P., Goldstein, L., Factor analysis of tongue shapes, 1977, *J. Acoust. Soc. Am*, 62: 693-707.

Heinz, J. M. and K.N. Stevens, On the relations between lateral cineradiographies, area functions, and acoustics of speech, 1965, *Proc. 4th International Congress on Acoustics*, Stuttgart: A44.

Johnson, K., *Acoustic and Auditory Phonetics*, Maldon, MA & Oxford: Blackwell, 2003.

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

Ladefoged, P. and I. Maddieson, *The Sounds of the World's Languages*. Oxford1996, U.K. & Cambridge, Massachusetts: Blackwell.

Laprie, Y. and collaborators, Projet ASPI, 2006, http://aspi.loria.fr/Save/survey-1.pdf

Lindblom, B. E. F. and J. E. Sundberg, Acoustical consequences of lip, tongue, jaw and larynx movement, 1971, *J. Acoust. Soc. Am.*, 50 : 1166-1179.

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

Maeda, S., An articulatory model of the tongue based on a statistical analysis', 1979, *J. Acoust. Soc. Am.* 65: S22.

Maeda, S., Acoustics of vowel nasalization and articulatory shifts in French nasal vowels' in M.K. Huffman and R.A. Krakow ed, *Nasals, nasalization and the velum: Phonetics and phonology*, San Diego: Academic Press, 1993, Vol 5: 174-167.

Maeda, S., Compensatory articulation in speech: analysis of x-ray data with an articulatory model, 1989, In *EUROSPEECH-1989*: 2441-2445.

Maeda, S., Phonemes as concatenable units: VCV synthesis using a vocal-tract synthesizer, in *Sound patterns of connected speech: Description, models and explanation*. In Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung 31, A. Simpson and M. Patzod eds, 1996, Kiel: 127-232.

Mermelstein, P., Articulatory model for the study of speech production', 1973, *J. Acoust. Soc. Am. 40*, 53: 1070-1082.

Ohala, J. J., Aerodynamics of phonology, 1997, Proc. 4th Seoul International Conference on Linguistics [SICOL]: 92-97.

Ouni, S. and Y. Laprie, . Studying articulatory effects through hypercube sampling of the articulatory space', 2001, International Congress on Acoustics, Rome, Italy.

Schwartz, J.L., Boë, L.J, Vallée, N., and Abry, C., The dispersion-focalization theory of vowel systems, 1997, *Journal of Phonetics* 25: 255-286.

Stevens, K. N. and A. House, Development of a quantitative description of vowel articulation, 1955, *J. Acoust. Soc. Am. 40*, 27: 484-493

Stevens, K.N., On the quantal nature of speech, 1989, *Journal of Phonetics*, 17.1/2: 3-45.

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

Straka, G., *Album phonétique*, 1965, Les Presses de l'Université de Laval, Québec.

Vaissière, J., *La phonétique*, Collection Que-Sais-Que ?, 2007, Presses Universitaires de France, Second edition.

Vaissière, J., Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of the contrast between the sounds in a language. In *Experimental Approaches to Phonology*, P.S. Beddor, M.J. Solé, and M. Ohala eds, 2007, Oxford: Oxford University Press, 54-72.

Willerman, R. and Kuhl, P., Cross-language speech perception of front rounded vowels: Swedish, English, and Spanish speakers, in H.T. Bunnell and W. Idsardi (ed.), 1996, *Proceedings of the International Conference on Spoken Language Processing ICSLP* 96.1: 442-445.

Yeou, M. (1996). *Une étude expérimentale des consonnes postérieures et pharyngalisées de l'arabe standard*, PhD Thesis, Université Paris3 Sorbonne Nouvelle, France.

Yeou, M. and S. Maeda, . Uvular and pharyngeal consonants are approximants: An acoustic modelling and study', 1995, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm: 586-589.

WEB LINKS

For the sounds:
http://www.personnels.univ-paris3.fr/users/vaissier/pub/China2008/

Jones's and Ladefoged's cardinal vowels:
http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html