Proposals for a representation of sounds based on their main acoustic-perceptual properties, Tones and Features

Jacqueline Vaissière

HAL Id: halshs-00675777

https://shs.hal.science/halshs-00675777

Submitted on 1 Mar 2012

Proposals for a representation of sounds based on their main acoustico-perceptual properties
Jacqueline Vaissière
Laboratoire de Phonétique et de Phonologie, UMR7018
Paris-France

1. Introduction

The behavior of speech sounds involves a large number of "natural", panchronic processes that take place at different points in time and space, in unrelated languages. More or less subtle synchronic, dialectal, stylistic, or allophonic variations may progressively become full-fledged sound changes. An overview of these sound changes brings out preferences for "natural" sounds, that are easier to perceive or produce, or more audible in noisy environments. These preferences are reflected synchronically in typologically preferred systems of contrasts, including cross-linguistic patterns in phonotactic restrictions. Additional evidence for these general tendencies comes from various sources, including widespread speech errors by children, hearing-impaired subjects or language learners, in ambient noise or in spectrogram reading (about panchronic phonology: Haudricourt 1969; on language universals, Greenberg [1966] 2005; on sound systems: Maddieson 1984; on the phonetics of sound changes: Ohala 1993, Blevins 2004; on universals in syllable structure: Clements and Keyser 1983, Vennemann 1988).

Phonetic sciences are concerned with explaining the panchronic behavior of speech sounds by contributing the widest possible range of plausible phonetic explanations of sound change. The constraints on sound patterns come from (i) the physiology of the speech production organs and the perception apparatus, (ii) the laws of aerodynamics and acoustics, and (iii) neurological-psychological facts. The constraints exert forces that are gradient; they interact in a nonlinear manner and may trade against each other. Thanks to progress in the understanding of the fundamentals of the speech process, and to technical advances in exploratory techniques, there now exists the potential for complex modelling which would incorporate all the different types of constraints, and which would improve gradually as new experimental evidence comes in. Earlier models that were developed in this spirit are the three-parameter model by Stevens and House (1955), and Fant's model (1960). Both approximate the vocal tract by three articulatory variables: place of articulation, degree of constriction and amount of lip rounding. Maeda's (1989) articulatory model continues in the same vein, but with more parameters. It is based on a statistical analysis of X-ray data, and allows for the study of compensatory phenomena in a more realistic way than the former three-parameter models. Other prominent models include Ohala's aerodynamic model of speech (Ohala 1997), the Task Dynamic model of inter-articulator coordination in speech (Saltzman and Munhall 1989), and vowel inventory simulations (Lindblom 1986, Schwartz et al. 1997).

This paper aims to propose a notation system based on the combination of acoustic and perceptual properties of sounds. The proposal is mainly based on Stevens's (1989), Fant's (1960) and Maeda's (1989) work and our own experience in spectrogram reading in different languages. The use of an articulatory model was instrumental in exploring the full potential of a given vocal tract in a more realistic manner than the previous models and in hearing the resulting percept (Vaissière 2008, in press). After the introduction, Section 2 provides a short summary of the articulatory, aerodynamic, acoustic and perceptual properties of sounds that have previously been put forward as explaining "natural" phonetic changes. Some relevant aspects of the acoustic theory of speech production (Fant 1960, Stevens [1998] 2000) will be recalled for the non specialists in acoustic phonetics in Section 3. Section 4 will present a method originally based on articulatory modelling for placing five vowels in the maximally stretched three-dimensional vowel space: these five vowels are proposed as references for language comparison or to transcribe detailed variations in the realisation of the phonemes (for a similar approach, see Lindblom and Sundberg 1969). The determination of the five vowels is intimately related to the Quantal Theory proposed by Stevens (1989). Since these vowels are meant to serve as references, they are compared with Daniel Jones' (1918) cardinal vowels and their rendition by Peter Ladefoged (the sounds corresponding to the cardinal vowels as uttered by Jones and

Ladefoged are available on UCLA internet web site: http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html). Section 5 describes possible uses of the same type of notation for enhancing the similarities between vowels and consonants and describing the effect of coarticulation.

2. The explanations of natural processes

This section provides a short review on the nature of proposed explanations for "natural" processes of sound change.

2.1. Articulatory-based explanations

Since Panini, in the 5th century BC, most of the sound changes in historical phonetics and natural processes have mainly been interpreted in terms of natural articulatory processes. The *distinctive features* are defined by the position of the articulators (Chomsky and Halle 1968), e.g. high, back, anterior, nasal, etc. The *International Phonetic Alphabet* (IPA) labels refer to articulation, such as height and backness of tongue body and lip position for the vowels. Finally, the basic *cardinal vowels* (Jones 1918) are also mainly defined in articulatory terms, at least according to their author.

The explanations based on articulation have proven to be very powerful. Minimum articulatory effort and economy of gestures (Lindblom 1983) lead to a decrease in the articulatory distance between the phonemes in a sequence. The overlapping of the gestures by the different organs required for the production of the successive phonemes (Hardcastle and Hewlett 1999) lead to a further reduction of the articulatory distance between the successive phonemes. The reduction of effort is not uniform and it is determined according to the prosodic status of each phoneme, mainly its position relative to word stress and to word boundaries. Sounds in word-initial and in syllable-initial position or in pre-stressed position are less likely to be lenited, i.e. reduced or suppressed. Being in a strong position leads to stronger constriction for the consonants and more opening for the vowels, resulting in a larger articulatory contrast between the onset consonant and the following vowel and to less coarticulatory phenomena. For a review on experimental studies on the effect of prosodic status on the speech organs, see Fougeron 1999.

A number of models emphasize the primacy of articulation over other aspects of speech: the listener decodes speech by identifying the underlying vocal tract gestures intended by the speaker (the Motor theory of speech production: Liberman and Mattingly 1985); the basic units of phonological contrast are articulatory gestures (Articulatory phonology: Browman and Goldstein 1992; Task-dynamic model: Saltzman and Munhall 1989).

2.2. Acoustic-based explanation

In their groundbreaking *Preliminaries to Speech Analysis*, Jakobson, Fant and Halle ([1952] 1967) viewed features as acoustic entities and defined them mainly in the acoustic domain. The criteria are the sharpness of the formant structure, the level of total intensity and the way the energy is concentrated in a central area of the spectrum, the range of frequencies where the energy is concentrated (e.g. the energy is concentrated in the low frequencies for grave sounds), the level of noise intensity, the presence of periodic low frequency excitation, the existence of additional formants and less intensity in existing formants (nasal/oral), etc. Reference to the acoustic properties of the sounds allows for the explanation of sound behaviours that are not explainable in articulatory terms. For example, the feature [grave] sheds light on the change of [x] into [f] (as in the final consonant of the English word *rough*), which are both [+grave] While there is no articulatory connection between the back of the tongue and the lips, both collaborate in lowering the energy. For Fant's later views on distinctive features proposed in the *Preliminaries*, see Fant 1973.

Stevens has developed a number of enlightening theories based on the acoustic properties of sounds. According to the Invariance theory, each distinctive feature has an invariant acoustic property. For example, the gross shape of the spectrum sampled at the consonantal release shows a distinctive shape for each place of articulation: a prominent mid-frequency spectral peak for velars, a diffuse-rising spectrum for alveolars, and a diffuse-falling spectrum for labials (Stevens and Blumstein 1978). The Quantal theory predicts that the languages of the world show a preference for regions of acoustic stability in the acoustic signal for phonemes. These regions correspond to quantal states where there are minimal acoustic consequences to the small perturbations resulting from the position of the articulators (Stevens 1989). According to Stevens' Enhancement theory, the distinctive features are often accompanied by 'redundant' features that strengthen the acoustic realization of distinctive features and contribute additional properties which help the listener perceive the distinction. For example, lip protrusion enhances distinctive [back] by lowering further the

second formant and therefore enhancing the contrast between [+back] and [-back] sounds; lip protrusion also serves to make post-alveolar [ʃ] more distinct from [s], by lowering the resonances due to the front cavity (Keyser and Stevens 2006).

2.3. Aerodynamic constraints

John Ohala has documented the aerodynamic laws underlying a number of sound changes related to voicing. Intra-oral air pressure build-up in non-sonorant consonants inhibits voicing. Non-sonorants in the world's languages tend to be uttered without vocal folds vibration (i.e. as voiceless). The phonologically voiced non-sonorants tend to devoice when they are long. Glides and high vowels have a greater tendency to devoice than comparable lower vowels because of a higher intraoral pressure. Tense vocal tract walls and back place of articulation inhibit longer voicing of the phonologically voiced stops by preventing expansion of the vocal tract volume necessary for the maintenance of a sufficient transglottal pressure; voiced velar consonants are accordingly missing more often than labial consonants in the inventory of the world languages. The voicing of a voiced fricative is gained at the expense of the energy of its frication: fricatives favor voicelessness (more than the corresponding stops), etc. (Ohala 1997, see also Passy 1890:161-162). The coupling of a side cavity, such as the nasal cavity, the tracheal cavity, the sub-lingual cavity or a lateral cavity or the back cavity in the case of fricatives is responsible for the presence of zeroes in the speech signals: nasals, fricatives, affricates, laterals, nasalized and breathy vowels share the presence of zeroes, which sheds light on some sound changes in which they pattern together. About the spontaneous nasalization of vowels in fricative context, see Ohala 1996.

2.4 Auditory-based explanation

Perception is known to play a role in preferred sound patterns. Vowel systems in the world's languages tend to maximize the perceptual space between vowels (or, in later versions of the theory, to ensure a *sufficient* space) independently of the ease or difficulty of their production (the Dispersion theory: Liljencrants and Lindblom 1972). The speaker adapts his/her way of producing speech to his/her estimation of the decoding capacities of the listener: the speaker will increase or decrease his/her articulatory effort depending on the context (the Hypo- and Hyperarticulation theory: Lindblom 1990). A large number of substitutions between sounds are explainable in simple auditory terms and interpreted as misparsing of the acoustic signal due to perceptual limitations in rapid speech (the Theory of misperception: Ohala 1981; see also Durand 1955). For example, Chang, Plauché and Ohala (2001) provide an interesting account of asymmetries in sound change based on asymmetries in perception. For a recent collection of papers on the importance of perception in shaping phonology, see Hume and Johnson 2001; for auditory based features, see Flemming 2002.

3. Acoustic theory of speech production

The acoustic theory of speech production offers a well-established method to relate a given articulatory configuration to the physical characteristics of the produced sounds. As mentioned in the introduction, the three basic components of the models simulating the relationship between articulation and acoustics are: (i) the location of the constriction formed by the tongue or the lips; (ii) the magnitude of the constriction; and (iii) the lip configuration (Stevens and House 1955, Fant 1960). More sophisticated models include the shape of the tongue, larynx height, length and shape of the constriction, more details on lip configuration (with a distinction between protrusion and rounding), or side cavities such as the nasal passage and secondary constriction(s) (Maeda 1996; Fant and Båvegård 1997).

For the sake of simplicity, the complex reality is maximally simplified in the present paper. A single nomogram (Figure 1, adapted from Fant, 1960: 82) is used to exemplify the principles underlying the effect of the speech organs' movements on the acoustics of the speech signal. A nomogram gives a rough but very useful indication of the behaviour of the first five formants when the narrowest passage in the vocal tract is moved from the lips (left in Figure 1) to the glottis (right), and when the lips are more or less rounded. It has been verified that such simplified modelling provides a useful approximation of the behavior of formant frequencies (for a comparison between Fant's nomogram, Maeda's model and the rendition of the nomogram by phoneticians, see Badin, Perrier and Boë 1990). In Figure 1, only the effect of strong rounding is represented. The third parameter, the degree of constriction, is fixed at 0.65 cm² in Figure 1. Five formants are visible under 5 kHz: the average spacing of the formants is 1 kHz for a male speaker; the spacing is wider for female (and child) speakers, because of their shorter vocal tracts. The five formants represent the so-called *F-pattern* (F1 to F5).

The five formants are not always visible on spectrograms because some of them may not be excited by a sound source or not sufficiently excited, or their intensity may be reduced because of the presence of anti-formants. For vowels, the (voice) source is at the glottis, and the whole F-pattern is excited. For obstruents, a noise source is created close to the narrowest point of constriction: this noise source excites mainly the resonances of the cavity between the constriction and the lips. One of the differences between vowels, glides and consonants is the size of the narrowest passage: The size of the constriction varies from large to zero, for vowels, glides, fricatives and stops. The F-pattern depends mainly on the tongue and lip configuration and remains approximately the same, independent of the size of the constriction and location of the source(s). The F-pattern is always calculable, for any vocal tract configuration, once the shape of the sagittal profile is known (see Fant 1960, for calculations based on X-ray data, for the complete set of Russian vowels and consonants). The place of articulation from the lips to the glottis determines an almost continuously varying aspect of segment patterns, reflecting the continuous movements of the speech articulators. Some discrete breaks in the F-pattern are due to change in type of source, or to sudden coupling of a side cavity, such as the nasal cavity, the tracheal cavity or a lateral cavity (Fant 1960; Stevens [1998] 2000).

The four circles in Figure 1 approximate some of the points where one of the first three formants reaches a local maximum (circle 1 for F3) or a local minimum (circle 1 for F4, circle 2 for F3, circles 3 and 4 for F2). The figure will be described in more detail further below.
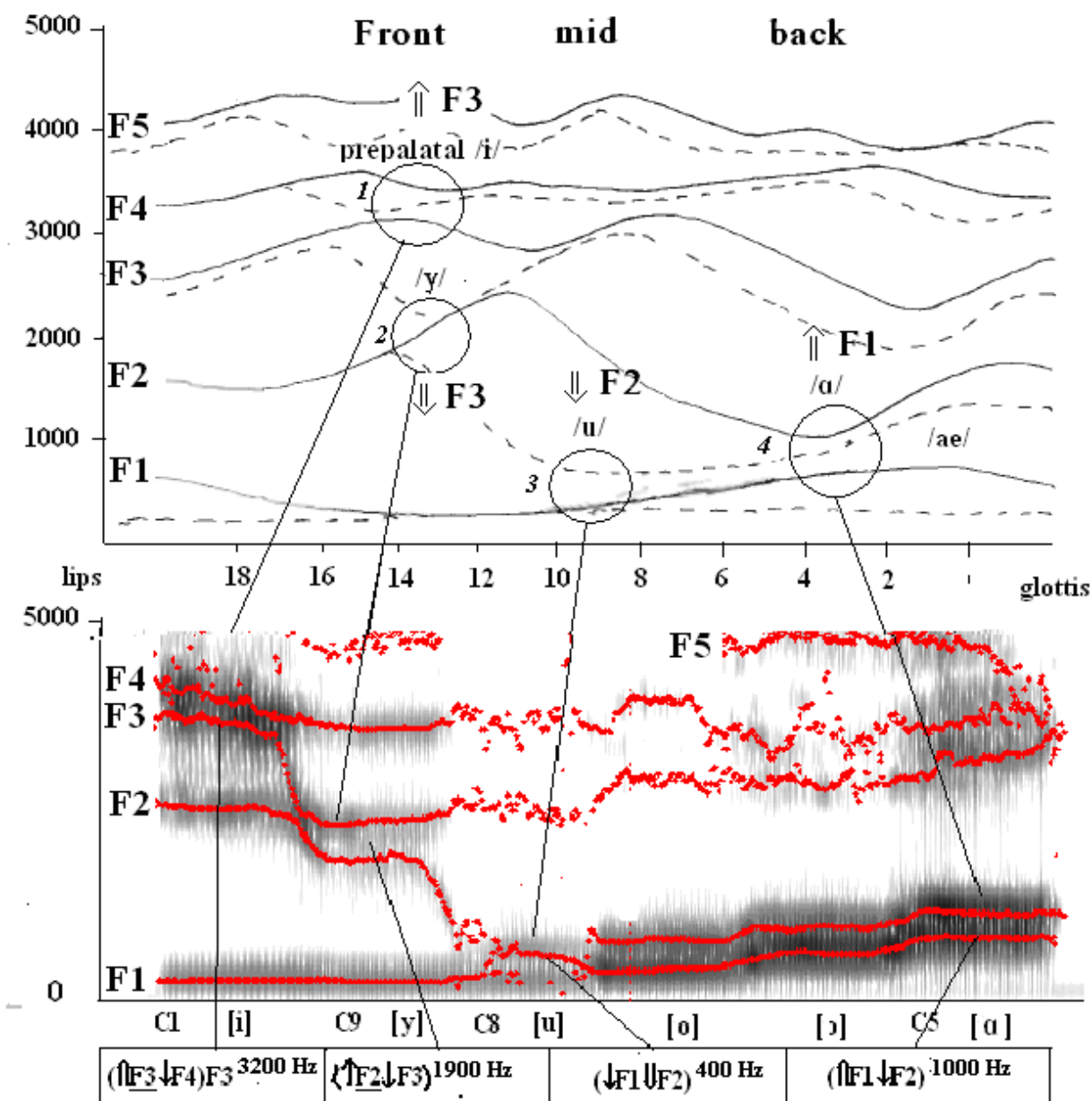
Figure 1: Top: Nomogram adapted from Fant using the three-parameter vocal tract model. The tongue constriction size is fixed at 0.65 cm^2. Plain lines and dashed lines correspond respectively to no lip rounding and to a 0.16 cm^2 lip area. Glottis at the right and lips at the left [after Fant 1960: 82]. See text for discussion of the circles. Bottom: spectrograms of the corresponding four cardinal vowels, as spoken by a female native speaker of French, plus the intermediate back vowels /o/ and /ɔ/ and their notation. See text for further discussion. The notation used below the spectrograms is discussed in Section 4.

The following remarks concern the points where one formant reaches a local maximum or a local minimum. First, when a formant is maximally high or low, it tends to converge with another formant. The points of converging formants are called focal points and correspond to quantal regions, as described by Stevens (Stevens 1989): relatively large changes in position of the constriction around the focal points will cause little change in the acoustic signal (at least for the frequency of the two formants concerned). When two formants converge, there is also an increase in their amplitude of 6 dB per halving their distance (Fant 1960:58). Furthermore, the closeness of the two formants and their increased amplitude create a sharp spectral salience in a well-defined frequency range. Two close formants are perceived as a single formant (Chistovich and Lublinskaya 1979). As is well known, vowel quality can be obtained by two-formant synthesis, F1 and F′2. F′2 is obtained by matching the quality of the vowel by one single formant, F1 being fixed. F′2 therefore gives an indication of the perceptual contribution of upper formants and represents an integrated value of F2, F3 and F4. F′2 seems to be attracted by clustered formants. For Swedish listeners, F′2 is the highest for /i/ (close to the cluster F3F4, closer to F4) and the lowest for /u/ (close to the cluster F1F2 and closer to F1). It is close to the cluster (F2F3) for /y/ and to (F1F2) for /ɑ/ (but closer to F2 than F1) (Carlson, Granström, and Fant 1970). (Swedish /y/ does not correspond to cardinal /y/, as will be discussed further below.) When there are no converging formants, F′2 is not attracted by a single formant. We may conclude then that the clustering of two formants inhibits the perceptual contribution of the upper formants and of F2 in the case of /i/ (observe the weaker amplitude of F2 than the cluster (F3F4) in Figure 2). Second, when two formants are converging, one of the converging formants may be extremely sensitive to both lip configuration and degree of constriction. They seem to be good points of departure to create new contrasts. Cardinal /i/ is characterized by converging (F3F4). The sensitivity of F3 to lip rounding for the very front constriction is employed to contrast /i/ and /y/ by lowering F3 (note that /i/ and /y/ have about the same F1 and F2, they differ by F3). The sensitivity of degree of constriction is employed to create glides, such as /j/, at every point where formants converge (see later).
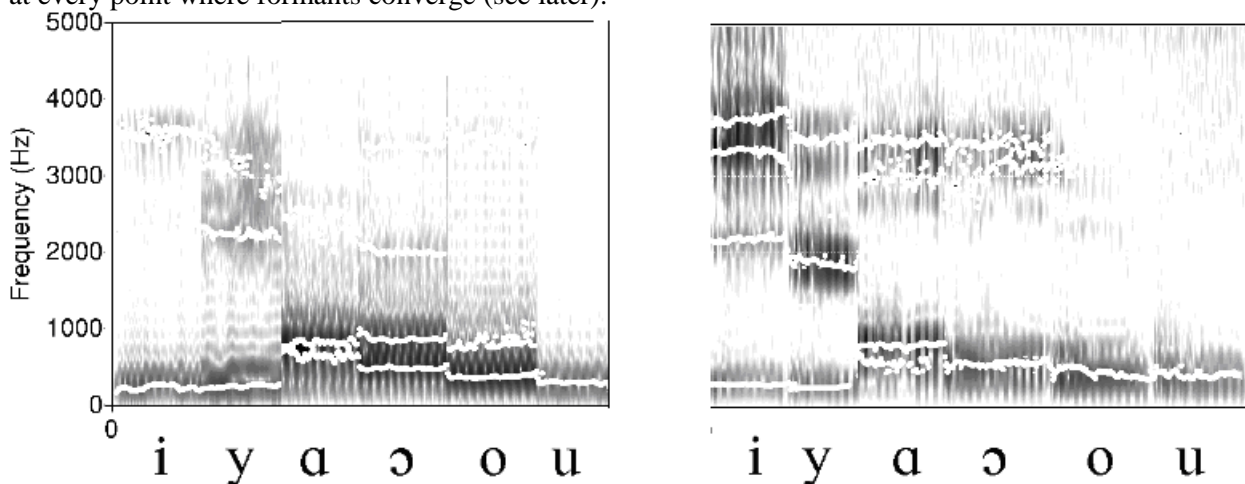

Figure 2: Spectrograms corresponding to six of the cardinal vowels as pronounced by Daniel Jones (left) and Peter Ladefoged (right) (The sounds can be found at http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html). The two converging formants often form a single peak visible in the spectrogram: (F3F4) for Jones's /i/, (F2F3) for Ladefoged's /y/, (F1F2) for Jones's /u/ and Ladefoged's /ɔ, o, u/.

Figure 2 illustrates the spectrograms of the cardinal vowels /i y ɑ ɔ o u/ as produced by Daniel Jones and imitated by Peter Ladefoged (the sounds are available at UCLA internet site). The vowels have converging formants as a common characteristic: F4 and F3 for /i/, F2 and F3 for /y/, and F1 and F2 for the back vowels. The vowels are very close to the corresponding six French vowels, except for Jones's /y/ which does not sound French and which does not have converging formants, F2 and F3, unlike

Ladefoged's /y/. /i y ɑ u/ correspond to the four encircled turning points in the nomogram represented in Figure 1.

4. Description of the cardinal vowels based on their acoustico-perceptual quality

The vowels are now described using our notation. In the expression $(\Uparrow \underline{Fn}\downarrow Fn+1)^{xHz}$, the parentheses (FnFn+1) indicate that the two formants Fn and Fn+1 converge; $\Uparrow$ (($\Downarrow$) in the expression Fn$\Uparrow$ ($\Downarrow$Fn) indicates that the formant Fn is maximally high (low) in frequency for that formant; $\Uparrow$ in the expression $\Uparrow$ ($\uparrow \underline{Fn}\downarrow Fn+1)^{xHz}$ indicates that the whole cluster formed by Fn and Fn+1 is made as high as possible (and where neither Fn nor Fn+1 represents a turning point). "x Hz" indicates the approximate location of the spectral concentration (its exact location depending on characteristics of the speaker, on particular his/her sex and vocal tract length); underlined <u>Fn</u> indicates that the formant is a resonance of the front cavity (and therefore especially sensitive to lip configuration). <u>Fn</u> is also excited by the supraglottic noise during fricatives and in the frication part of stops). The downward arrow $\downarrow$ in $\downarrow$Fn indicates that the formant is low.

Table 1: Correspondence between the focal vowels and four of the cardinal vowels

Place of constriction		Spread lips	Rounded lips
Back		$\Uparrow (\uparrow F1\downarrow F2)^{1000Hz}$ C5[ɑ]	
Mid			$(\downarrow F1\Downarrow F2)^{400Hz}$ C8[u]
Front	Prepalatal	$(\Uparrow \underline{F3}\downarrow F4)F3^{3200Hz}$ C1[i]	$(\uparrow \underline{F2}\downarrow F3)^{1900Hz}$ C9[y]

Cardinal vowel No. 1: C1[i] = prepalatal $(\Uparrow \underline{F3}F4)^{3200Hz}$

$(\Uparrow \underline{F3}F4)^{3200Hz}$ is a point where F3 and F4 converge (circle 1 in Figure 1) at about 3200 Hz (for a male speaker, higher for the female or a child speaker), and where F3 represents a local maximum. F3 is a half-wave resonance (no close end) of the front cavity, which is made as short as possible to obtain the highest possible F3 value. In that position, neither F1 (a Helmholtz resonance) nor F2 (a half-wavelength resonance of the back cavity) is independently controllable. F2 and F3 correspond to two half-wavelength resonances, the type of resonances that produces the highest resonance frequency. The $(\Uparrow \underline{F3}\downarrow F4)F3^{3200Hz}$ vowel corresponds to the cardinal /i/ produced by D. Jones and P. Ladefoged (see bottom of Figure 2), to French /i/ (Schwartz et al 1997, Vaissière, 2007), and to Swedish /i/ (Fant 1973:96, 2004:29). As Jones stated, cardinal /i/ is the sound in which the raising of the tongue is as far forward as possible and as high as possible, and the lips are spread. It does not correspond to midpalatal /i/, often observed in English (see Delattre 1965 for an X-ray study comparing French and English; Gendrot, Adda-Decker and Vaissière 2008 for comparison on statistical data concerning /i/ in a large number of languages; Willerman and Kuhl 1996 for a perception study showing differences in identification of /i/-like stimuli between English and Swedish listeners). Prepalatal /i/ has a higher F3 and lower F2 than mid-palatal /i/. A vowel with the highest F2 will be represented in our notation as $(\Uparrow \underline{F2})^{2500Hz}$. This (non cardinal) /i/-like sound could be taken as a reference, with an F2 maximally high (around 2500 Hz), corresponding to a constriction at about 11cm from the glottis, but where F2 is clustered neither with F3 nor with F4. When F3 and F4 are clustered, F1 and F2 amplitude is minimal (as pointed out above concerning Figure 2), again enhancing the acuteness of the vowel: the lower F3 in midpalatal /i/ leads to a "duller" quality than in prepalatal /i/. In languages using prepalatal /i/, /i/ does not necessarily have the highest F2 as compared to the other vowels (see an example in Swedish, Fant 1973: 96, where /e/ has higher F2 than /i/). Raising the larynx favors a high F2, by shortening the back cavity, but does not favor a low F1 (a Helmholtz resonance), because it reduces the volume of the back cavity. It is thus preferable to widen the tongue root to increase the volume of the back cavity (to lower F1), while keeping the back cavity short (for a high F2).

Cardinal C9[y]= $(\uparrow \underline{F2}\downarrow F3)^{1900Hz}$

$(\uparrow \underline{F2}\downarrow F3)^{1900Hz}$ also corresponds to the narrowest passage in the prepalatal region, where F3 is most sensitive to rounding. For the production of /y/, the lips are rounded, but moderately protruded when compared to the rounding necessary to create the cardinal vowel /u/. The lengthening of the front cavity allows for an abrupt decrease in F3 frequency. F3 becomes clustered with F2, creating a spectral peak around 1900 Hz, after F2 has become a resonance of the front cavity. Further protrusion of the lips would lower F2 and there would be no clustering with F3, resulting in a vowel quality that would not sound like /y/. Front $(\uparrow \underline{F2}\downarrow F3)^{1900Hz}$ does

not correspond to Jones's /y/, nor to Swedish /y/ (where F3 is equidistant from F2 and F4, Fant 1973: 98). However, it clearly corresponds to the rendition of cardinal vowel /y/ by Peter Ladefoged and to French /y/. Note that languages contrasting /i/ and /y/ seem to prefer a prepalatal position for both (Wood 1986), but this is not true of Swedish (as described by Fant 1973: 94-99); there is lip protrusion in French and German, but not in the Scandinavian languages (Malmberg 1974:139). As far as I have observed on spectrograms, German /y/ does not correspond to $(\uparrow F2 \downarrow F3)^{1900Hz}$ either: F2 is separated from F3. The notation for Swedish /y/ is given in Vaissière 2007.

Cardinal C8[u]: $(\downarrow \underline{F1} \Downarrow F2)^{400Hz}$

$(\downarrow F1 \Downarrow F2)^{400Hz}$ corresponds to the narrowest passage in the middle of the vocal tract (at about 6.5 cm, Fant and Båvegård 1997). The use of Maeda's model shows that C8[u] is the vowel with the lowest possible concentration of energy that a human vocal tract is capable of producing: F1 and F2 correspond to two Helmholtz resonances, the type of resonances that produces the lowest resonance frequency. It could be represented by $\Downarrow (\downarrow \underline{F1} \downarrow F2)^{400Hz}$. A strong rounding of the lips allows for a decrease in F2, which reaches its minimum. F2 of /u/ may be considered mainly as a resonance of the front cavity (Fant 1960: 211). Strictly speaking, the narrowest passage for /u/ is much fronter than for /ɑ/, in the velar region. $(\downarrow F1 \Downarrow F2)^{400Hz}$ corresponds to Jones's cardinal vowel /u/, to its rendition by Ladefoged, to the French and Swedish vowel /u/, but not to the English /u/, which usually has a higher F2. French /u/ does not have the same spectral quality as Spanish /u/, which is close to French /o/.

Cardinal C5 [ɑ]: $\Uparrow (\uparrow F1 \downarrow F2)^{1000Hz}$

$\Uparrow (\uparrow F1 \downarrow F2)^{1000Hz}$ is a sound where F1 is very high and still clustered with F2, creating a sharp peak around 1000 Hz. F1 is not exactly maximal, and F2 could be made lower. The location of the constriction corresponds to the highest clustering (F1F2). Note that a constriction at the root of the tongue leads to an even higher F1 (see Figure 1), but to a separation of F1 and F2, since it raises the frequency of F2. A constriction at the root creates an /æ/-like sound (Fant and Båvegård 1997). C5[ɑ], [a] and [æ] share a high F1, but strictly speaking, only C5[ɑ] is a quantal vowel. In the two other vowels, the first two formants are separated and do not sound like a "back" vowel.

Mid vowel [ɚ]= $(F2 \Downarrow F3)^{1500Hz}$

Another vowel, which is not considered cardinal, nonetheless represents an extreme in terms of F3 which gets as low as 1500 Hz. Three constrictions are necessary for the production of such a low F3 since there are three points along the vocal tract where the volume velocity nodes of F3 are located (Chiba and Kajiyama 1941). The production of $(F2 \Downarrow F3)^{1500Hz}$ is achieved by a constriction in the pharyngeal region, lip rounding and a bunching of the tongue toward a node corresponding to the third resonance.

Creating the back (F1F2) series /u o ʔ ɑ/

The whole back series, /u o ʔ ɑ/, is characterized by the clustering of the first two formants, and by weak intensity of the upper formants. The series can be synthesized using a single formant at equal intervals in frequency between /u/ and /ɑ/ (Delattre et al. 1952). To keep the first two formants close together, the tongue constriction has to move back from /u/ to /ɑ/ synchronously with the delabialization gesture. X-ray data show that the continuum /u o ʔ ɑ/ corresponds to a backing of the constriction and not to an increase in the area of the constriction (Wood 1979). When the constriction is in the back region of the vocal tract, jaw opening has much less effect on the formants than when it is in the front. Strictly speaking, the vocal tract is as "closed" for /ɑ/ as for /u/, but the highest point of the tongue is actually higher for /u/ than for /ɑ/.

Creating the series /C2/e/, C3/ɛ/ and C4/a/

Unlike for the vowels described above, no formants are regrouped for these vowels. The constricted part is less narrow (Fant and Båvegård 1997) than for the focal vowels. Since they are not focal, and do not correspond to turning points, these vowels are more difficult to define in acoustic terms.

5. From vowels to glides and to consonants: Coarticulation processes

5.1 From vowels with a strong constriction to glides

Table 2 illustrates the specification of the glides corresponding to the point vowels described earlier: /j/, /ɥ/, /ɹ/, /ʁ/, /w/ have F-patterns similar to /i/, /y/, /ɚ/, /ɑ/ and /u/, with clustered formants. The formant frequencies of glides are more extreme: as the constriction is made tighter in the front region of the vocal tract, F2 gets higher than for the corresponding vowel; when the constriction is closer to the glottis (as for /ʁ/ and /w/), it gets lower (on the effect of reducing a constriction, see Fant 1960: 81). When extracted from the sequences /iji/, /yɥy/, /ɚɹɚ/, /ɑʁɑ/, /uwu/, the portions corresponding to /j/, /ɥ/, /ɹ/, /ʁ/, /w/ are respectively identified as the vowels /i/, /y/, /ɚ/, /ɑ/ and /u/.

The palatal approximant /j/, palatal fricatives, palatalized liquid /ʎ/, and palatalized allophones of /l/ or /g/ share a low F1 and high F2, and (F3F4) are often clustered. Their acoustic similarity is generally not perceived. However, if the consonantal portion of /lili/ is extracted and presented to naïve listeners, it is perceived as having a vowel quality close to /i/. Similarly, if some portion of the glide /j/ is left before /i/, the stimulus is perceived as [gi] (Vaissière, 2006). As will be discussed below, all phonemes sharing a similar F-pattern will have the same effect on the surrounding phonemes.

Table 2: F-pattern of the glides in relation to the corresponding point vowels.

vowel	Corresponding glide	Type of clustering	Main effect on the surrounding phonemes
i	j	High (F3F4)	F1↘ F2↗ F3↗
y	ɥ	High (F2F3)	F1↘ F2↗ F3↘
ɚ	ɹ	Low (F2F3)	F3↘
ɑ	ʁ	High (F1F2)	F1↗ F2↘
U	w	Low (F1F2)	F1↘ F2↘

5.2. Stops and fricatives consonants

The same type of F-pattern as illustrated for vowels and glides applies to more constricted vocal configurations, such as for fricatives and stops. The parameters used to describe the vocal tract and lip configurations for vowels and glides pertain also to the description of the tongue and lip configurations for fricatives and stops. Other parameters pertaining to the length and shape of the constriction may improve the modelling (Maeda 1996; Fant and Båvegård 1997) but such details are not relevant for our present purpose. As for vowels and glides, the F-pattern for stops and fricatives contains about 5 resonances up to 5 kHz (see for example Fant 1973:100-139 for calculations of the F-pattern of stops in CV syllables).

In contrast to oral vowels, the entire F-pattern is not excited in fricatives and stops. Simplifying slightly, we can consider that only the resonances due to the cavity between the constriction and the lips are excited. The effective length of that cavity depends on lip rounding and protrusion, on the front-back position of the tongue and on the shape of the constriction. Depending on the shape of the constriction, for example, the type of resonance may be a half-wavelength type (as for /t/) or a quarter-wavelength type (as for /k/). For the same length of the cavity, half-wavelength type resonances are twice as high as quarter-wavelength type resonances. Large compensation manoeuvers are therefore possible, which are easy to understand. Figure 3 (top) represents the resonances due to the front cavity on the nomogram illustrated in Figure 1. As the constriction moves from the lips (no formants excited) to the pharyngeal region, the cavity in front of the constriction tends to become longer, and as a consequence of this, lower and lower formants are excited. The lower formants up to F5 are not excited during the labials, because there is no front cavity. The formants above F5 are excited for the anterior consonants (dental and alveolar). F3 is excited in the case of post-alveolar constriction and F2 in the case of a pharyngeal constriction; again, F3 is excited when the constriction is close to the root, e.g. for /ħ/.

Figure 3 (bottom) illustrates the spectrograms corresponding to /ki/, /ke/, /ka/, /kɑ/ and /ku/, where the constriction location of /k/ adjusts from anterior to posterior due to perceptual requirements. Lower and

lower formants are excited as the constriction moves from very front to velar. Note that the lower resonance in the case of /ku/ as compared to /kɑ/, which has a more backed constriction, is most likely due to rounding: the length of the front cavity is longer in the case of /ku/ than in the case of /kɑ/.
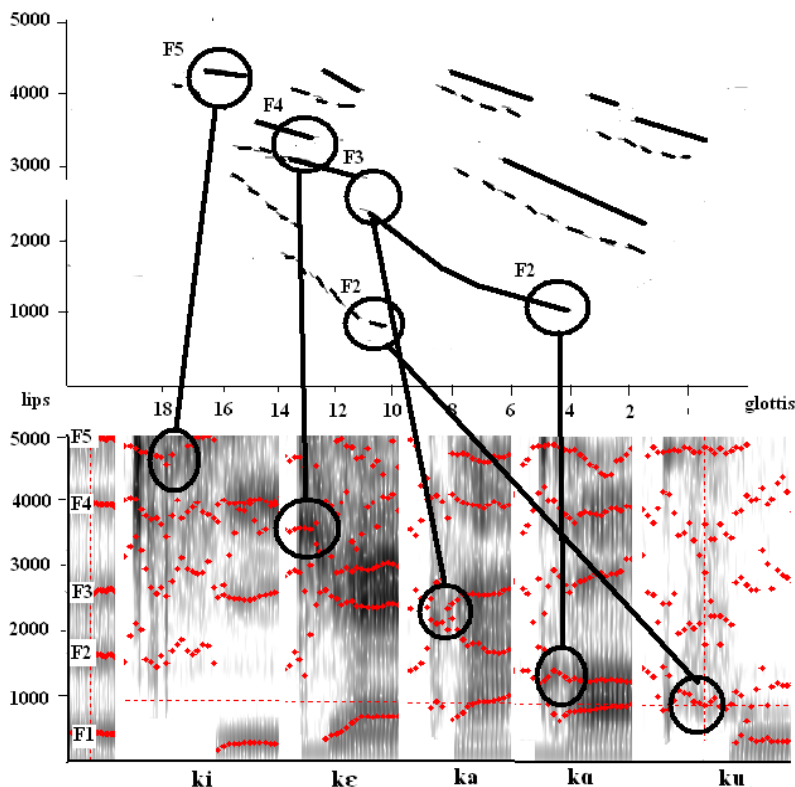

Figure 3: Top: nomogram with the representations of the resonances due to the front cavity extracted from the complete nomogram shown in Figure 1. Bottom; spectrograms corresponding to /ki/, /ke/, /ka/, /kɑ/ and /ku/. See text.

Depending on the relative size of the constriction area and the glottis opening, approximants may become fricatives and vice versa. For example, the realisation /j/ may be accompanied by noise if the constriction is made tighter: the higher formants are excited by the noise and the lower formants by the glottal source; the creation of noise is not favourable to the maintenance of voicing; it may be devoiced and become acoustically a fricative (only the formants in front of the constriction are excited). Uvular and pharyngeal fricatives, when voiced, have the characteristics of approximants (Yeou and Maeda 1995), etc. The tendency for the constriction of a consonant to be tighter or less tight than expected, and the opening of the glottis depends on the prosodic status of the phoneme. Most of the variations make excellent sense from an acoustic point of view and the gradient changes can be modelled.

5.3 Coarticulatory processes

The F-pattern for a phoneme is the result of the coarticulation of the tongue and lip configurations of the surrounding phonemes and those required for the phoneme (Öhman 1966). The direction and extent of coarticulatory overlapping are language-specific (Manuel 1990) and depend on a number of factors, such as the duration of the phoneme and the prosodic status of the phoneme. Coarticulation leads to the neutralisation of certain contrasts in given contexts, and possible sound changes.
The effect of coarticulation, and well described phenomena such as palatalization, or pharyngealization can be accounted for by similar principles. The F-pattern of consonants with a secondary articulation, such as palatalized or pharyngealized consonants, show the same influences relative to the F-pattern for the consonants with a single place of constriction. Palatalization, labialisation and retroflexion modify the effective length of the front cavity, and therefore the frication part of stops and fricatives. Pharyngealization, on the contrary, causes a minor change to the resonance patterns in front of the main constriction because the

9

secondary place of articulation is at the larynx and does not have much influence on the shape of the front cavity (Fant 1960: 219). Table 3 summarizes these effects. A fronted position of the tongue due to the phonological palatalization of consonants (such as in Russian, see Fant 1960: 220-221; Fant 1973: 69), or to surrounding front phonemes, always has the effect of lowering F1 (such as in /tat/, where /a/ tends to be perceived as /ɛ/ if extracted), and to raise the formants due to the front cavity (generally F2) (such as in /tut/, where /u/ is centralized). Similarly, a backed position of the tongue due to phonological pharyngealization (as in Arabic) or to surrounding back phonemes always has the effect of raising F1 (as in /ʁuʁ/, where /u/ tends to be perceived as /o/ if extracted) and to lower the formants due to the back cavity.

Table 3: The columns group together the phenomena which have the same raising or lowering effects on the first four formants, due to a similar tongue or lip configuration, the influences of phonetic context, and the effects of secondary constrictions.

	/i/-ness	/ɑ/-ness	/u/-ness	others
	F1↘ F2↗F3↗	F1↗ F2↘	F1↘ F2↘	F3↘ or F4↘
	closing of open phonemes and fronting of back phonemes	opening of closed phonemes and backing of front phonemes		
Tongue or lip configuration	fronted	retracted	tongue mid (+ labialisation)	retroflexed
1) Contextual influences	front consonant or vowel	back consonant or vowel	round consonant or vowel	retroflex consonant or vowel
	palatal	pharyngeal	labio-velar	retroflex
2) Secondary constriction	palatalized	pharyngealized	labio-velarized	retroflexed
Processes	palatalization	pharyngealization	labialization and velarization	F3↘ : palatal retroflexion F4↘: alveolar retroflexion

To conclude, Table 4 summarizes the commonalities between vowels and consonants, and the gestures for achieving extreme (low or high) values for the first three formants. In short, there is one main gesture for lowering F1 (fronting of the constriction), two for manipulating F2 (backing and rounding) and three for F3 (backing, rounding and retroflexion).

Table 4: Combined gestures causing one of the first three formants to be maximally high or low. The values in Hz are approximate.

Lowest possible formant		Highest possible formant	
⇓F1	Narrowest constriction at the anterior part, advanced tongue root, lowered larynx, raised velum (larger pharyngeal velum) + small lip opening (⇓F1)/i/ (≈300 Hz) All stop consonants F1 = Helmholtz resonance	⇑F1	Narrow constriction at the far back part, constricted tongue root + large lip opening (⇑F1): /ʔ/ (≈700 Hz) Uvular and pharyngeal consonant (/χ/ /ʁ/) F1= Quarter wave, back cavity resonance
⇓F2	Narrow constriction at the middle (velar) region + lip rounding and protrusion (⇓F2): /u/ (≈700 Hz) Labio-velar consonants F2 = Helmholtz resonance	⇑F2	Narrow constriction at the mid-palatal region + lip spreading + glottal constriction (⇑F2) Mid-palatal /i/ (≈2300 Hz) Mid-palatal consonants F2 = half wave, back cavity resonance
⇓F3	Narrow constriction at the back (pharyngeal) region + bunching of the tongue, retroflexion + lip rounding and lip protrusion (⇓F3): /ɚ/ (≈1500 Hz) Retroflex consonants Front cavity resonance	⇑F3	Narrow constriction at the front (apical and prepalatal) regions + lip spreading + (larynx lowering) (⇑F3): prepalatal /i/ /j/ /ç/ (≈3000 Hz) Pre-palatal consonants F3 = half wave, front resonance cavity

Conclusion

The articulatory description of the phonemes is extremely useful, but sometimes difficult to achieve in sufficient detail. Vowels defy articulatory description because they do not have a precise place of constriction. For consonants, the shape of the tongue may play a role; an acoustic contrast such as *plain* vs. *flat* may be produced by a set of articulatory manoeuvers from different parts of the vocal tract, which conspire to produce a certain percept for the consonants.

Phonetic transcription using IPA has proven useful, but in practice, transcription raises fundamental issues since the choice of IPA symbols depends in no small part on the transcribers' native language and on the instructions that they received during their training. In addition to the set of symbols for vowels and consonants, the IPA proposes diacritics to transcribe some differences between similar sounds; however, in order to describe fine differences such as that between the vowels transcribed as /u/ in French, English and Spanish, there is a clear need of well-established references for comparison. The cardinal vowels devised by Daniel Jones can be used as references but there are disturbing discrepancies between Daniel Jones's production, and the rendering of the same vowels by Peter Ladefoged. I have shown that some of the cardinal vowels have a clear acoustic definition and correspond to quantal regions as described by Ken Stevens. They are good candidates to be defined and used as reference vowels. The quantal vowels as described in Stevens's Quantal theory (with converging formants) may not correspond precisely to the most frequent vowels in the world's languages, but this does not detract from their usefulness as references in the description of vowel systems.

The specification of the phoneme in terms of distinctive features does not always reflect the acoustic-perceptual similarity between the sounds: the (only) back consonant in French is actually acoustically close to the back vowel /ɑ/; /l/ and /g/ in /i/ context share acoustic characteristics with /i/ and /j/ (such as a clustering of (F3F4), visible during /g/ when the closure is not complete): their short acoustic distance is not reflected in their definition (see examples in Vaissière 2007), but it is reflected in sound changes.

Acoustic description, based on observations of the data on modelling is a welcome addition to a description in terms of articulation and in terms of distinctive features. Acoustic description is sometimes done in the literature, but it is often incomplete. The habit of using only the two first formants to represent vowels still persists, but it is not entirely justified, at least for front vowels: the notion of F′2 has been well established and F3 plays a very large role in languages such as Swedish and French. The correlation between front/back constriction and F2, on the one hand, and high-low and F1, on the other is overestimated, whereas the role of the lips (in determining F2 of the back vowels) and the relative amplitude of the formants (which plays a role for contrasting oral and nasal vowels) is often neglected. The lack of information on F3 and F4 makes it difficult to determine the position of a vowel relative to the cardinal vowels.

The point vowels as defined here may go beyond the extreme vowels produced by the speaker in his/her native language, even when its vocalic triangle is maximally stretched. The real time visualisation of the formants when he/she tries to utter the point vowels as defined in the present article is very useful. The vowels in his/her own language may be located relative to these point vowels. The point vowels may then be used for speaker normalisation. The use of a common descriptive apparatus for all sounds brings out the continuity between vowels, glides, fricatives and stops. A solid grounding in the laws of acoustics is also a promising basis for studies of coarticulation. Clearly, these constitute challenges for future research.

References

Badin, Pierre, Pascal Perrier, and Louis-Jean Boë
1990   Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustic Society of America* 87: 1290-1300.
Blevins, Juliette
2004   *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
Boë, Louis-Jean, Pascal Perrier, Bernard Guérin, and Jean-Luc Schwartz
1989   Maximal vowel space. *First European Conference on Speech Communication and Technology (EUROSPEECH-1989)*. 2281-2284.
Browman, Catherine P., and Louis Goldstein
1992   Articulatory phonology: An overview. *Phonetica* 49: 155-180.
Carlson, R., Bjorn Granström, and Gunnar Fant
 1970   Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Status Report* (STL-QPSR 2/3*)*: 19-35.
Chang, Steve S., Madelaine C. Plauché, and John J. Ohala
2001   Markedness and constraint confusion asymmetries. In: Hume Elisabeth and Keith Johnson (eds.). *The role of speech perception in phonology*: 79-102. San Diego CA: Academic Press.
Chiba, T. and M. Kajiyama
1941   *The Vowel: Its Nature and Structure.* Tokyo: Tokyo-Kaiseikan Publishing Company.
Chistovich, Ludmilla A., and Valentina V. Lublinskaya
1979   The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research* 1: 185-195.
Chomsky, Noam, and Morris Halle
1968   *The Sound Pattern of English.* New York: Harper & Row.
Clements, Nick, and Samuel Jay Keyser
1983   *CV-phonology: A Generative Theory of the Syllable*. Cambridge. MIT Press.
Delattre, Pierre
1965   Comparing the vocalic features of English, German, Spanish and French. *International Review of Applied Linguistics* 2: 71-97.
Delattre, Pierre, Alwin Liberman, Franck Cooper and L. J. Gerstman
1952   An experimental study of the acoustic determinants of vowel color. *Word* 8: 195-210.
Durand, Marguerite
1955   Du rôle de l'auditeur dans la formation des sons du langage. *Journal de Psychologie Normale et Pathologique* 52: 347-355.
Fant, Gunnar
1970   Reprint. *Acoustic Theory of Speech Production, with Calculations Based on X-ray Studies of Russian Articulations*. The Hague: Mouton. Original edition, The Hague: Mouton, 1960.
1973   *Speech Sounds and Features*. Cambridge. MA and London, UK: MIT Press.
2004   Reprint. The relations between area functions and the acoustical signal. In: Gunnar Fant (ed.), *Speech Acoustics and Phonetics*: *Selected Writings*. Dordrecht: 29–57. Original, *Phonetica* 37: 55–86, 1980.
Fant, Gunnar, and Mats Båvegård
1997   Parametric model of VT area functions: Vowels and consonants. *Speech, Music and Hearing‚ Quarterly Progress Status Report*: Stockholm. 38(1): 1 -20.
Flemming, Edward S.
2002   *Auditory Representations in Phonology*. New York: Routledge.
Fougeron, Cécile
1999   Prosodically conditioned articulatory variations: a review. *U.C.L.A. Working Papers in Phonetics* 97: 1-74.
Gendrot, Cédric, Martine Adda-Decker, and Jacqueline Vaissière
2008   Les voyelles /i/ et /y/ du français: aspects quantiques et variations formantiques, Proceedings of the Journées d'Etude de la Parole, 2008, Avignon, France.
Greenberg, Joseph H.
2005   Reprint. *Language Universals: With Special Reference to Feature Hierarchies*. Publisher (reprint), and city??? Original edition, The Hague: Mouton, 1966.
Hardcastle, Williams J., and Nigel Hewlett
1999   *Coarticulation*, *Theory*, *Data, and Techniques*. Cambridge: Cambridge University Press.

Haudricourt, André

1969　La linguistique panchronique nécessaire à la linguistique comparée, science auxiliaire de la diachronie sociologique et ethnographique. *Ethnies* 3: 23-26.

Hume, Elisabeth, and Keith Johnson (eds.)

2001　*The role of speech perception in phonology*. San Diego CA: Academic Press. 79-101

Jakobson, Roman, Gunnar Fant, and Morris Halle

1967　*Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Cambridge, MA: MIT Press. Original edition, Cambridge, MA: MIT Press, 1952.

Jones, Daniel

1918　*An Outline of English Phonetics*. Cambridge: Cambridge University Press.

Keyser, Samuel Jay, and Stevens, Kenneth N.

2006　Enhancement and overlap in the speech chain. *Language* 82(1): 33-63.

Liberman, Alwin M., and Ignatus G. Mattingly

1985　The motor theory of speech perception revised. *Cognition*. 21: 1-36.

Lijljencrants, Johan and Bjۤorn Lindblom.

1972　Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48: 839-62.

Lindblom, Björn

1986　Phonetic universals in vowel systems. In: John Ohala and J. J. Jaeger (eds.). *Experimental Phonology,* 13-44. Orlando, FL: Academic Press..

1983　Economy of speech gestures. In: Macneilage, Peter .F. (ed.). *The Production of Speech.* New York: Springler-Verlag.

1990　Explaining phonetic variation: a sketch of the H&H theory. In: William Hardcastle, and Alain Marchal (eds.), *Speech Production and Speech Modelling* 403-439.Kluwer. Dordrecht 1990.

Lindbom, Björn, and Johan Sundberg

1969　A quantitative theory of cardinal vowels and the teaching of pronunciation. *Speech Transmission Laboratory Quarterly Progress Status Report*: Stockholm. 10: 19-25.

Maddieson, Ian

1984　*Patterns of Sounds*. Cambridge: Cambridge University Press.

Maeda, Shinji

1989　Compensatory articulation in speech: analysis of x-ray data with an articulatory model. *First European Conference on Speech Communication and Technology (EUROSPEECH-1989)*. 2441-2445.

1996　Phonemes as concatenable units: VCV synthesis using a vocal-tract synthesizer. In: Adrian Simpson and M. Patzod (eds.), *Sound Patterns of Connected Speech: Description, Models and Explanation.* Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel. 31: 127-232.

Malmberg, Bertil

1974　*Manuel de phonétique générale.* Paris : Editions A. & J. Picard.

Manuel, Sharon. Y.

1990　The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*. 88: 1286-1298.

Ohala, John

1981　The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.). *Papers from the Parasession on Language and Behavior.* Chicago: Chicago Linguistic Society, 178 - 203.

1993　The phonetics of sound change. In Charles Jones (ed.). *Historical Linguistics: Problems and Perspectives*. London: Longman, 237-278.

1996　Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*. 99: 1718–1725.

1997　Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL]*, 92-97.

Öhman, Sven .E.G.

1966　Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–168.

Passy, Paul

1890　*Etude sur les changements phonétiques et leurs caractères généraux.* Paris: Librairie Firmin-Didot.

Saltzman, Elliot L., and Kevin G. Munhall

1989　A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333-382.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry

1997　The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25: 255-286.

Stevens, Kenneth N.

1989　On the quantal nature of speech. *Journal of Phonetics* 17:3-45.

2000　*Acoustic Phonetics*. Cambridge, MA and London, UK: MIT Press. Original edition: Cambridge, MA and London, UK: MIT Press, 1998.

Stevens, Kenneth N., and Arthur House

1955   Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America* 27: 401-493.

Stevens, Kenneth, and Sheila Blumstein

1978   Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64: 1358-1368.

Vaissière, Jacqueline

(In press)   Articulatory modeling and the definition of acoustic-perceptual targets for reference vowels , *The Chinese Phonetics Journal*.

2007   Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of sounds across languages. In: Maria Josep Solé, Patricia Beddor, and Manjari Ohala. *Experimental Approaches to Phonology*: 54-71. Oxford: Oxford University Press.

2008   On acoustic salience of vowels and consonants predicted from articulatory models. *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers,* Beijing, China.

Vennemann, Theo

1988   *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin, New York, Amsterdam: Mouton de Gruyter.

Willerman, Rachel, and Patricia Kuhl

1996   Cross-language speech perception of front rounded vowels: Swedish, English, and Spanish speakers. In: Tim Bunnell and W. Idsardi (eds.). *Proceedings of the International Conference on Spoken Language Processing (ICSLP)* 96(1): 442-445.

Wood, Sydney

1986   The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *Journal of the Acoustical Society of America* 80: 391-40.

1979   A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 7:25–43

Yeou, Mohamed and Shinji Maeda

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