Control and representations in speech production

Pascal Perrier

Control and representations in speech production

Pascal Perrier
Institut de la Communication Parlée, UMR CNRS 5009, Institut National Polytechnique de Grenoble & Université Stendhal, Grenoble, France

In this paper the issue of the nature of the representations of the speech production task in the speaker's brain is addressed in a production-perception interaction framework. Since speech is produced to be perceived, it is hypothesized that its production is associated for the speaker with the generation of specific physical characteristics that are for the listeners the objects of speech perception. Hence, in the first part of the paper, four reference theories of speech perception are presented, in order to guide and to constrain the search for possible correlates of the speech production task in the physical space: the *Acoustic Invariance Theory*, the *Adaptive Variability Theory*, the *Motor Theory* and the *Direct-Realist Theory*. Possible interpretations of these theories in terms of representations of the speech production task are proposed and analyzed. In a second part, a few selected experimental studies are presented, which shed some light on this issue. In the conclusion, on the basis of the joint analysis of theoretical and experimental aspects presented in the paper, it is proposed that representations of the speech production task are multimodal, and that a hierarchy exists among the different modalities, the acoustic modality having the highest level of priority. It is also suggested that these representations are not associated with invariant characteristics, but with regions of the acoustic, orosensory and motor control spaces.

1. Introduction

The concept of representation is a key concept in language and speech research. However, it can have different meanings, according to the specific field of interest that is considered. Representations in language can be *lexical*, when they refer to the verbal characterization of the world, or *phonological*, if they aim at describing speech sequences in an invariant and segmental way. Representations in speech communication can also be *motor*, if motor control processes underlying speech production are under investigation, *spectro-temporal*, if the

characteristics of the acoustic speech signal are analyzed, or *auditory*, if the focus of the research is the perceptual processing of speech. In this paper, we will not address all the different kinds of meaning of this concept. The focus of the paper concerns the representations of the speech production task which could be elaborated in the human brain, with the purpose of setting up the motor commands and generating the movements of the vocal source and of the vocal tract articulators that will permit the production of intelligible speech sequences.

In cognitive sciences (see Jeannerod, 1994, for a challenging tutorial), the term *representation* relates to the mental imagery process underlying brain functions of human beings in their interactions with the external world surrounding them. When the brain function of interest is the production of speech, the external world includes the peripheral speech apparatus (vocal folds, jaw, tongue, velum, lips and the vocal tract as a whole) as well as the acoustic speech signal. Hence, from this perspective, the terms *representations in speech production* refer both to the mental imagery of the measurable physical characteristics associated with the articulation of intelligible speech sounds (i.e. muscle activities, articulators' positions, geometrical shape of the vocal tract, together with the spectro-temporal characteristics of the acoustic signal), and to the mental imagery of the peripheral apparatus itself (i.e. of the muscle anatomy, of the relations between muscle activations, articulatory positions and characteristics of the speech signal, of the dynamical and biomechanical articulators' properties …). In other words, studying representations in speech production means trying to find answers to the two following questions:

1. How does the human brain characterize what a speaker wants (probably unconsciously) to generate in the physical world when he/she produces speech? Articulatory positions? Spectral properties? Temporal properties?
2. How does the human brain characterize the relations between motor commands and the expected objectives of speech production in the physical world?

The focus of this paper will be limited to the question of the nature of the *representations of the speech production task* in the speaker's brain (i.e. question 1). In the first part of the paper, fundamental theoretical aspects will be presented via a summary of the four main speech perception theories. In the second part, a few selected experimental studies published in the literature will be described in order to show how these theories can be questioned and used to understand the representations of the speech production task. In these two parts of the paper, the different theoretical hypotheses and the interpretations of the experimental works will be presented as objectively as possible. In the conclusion, I will set out my own interpretation of this whole theoretical and experimental material in terms of representations.

As concerns the second question, the one of the nature of the *representation of the peripheral speech apparatus* in the speaker's brain, readers could refer to Guenther (1995) and Guenther *et al.* (1998) for theoretical studies about the use of these representations in speech motor control, to Perkell *et al.* (1997, 2000) for experimental evidences supporting the hypothesis of the existence of such representations in the brain, and to Laboissière *et al.* (1996), Perrier *et al.* (2003) and Perrier (2005) for theoretical and modeling studies of the complexity of these representations. More generally, theoretical foundations of the concept of *internal representation of the motor system* in the brain can be found in Kawato *et al.* (1987), and Jordan & Rumelhart (1992), while the controversy between Gomi & Kawato (1996) and Gribble *et al.* (1998) illustrates well the debates about the nature and the complexity of these representations.

2. What do models of speech perception tell us about the representation of the speech production task?

The complexity of the issue of the speech production task's representation arises essentially from the combination of three factors: the truly perceptual nature of the ultimate objective of speech production; the multimodality of the physical correlates of this perceptual objective; the many-to-one relations between the articulatory and the acoustic domains of speech production.

The ultimate objective of speech production is not defined in the directly accessible and measurable physical world, but in the brain of the listeners. Indeed, speech signals have no meaning by themselves. They only make sense in relation with the perception that a listener can have of them. In other words, speech production does not ultimately aim at producing specific phenomena in the physical environment of the speakers, such as movements or spectral properties of the acoustic signal, but at transmitting a code that can be interpreted by the listeners. To do so, speech production does use physical carriers, which are both articulatory movements and acoustic signals. In natural speech, articulatory movements and the acoustic signal are obviously strongly coupled, since articulatory movements are the source of the acoustics and determine its spectro-temporal characteristics. However, there is some experimental evidence suggesting that, when both modalities are available, they are both taken into account and processed in speech perception, even if they don't match with each other as it is the case in audiovisual illusions (McGurk & MacDonald, 1976). In summary, there is no way to measure the characteristics of the ultimate objective of speech production, because it is perceptual and, then, related to the listeners, but multimodal correlates of this objective (i.e. articulatory, gestural, acoustic or/and aerodynamical) can be found in the physical world.

This multimodal nature constitutes the second factor of complexity in attempts to characterize the speech production task's representations, because the characteristics in the different modalities are not simply different faces of the same object. Indeed, the many-to-one nature of the relations between the different physical domains of speech production has been shown many times (see for example Atal *et al.*, 1971). Thus, numerous muscle activations can underlie the same articulatory configuration, various articulatory positions can produce similar vocal tract shapes, and a number of vocal tract shapes can be associated with very similar characteristics of the acoustic speech signal. This is the third major source of complexity.

The search for physical correlates of the perception of speech categories (for example of phonemes) has been one of the crucial issues of speech production and speech perception research for the last 3 decades. Four major theories have been proposed in the literature; they have at the same time served as rationales for a very large amount of experimental and modeling work, and been at the core of numerous controversial debates (Perkell & Klatt, 1986; McGowan & Faber; 1996): the *Acoustic Invariance Theory* (Stevens & Blumstein, 1978; Blumstein & Stevens, 1979), the *Motor Theory* (Liberman et al., 1967; Liberman & Mattingly, 1982), the *Direct Realist Theory* (Fowler, 1986; 1991), and the *Adaptive Variability Theory* (Lindblom, 1988; 1990). In the two following subsections, the main hypotheses of these reference theories will be summarized.

2.1. Acoustic Invariance Theory and Adaptive Variability Theory: the object of speech perception is acoustic

Stevens (Stevens & Blumstein, 1978; Blumstein & Stevens, 1979) proposed that speech perception would be based on invariant properties of the acoustic signal:

> *"* [...] *there is an acoustic invariance in the speech signal corresponding to the phonetic features of natural language. That is, it is hypothesized that the speech signal is highly structured in that it contains invariant acoustic patterns for phonetic features, and these patterns remain invariant across speakers, phonetic contexts, languages.* [...] *the perceptual system is sensitive to these invariant properties. That is, it is hypothesized that the perceptual system can use these invariant patterns* [...] *to process the sounds of speech in ongoing perception"* (Blumstein, 1986, p. 178).

Typically, these acoustic invariants could be formant patterns for vowels or the spectral shape of the burst for plosives. Stevens does not deny a possible role of

articulatory information in speech perception, but not with a primary status, and only in addition to the basic process based on the processing of acoustic events:

> *"In fact the occurrence of acoustic events arising from implementation of* [phonological] *features could provide landmarks that guide the search for other features that are more directly the result of manipulation of particular articulators"* (Stevens, 1996, p. 1693).

For example, knowledge about the articulatory-acoustic relations could be helpful when the acoustic information is ambiguous or not complete:

> *"The listener must also know how to access items in the lexicon based on partial information and must also know which kinds of modifications are permitted in the sound* [...] *and which are not. Speech production clearly can play an important role in acquiring these sources of knowledge* [...] *"* (Stevens, 1996, p. 1693).

In coherence with the acoustic invariance theory, and also in strong support of it, the *quantal theory of speech* (Stevens, 1972; 1989) proposes that structure of phonological systems in the world languages would have been determined by the non-linearities of the articulatory-acoustic relation, in order to associate phonemes with the most stable acoustic patterns. Thus, the articulatory configurations would have been selected in order to minimize the acoustic variability associated with the articulatory inaccuracy existing in ongoing production of speech, and to ensure the best achievement of the acoustic features characterizing the phoneme.

The Adaptive Variability theory of Lindblom (Lindblom, 1988; 1990) defends also the primacy of acoustics over articulation for the characterization of the physical correlates of speech perception.

> *There is at present no evidence suggesting that gestures have* [any] *particular advantage over acoustic patterns.* [...] *articulatory recovery* [...] *does not seem like a compelling alternative to exploiting acoustic/auditory systematicities in efficiently precompiled form"* (Lindblom, 1996, p. 1690).

However, as indicated by its name, the Adaptive Variability theory rejects the hypothesis of the existence of any physical invariant whether in the acoustic space or in the articulatory one.

> *"Looking for invariance cannot be seen as a phonetic problem. It is not a signal analysis problem at all. The invariance of linguistic categories is ultimately to be defined only at the level of listener comprehension."* (Lindblom, 1988, p. 160).

Thus, the physical realizations of phonetic units would be fundamentally variable, depending on the phonetic context, on the speaking style, on the speaking rate, and, more generally, on the speaking condition. However, this variability would also be controlled and adapted by the speaker who evaluates what it is necessary to generate, in order ensure a good perception of the message.

> *"Intraspeaker phonemic variation is genuine and arises as a consequence of the speaker's adaptation to his judgment of the need of the situation. In the sense of the biologist's term speech is an adaptive process"* (Lindblom, 1988, p. 163).

In spite of this physical variability, a correct perception of invariant phonetic categories should be possible:

> *Any sets of intelligible pronunciations are, by definition, articulatorily, acoustically, and auditorily equivalent with respect to the goal of perceiving the given lexical item correctly, but that does not logically entail assuming articulatory, acoustic or auditory invariance in the phonetic behavior. The common point of these examples is that an invariant (non signal) end is reached by variable (signal) means* (Lindblom, 1996, p. 1685).

To achieve a correct perception, the speech perception system would not only take into account the information carried by physical speech signals, but also information about the conditions under which speech is produced. Thus, the Adaptive Variability Theory assumes

> *"that, in all instances, speech perception is the product of both signal-driven and signal independent information, that the contribution made by the signal-independent processes show short-term fluctuations, and that speakers adapt to those fluctuations. It says that* [...] *adaptive behavior is the reason for the alleged lack of invariance in the speech signal"* (Lindblom, 1990, p. 431).

Thus, according to this theory, the physical correlates of speech perception would be variable acoustic properties that would have a "*sufficient discriminative power*" (Lindblom, 1990, p. 431) to allow the identification of the different phonetic classes when contextual information is taken into account. For example, for vowels, the physical correlates could be the (F1, F2) formant patterns that would not be interpreted independently, but relatively to each other by taking into account the limits of the maximal acoustic (F1, F2) vowel space that can be produced by the speaker for the considered language and under the considered speaking conditions. The question of the mechanisms permitting the integration of contextual information in order to predict the size of the maximal vowel space, is still an unsolved question.

2.2. Motor Theory and Direct-Realist Theory: the object of speech perception is articulatory

The Motor Theory and the Direct-Realist Theory defend the idea that the objects of speech perception would be in the articulatory domain. Thus, along the lines of Stetson, they both suggest that "*Speech is rather a set of movements made audible than a set of sounds produced by movements.*" (Stetson, 1928, p.29). However, these two theories strongly differ about two important points. First, the Motor Theory does not assume the existence of a measurable articulatory invariant (i.e. of a physical invariant), while the Direct-Realist Perception Theory does. Second, and it is a consequence of the first point, the Motor Theory assumes the existence of a speech specific perceptual processing (a hypothesis classically summarized with the sentence "*Speech is special*"), while the Direct-Realist Theory is based on general human perception principles proposed by Gibson (1966). These points will be further developed below.

The Motor Theory rejects the idea that the object of speech perception would be in the acoustic domain, because

> *"*[...] *there is typically a lack of correspondence between acoustic cue and perceived phoneme, and in all cases it appears that perception mirrors articulation more closely than sound"* (Liberman *et al*., 1967, p. 453).

According to this theory, the acoustic signal would rather be for the listener

> *"a basis for finding his way back to the articulatory gestures that produced it, and thence, as it were, to the speaker's intent"* (Liberman *et al*., 1967, p. 453).

However, it should be noted that this theory does not pretend that the perception of a phoneme is associated with the existence of a measurable invariant, such as a specific set of articulatory positions or a specific vocal tract shape. Its authors rather suggest that the invariant would be at the level of the motor commands, and not in the physical external world:

> *"The invariant is found far down in the neuromotor system, at the level of the commands to the muscles"* (Liberman *et al*., 1967, p. 454).

This hypothesis was refined almost 20 years later in the "*revised*" version of the Motor Theory (Liberman & Mattingly, 1985) by introducing a clear link between the speaker's intent and physical phonetic characteristics:

> *"The objects of speech perception are the intended phonetic gestures of the speakers, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations. These gestural commands are the physical reality underlying*

> *traditional phonetic motions – for example " tongue raising", "tongue backing", "lip rounding" and "jaw raising" – that provide the basis for phonetic categories"* (Liberman & Mattingly, 1985, p. 64).

According to these authors, in speech production the existence of invariant motor commands associated with an intended phonetic gesture does not imply that any invariance exists at the articulatory or at the acoustic level. Indeed, the successive gestures necessary for the production of a sequence of phonemes are not produced purely sequentially. They partly overlap each other in time, in such a way that an invariant intended gesture will generate various movements and various acoustic signals because of two main factors: first, the nature of the preceding and following phonemes, and, second, the speaking rate determining the time overlap between successive gestures. This is the consequence of coarticulation. A strong feature of the Motor Theory consists in the fact that, thanks to the concept of overlap between invariant intended gestures, the observed physical variability of speech signals is fully compatible with the concept of phoneme related invariance. However, at the same time, the absence of congruence between the motor commands underlying an intended phonetic gesture and the associated measurable articulatory or acoustic properties, raises the question of how an intended gesture can be recovered by a listener. The solution proposed by the Motor Theory is that the perception of speech is special, and based on the use of a specialized "*phonetic module*" in the brain. This module would describe the very complex acoustic consequences of gestural overlaps in speech production, in order to infer the sequence of intended gestures from the acoustics.

> *"Incorporating a biologically based link between perception and production, this specialization prevents listeners from hearing the signal as an ordinary sound, but enables them to use the systematic, yet special, relation between signal and gesture to perceive the gesture. The relation is systematic because it results from lawful dependencies among gestures, articulator movements, vocal-tract shapes, and signal. It is special because it occurs only in speech"* (Liberman & Mattingly, 1985, p. 67).

Thanks to the "*phonetic module*",

> *"Speech perception is immediate, no cognitive translation from patterns of pitch, loudness, and timbre is required"* (Liberman & Mattingly, 1989, p. 489).

As said above, this last point is in strong disagreement with the Direct-Realist Theory of speech perception elaborated by Fowler (1982, 1986) who suggests

that the basic mechanisms of speech perception are just the same as the ones underlying visual or tactile perception:

> *"An informational medium, including reflected light, acoustic signals and the perceiver's own skin, acquires structure from an environmental event specific to certain properties of the event; because it acquires structure in this way the medium can provide information about the event properties to a sensitive perceiver. A second crucial characteristic of an informational medium is that it can convey its information to perceivers by stimulating their sense organs and imparting some of its structure to them"* (Fowler, 1986, p. 5).

In speech perception the informational medium is the acoustic signal, and the event, the source of information, is the articulating vocal tract. Fowler considers the vocal tract itself and not the motor commands that are at the origin of its shaping:

> *"* […] *studies of the activities of individual muscles or even individual articulators will not in themselves reveal the systems that constitute articulated phonetic segments"* (Fowler, 1986, p. 5).

Hence, from this perspective also the Motor Theory and the Direct-Realist Theory strongly differ. The Motor Theory suggests that listeners perceive the speaker's intent, even if this intent is hidden by the gestures associated with the surrounded phonemes, while, according to the Direct-Realist Theory, the object of perception is a set of actual characteristics of the vocal tract. What are these characteristics? Fowler does not give an answer, but she assumes that they are produced anyway, whatever the context and in spite of the observed variability of the articulatory patterns associated with the production of a phonetic segment:

> *"* [...] *from an event perspective, the primary reality of the phonetic segment is its public realization as vocal-tract activity"* (Fowler, 1986, p. 10).

This "*public*" vocal-tract activity could be perceived directly in the acoustic signal without any complex cognitive or non-cognitive processing, and it would be the direct image of the mental intention of the speakers:

> *"* [...] *the idea that speech production involves a translation from a mental domain into a physical, non-mental domain such as the vocal tract must be discarded.* [...] *we may think of the talker's intended message as it is planned, uttered, specified acoustically, and perceived as being replicated intact across different physical media from the body of the talker to that of the listener"* (Fowler, 1986, p. 10-11).

A consequence of this "*direct*" conception of the speech production-speech perception system lies in the fact that the invariant at the phonological level not only appears as a vocal tract activity, but also in the acoustic signal itself as "*specifiers or invariants*" (Fowler, 1996, p. 1731). Hence, from this perspective, the Direct-Realist Theory does agree with the Invariance Acoustic Theory, and this statement logically raises the following question: Why does Fowler defend the hypothesis of the perception of invariant vocal tract properties via the acoustic signal, rather than Stevens' hypothesis of the perception of invariants in the acoustic signal? There are two main reasons that justify this theoretical approach. First, speech perception should not obey different rules than other animal perception systems

> *"Perceptual systems have a universal function. They constitute the sole mean by which animals can know their niches.* [...] *even though it is the structure of the media (light for vision, skin for touch, air for hearing) that sense organs transduce, it is not the structure of those media that animals perceive. Rather, essentially for their survival, they perceive the components of their niche that caused the structure"* (Fowler, 1996, p. 1732).

Second, theoretical models of the different levels of human communication with language have to be as congruent and compatible with each other as possible, in order to offer a coherent theoretical framework, in which general models integrating interactions between these different levels can be developed (Fowler, 1996).

2.3. Conclusions for the representation of the speech production task in the speaker's brain

It is common sense to say that speech is produced to be perceived and that the relevance of physical characteristics of speech should only be assessed from this perspective. However, from a speech motor control perspective this common sense tells us also that the task of the speaker should be to generate in the physical world information that listeners will be able (1) to perceive and (2) to interpret in terms of phonetic categories and/or in lexical and semantic terms. Consequently, depending on the speech perception model, representations of different natures can be proposed for the speech production task in the speaker's brain.

The Acoustic Invariance Theory suggests that representations should be associated with absolute invariant temporal and/or spectral characteristics of the acoustic signal. Thus, the production of French rounded vowel /u/ could be represented as a low frequency (300Hz, 800Hz) point in the (F1, F2) space,

while producing the stop consonant /k/ could mean generating a short burst with a maximum of energy around 2.5 kHz.

The Adaptive Variability Theory suggests something more complex associating, on the one hand, some kind of acoustical characteristics, such as formant patterns or burst spectrum, that could vary within certain limits as the result of a permanent negotiation between speaker and listener, and, on the other hand, some kind of extra-linguistic information about the speaker, the speaking style or the speaking rate. From this perspective, thinking about the representations of the speech production task in the speaker's brain implies thinking about the terms of the speaker-listener negotiation and about the implementation of this negotiation in the brain.

If the Motor Theory is right, the speaker should have in mind the production of a sequence of overlapping phonetic gestures. Thus, producing a rounded vowel followed by a nasal labial stop would imply, for example, to generate with a certain time overlap, a combination of a lip movement toward protruded lips and a lowering of the larynx, for the rounded vowel, and a movement toward closed lips associated with a lowering of the velum, for the stop. There is no requirement for the speaker to actually achieve these articulatory goals. It is just necessary for them to send the appropriate commands to the motor system, the final movements depending on the gestural overlap in time.

If we follow the Direct-Realist Theory, the speaker should have the objectives to achieve a number of specific characteristics of the vocal tract. For example, producing the vowel /u/ should mean achieving a constriction in the velar region of the vocal tract together with rounded lips.

Why is it important, in terms of speech motor control, to be able to make a choice among these different hypotheses? Indeed, after all, when a French /u/ is produced, we do actually observe at the same time, rounded lips, a constriction in the velar region and a low frequency (F1, F2) pattern. Hence, why do we care whether the speaker's objective was to produce the articulatory or the acoustic characteristics? It is because of the non-linear and non bi-univocal characteristics of the relations between the articulatory and the acoustic domains of speech production.

Indeed, the non-linearity generates a warping of the relations between configurations within a space, when one moves from a space to the other. Thus, configurations that are very close in one domain, could be far from each other in another domain. This is well illustrated by the French vowels /i/ and /y/, which are quite close in the space of the first three formants, and are very well separated in the articulatory domain along the lip rounding dimension. As a consequence, if we suppose that control strategies underlying the production of speech sequences could involve a minimization of distance in a given space, the resulting optimal strategy could be different according to which space is

considered. Similarly, requirements in terms of accuracy and stability of the control could be very different according to which physical space is taken into account to measure accuracy and stability.
The non bi-univocal characteristic of the articulatory-acoustic relations could also largely influence speech motor control strategies. Indeed, a specific configuration in one domain can be associated with a number of configurations in the other domain. Thus, a given formant pattern can be produced by several combinations of jaw and tongue positioning. Similarly, a given jaw aperture can be generated by different recruitments of the jaw muscles. In summary, the number of degrees of freedom in the achievement of an objective is not the same depending on the space where the objective is defined. This implies, in particular, that different compensation strategies could be involved, and this would generate different coarticulation mechanisms…
It should be also noted that the issue of the nature of the representation of the task is an important issue not only for speech production but for motor control in general, and it has been shown to be crucial to understand the motor control strategies underlying human movements. For example, as concerns target pointing tasks with a finger, many research works have studied whether movements are controlled in the geometrical space of the finger position, in the space of the joint angles (wrist, elbow, shoulder) or in the space of the torques at the joints. The challenges of this research were well illustrated by the studies carried out by Soechting & Lacquaniti (1981), Wolpert *et al.* (1995) or Sabes & Jordan (1997).

3. Some insights into the nature of speech task representations from recent experimental studies

The four speech perception theories described in Section 2, and the corresponding hypotheses about possible speech production task's representations in the speaker's brain, are very controversial and still at the center of numerous debates. It is not the purpose of this paper to present an exhaustive review of the arguments in favor of or against each of them. Numerous, very interesting discussions were published in the literature about this topic, in particular in the book *Invariance and Variability in Speech Processes* (Perkell & Klatt, 1986), in two special issues of the *Journal of Phonetics*, the one centered on the Direct-Realist Theory for speech perception and on the Task Dynamics for speech production (*Journal of Phonetics, 14*, Vol. 1, 1986) and the one devoted to Stevens' Quantal Theory of Speech (*Journal of Phonetics, 17*, Vol. 1/2, 1989), and in the group of papers published in 1996 in the *Journal of the Acoustical Society of America* after a special session entitled *Speech Recognition and Perception from an articulatory point of view* held during spring 1994 in the

ASA meeting (*Journal of the Acoustical Society of America, 99(3),* 1680-1741). Two books respectively published by Alvin Liberman (Liberman, 1996) and by Ken Stevens (Stevens, 1998) offer numerous details about the theories defended by their authors. Two critical tutorial papers in the field of speech perception, Schwartz *et al.* (2002) and Hawkins (2004), should also be recommended. Finally, in order to close this list of publications related to the notion of representations, it is necessary to mention Sock's contribution, which totally rejects the concept of representation in speech (Sock, 2001).

To illustrate the content of the debates and the kind of studies that aim at clarifying the nature of the representations of the speech production task in the speaker's brain, a few selected experimental works will be presented in this section. First, three experimental studies supporting the hypothesis of acoustic representations will be presented. Then, an experiment suggesting the existence of strong articulatory specifications for the speech production task will be described. Finally, in a more speculative approach, we will see how the recent discovery of *mirror neurons* in monkeys' brain could offer new perspectives.

3.1. The Lip Tube experiment

In the *Lip Tube experiment*, Savariaux *et al.* (1995, 1999) perturbed the production of the French [u], pronounced in isolation, by introducing a 2 cm diameter tube between the lips of 11 subjects speakers of French. This induced a strong increase of the lip area, and limited the range of variation of jaw position. After the insertion of the tube, the subjects could train 19 times, in order to find out how to compensate for the perturbation, if they felt that compensation should be made. Then, by the twentieth repetition, they were asked to pronounce /u/ with the lip tube one more time but with the strategy that they considered to be the best among the 19 preceding training trials. Compensatory strategies were observed in the articulatory domain with sagittal cineradiography.

The production of /u/ in French is normally achieved with very rounded and protruded lips associated with a back and high tongue position generating a vocal tract constriction in the palato-velar region. However, the classical acoustic theory of vowel production (Fant, 1960) predicts that the same (F1, F2) pattern could be also produced with open lips and with a back tongue position generating a vocal tract constriction in the velo-pharyngeal region. By inserting the tube between the lips, Savariaux *et al.* wanted to test how the subjects reacted to the perturbation, with the following hypothesis in mind: if the subjects move their tongue back in order to generate a velo-pharyngeal articulation, it would support the hypothesis of an *acoustic* representation of the speech task. On the contrary, if the subjects do not change anything to their tongue position, or if they produce changes that are not compatible with an enhancement of the

(F1, F2) pattern, it would rather support an *articulatory* representation of the speech production task.
A systematic analysis of the acoustic signal in terms of pitch and (F1, F2, F3) formant patterns was performed, and perceptual tests were run to evaluate the perceptual quality of the /u/ produced under perturbed conditions. Results were as follows. First, none of the subjects could compensate for the perturbation in the first trial, and, in case of compensation, a number of trials were always necessary to achieve it. Second, only one subject actually produced a velo-pharyngeal constriction and could compensate for the perturbation in the (F1, F2) plane. Three other subjects, still keeping their tongue in the palato-velar region, moved it back sufficiently to generate an improvement of the (F1, F2) formant pattern, which, in combination with the pitch frequency, permitted the production of a well perceived /u/: the backward tongue movement permitted to limit the F2 variation, while an increase of the pitch maintained (F1-F0) sufficiently low. Third, all the remaining subjects tested a number of new articulatory strategies during the 19 trials of the training phase. Some of them provided a small improvement of the perceptual quality of their /u/ and actually moved their tongue slightly backward. Some of them did not, but none produced a forward movement of the tongue which would have led to a worse (F1, F2) pattern and to a decrease of the perceptual quality of their /u/.
These observations support largely the hypothesis that, in all cases, the compensatory maneuvers were elaborated in order to generate (F1, F2) formant patterns as close as possible to the normal patterns. Hence, they speak in favor of an acoustic nature of the representation of the speech production task in the speaker's brain. The perceptual tests also permitted to suggest that acoustic representations of vowels could be associated with regions of a space combining the formant patterns and the pitch frequency. Last, it was observed that, at the end of the training phase, 3 subjects finally had selected the original palato-velar articulation as the best one of their 19 trials, after they had stated that they could not enhance the bad (F1, F2) pattern and the bad perceptual quality of their /u/. This suggests that these canonical articulations could also be part of the representation of the speech production task.

3.2. The Dental Prosthesis experiment

Jones & Munhall (2003) investigated for six native speakers of Canadian English the contribution of the auditory feedback to the process of adapting for a geometrical perturbation of their vocal tract during the production of the fricative /s/ in the context of the word /tɑs/. The perturbation consisted in a dental prosthesis that lengthened the upper incisor teeth between 5 and 6 millimeters, without affecting the subjects' bite. As compared to the lip tube, this

perturbation has the noticeable advantage of not modifying at all the normal articulation and the normal proprioceptive and tactile information within the vocal tract. In other words, with this perturbation the natural tongue and jaw positions underlying the production of /s/ are still possible and the corresponding proprioceptive feedback is not altered.

On the contrary, in the acoustic domain this perturbation has a noticeable impact. In the production of /s/, the noise source arises from a jet of air, generated by the vocal tract constriction, and hitting the surface of the front teeth (Shadle, 1989). This source of noise excites essentially the small front cavity of the vocal tract located between the constriction and the lips. This is at the origin of a maximum of energy in the high frequency domain of the speech spectrum. According to Jones & Munhall (2003) the lengthening of the upper incisor teeth essentially induces an enlargement of the front cavity, and, thus, a lowering of the frequency of the maximum of energy in the spectrum. As a consequence, in the absence of compensation, /s/ pronounced with the dental prosthesis is expected to sound more like /ʃ/.

Each experimental session consisted of two sub-sessions, and each of these sub-sessions consisted of 15 blocks of 10 repetitions of /tɑs/ under 4 different conditions: (C1) normal condition; (C2) without the dental prosthesis in the mouth and with a masking of the auditory feedback with a white noise; (C3) with the prosthesis in the mouth and with masked auditory feedback; (C4) with the prosthesis in the mouth and with normal auditory feedback. The ordering of the 15 blocks was as follows: C1, C2, C3, C4, C3, 4 alternations (C4-C3), C2, C1. The acoustic production of /s/ was evaluated in the spectral domain by measuring the ratio between the slope of the spectral envelope below 2.5 kHz and the slope between 2.5 kHz and 8 kHz, and its perceptual quality was rated by perceptual tests carried out by 16 listeners.

Results are as follows. In the first block in condition C3, the acoustic production of /s/ was altered by the dental prosthesis, and the spectral impact conformed to the theoretical predictions: /s/ sounded more like /ʃ/. No improvement was observed during the 10 repetitions in this block. Compensation started only during the first block in condition C4, when auditory feedback became available. Afterwards, in the sequence of (C4-C3) alternations an improvement was generally noted within each block with or without auditory feedback, but the improvement was larger in the presence of auditory feedback. In addition, a learning effect was observed, since the improvement increased continuously across the 5 repetitions of the alternations (C4-C3), and since the improvement obtained at the end of sub-session 1 was maintained during sub-session 2.

The observation of the major role of auditory feedback in the compensation process supports the hypothesis of an acoustic representation of the speech production task in the speaker's brain. The results also suggest that, for

fricatives, the spectral steepness difference could be a good physical correlate of the acoustic representation. In addition, the learning effect observed within each session and across them, together with the fact that, once compensation was initiated with auditory feedback, improvement was also possible in the absence of it, suggest that speakers were immediately able to transpose requirements in the acoustic domain into articulatory terms. This suggests that the primary acoustic representation of the speech production task could be immediately associated with a secondary representation in the articulatory domain, which is more proximal for the speaker.

3.3. Velocity dependent perturbations of jaw movements

With a robotic device connected to the mandibular teeth and controlled by computer, Tremblay *et al.* (2002) delivered velocity dependent mechanical perturbations to the jaw of subjects in speech and non speech conditions. Perturbing forces were applied in the sagittal plane along an axis parallel to the occlusal plane, in the direction of jaw protrusion. Analyses of kinematic data of these perturbations induced a change of the motion path of the jaw, and thus a change of the somatosensory feedback. The larger the velocity the stronger the perturbation force applied to the jaw, and, then, the change provided to the motion path. Three different conditions were tested: production of the utterance [siat] slowly and clearly with vocalization; articulation of the same utterance without vocalization (silent speech) and still slowly and clearly; non-speech jaw movement that matched the amplitude and duration of the two speech conditions. For each condition the session started with 20 repetitions of the task without perturbation; it continued with 20 repetitions of the task with perturbation, and finally the perturbation was removed and the task was again repeated 20 times.

Results were as follows. In the first trials following the introduction of the perturbing force field, a noticeable modification of the motion path of the jaw was observed for all subjects and for the three conditions. For the two speech conditions (vocalized and silent), after training, an adaptation to the perturbation was observed: after a few trials, the motion path of the jaw became similar to the one produced without the perturbation. In addition, still for the two speech conditions, an after-effect was noticed, since, once the perturbation was removed, a few trials were again necessary for the subjects to go back to their normal jaw movements. In the non-speech condition no adaptation was observed. In order to understand why the speech conditions induced a specific behavior of the subjects, the authors assessed whether the perturbation of the jaw path did provide changes to the speech acoustics. For that, they measured and compared for vocalized speech the frequencies of the first two formants

during the transition from [i] to [a] under 4 conditions: (1) in normal condition, before the introduction of the perturbation; (2) at the beginning of the perturbed condition; (3) at the end of the perturbed condition; (4) at the end of the normal condition after the perturbation was removed. No significant differences were observed between the different conditions. Perceptual tests were also carried out and no systematic distinction could be made by the listeners between the stimuli produced in the different conditions. The last two results speak against the hypothesis of an adaptation process guided by the non-achievement of specific acoustic or perceptual goals.

In this experiment, since both speech and non speech movements used the same articulators in very similar ranges of displacement and duration, the differences observed between these two categories of movement in the impact of the perturbation cannot be attributed to any peripheral phenomenon, such as muscle mechanics or jaw dynamics. They have to be associated with differences in the motor control at the origin of the movements. According to Tremblay *et al.* (2002), they reflect differences in the specification of the goals in the articulatory domain: the time variation of the somatosensory information during the movement is part of the goal for speech production, while it is not the case for non speech movements.

These observations support the hypothesis of an articulatory representation of the speech production task in the speaker's brain.

3.4. Mirror neurons

Before concluding, it seems important to mention an extremely interesting finding that was recently made in neurophysiology for monkeys, namely the *mirror neurons* (Rizzolatti *et al.*, 2001). Indeed, this finding could become a strong support for the Motor Theory of speech perception, if similar findings could be made for human subjects in the future.

Rizzolatti and colleagues have discovered, in area F5 of the premotor cortex of macaque monkeys, neurons that are activated when a monkey grasps food with its hand, and also when the monkey does not move but observes an experimenter grasping the food with his hand. In other words, the discovery of the mirror neurons shows that the observed action leads to resonance in the internal neural circuit of the observer, which is normally activated during execution of a similar action. It should be noted that these neurons are not activated in visual perception if the observed movement does not belong to a category of movement that the monkey is able to produce, and to produce for an identified purpose. Thus, mirror neurons constitute a neural system matching action observation with action execution. It was suggested that this matching system could be at the basis of action understanding.

More recently, mirror neurons associated with the production and the observation of orofacial movements, such as lip-smacking or lip protrusion to take food, were observed in a brain area of macaque monkeys which is close to the Broca area of human brain (Ferrari *et al*, 2003). According to Rizzolatti & Craighero (2004),

> "*There are no studies in which single neurons were recorded from the putative mirror-neuron areas in humans.*[…] *There is, however, a rich amount of data proving, indirectly, that a mirror-neuron system does exist in humans. Evidence of this comes from neurophysiological and brain-imaging experiments*" (p. 174).

These experiments have shown in particular that when humans observe another human who is achieving a motor task, their motor cortex becomes active, even if no movement is actually produced. This suggests the existence of

> *"a neurophysiological mechanism* [in humans] *that creates a common* […], *non arbitrary, semantic link between communicating individuals"* (Rizzolatti & Craighero, 2004, p. 183; see also Rizzolatti & Arbib, 1998, for more details related to this hypothesis for speech).

Such a mirror-neuron system could be the basis of sensorimotor representations of speech. This is why Rizzolatti and colleagues' discovery is often seen as a potential support for the Motor Theory of speech perception.

4. Conclusions

In a theoretical approach assuming that the characteristics of the speech production and speech perception systems are the results of a strong mutual interaction, it was proposed to link the representations of the speech production task in the speaker's brain with the potential objects of speech perception. In pursuit of this aim, four reference speech perception theories were analyzed. In doing so, we ended up with three main questions:

1. Are the speech production task's representations acoustic or articulatory?
2. Do they correspond to invariant or to variable characteristics?
3. Is it necessary to actually achieve these characteristics in ongoing speech production?

To illustrate how experimental studies could help us finding (partial) answers to these questions, the results of a few recent perturbation experiments and of a neurophysiological study of animals were presented. What did we learn?

First of all, the representation of the speech production task in the speakers' brain is probably not purely acoustic and not purely articulatory. Evidence was

found in the lip tube experiment and in the dental prosthesis experiment that these representations have an acoustic component. At the same time, the velocity dependent jaw perturbations demonstrated the existence of an articulatory component in these representations. Hence, our proposal is that speech production task's representations are multimodal, i.e. both acoustic and articulatory.

However, it is important to emphasize that the acoustic and the articulatory modalities do not seem to be equally important, and that a hierarchy seems to exist among them. Indeed, some experimental works that were proposed in this paper (the lip tube and dental prosthesis experiments) show clearly that when both articulatory and acoustic characteristics of normal speech are modified by external perturbations, the speakers elaborate new strategies, in order to correct the acoustical output as their main priority. Compensation did only correct the articulatory configurations when the perturbations did not endanger the achievement of the acoustical goal (see the velocity dependent jaw perturbation experiment). We are not aware of experiments where the speakers accepted changes in the acoustical output, in order to preserve specific articulatory properties. Hence, it can be logically hypothesized that the acoustic component of the speech production task's representation is essential, primary, and that the articulatory component are of secondary importance.

The observation of learning in the alternations of the C4 and C3 blocks in the dental prosthesis experiment, and in particular the fact that improvements were also observed during the C3 blocks without auditory feedback, suggests that the articulatory component of speech production's representations could emerge from a learning guided by the acoustic representation, and that this learning could be very fast for adult speakers. The emergence of this secondary component of the task's representation as a correlate of the primary acoustic component could be, for the speakers, a way to simplify the control by projecting a distal objective (the acoustic product in the external world) into a more proximal one (in the orosensory domain). For speech this could be a particularly efficient way to simplify the control, because the transformations from the orosensory domain to the acoustic one are non-linear and non bi-univocal. In the continuity of this hypothesis, it can assumed that, within the articulatory domain, the speaker could also learn a representation in terms of motor commands associated with the orosensory goals. In the light of the role of mirror neurons in visual perception in monkeys, this projection in the motor domain could provide an efficient framework for identification and classification of phonetic units. Thus, after speech learning, the representation of the task could consist of components in the motor control domain, in the orosensory domain and in the acoustic domain, with an increasing importance from the motor control component to the acoustic one. In normal speech

production, these three levels of representation are equivalent. Planning and monitoring of speech production could thus be made in either of these domains, or in a hybrid domain based on complex, possibly phoneme-dependent combinations of the three components. However, when perturbations modify the speech conditions, so that the goals of different nature cannot be matched simultaneously, priority will be given to the achievement of the acoustic goals.

Are these representations associated with invariant characteristics? The velocity dependent jaw perturbations experiment suggests the existence of an absolute invariant in the orosensory domain, since the motion path of the jaw in its whole is quite perfectly reproduced from repetition to repetition and across experimental conditions. The dental prosthesis experiment suggests that acoustic representations could be associated with a relative invariant, which describes relations between physical characteristics of speech (in this case the steepness ratio between the slopes of the spectral envelop in the low and in the high frequency domain of the speech spectrum). The lip tube experiment suggests that speech goals would be regions of the acoustic space combining F0, F1 and F2, rather than relative or absolute invariants. This last hypothesis is more compatible with the well-known articulatory and acoustic variability of natural speech than the invariant hypothesis. Hence, in agreement with Guenther *et al.* (1998) (see also Keating, 1988, for a first proposal along these lines), our suggestion is that representations of the speech production task associate phonetic units with specific regions in the motor, orosensory and acoustics domains. The size of these regions could be variable according to the phonetic unit and, also, to the speaking style. This last hypothesis could explain intrinsically a part of the observed variability of speech signals.

Is it part of the specification of the speech task for the speaker to generate motor, orosensory and auditory characteristics that are in these regions? According to the hypothesis that we proposed about a hierarchy among the different levels of representation of the speech production task, the answer should be negative for what concerns the motor and the orosensory domains. For the acoustic domain, things are less clear. The different experiments that were presented in this paper do not permit an answer, since they did not involve changes in speaking rate, speaking style or clarity. The answer is strongly dependent on whether and to what extent the human speech perception system could be able to recover intentions in motor tasks that are not achieved. The mirror neurons, if they exist in the human brain, could participate to such an intention recovering process, since the gesture identification and classification based on these neurons seem to be related to the gesture intentionality. From this perspective the projection of the primary acoustic component of speech production task's representation into

the motor control domain would be particularly helpful. Other proposals have also been made involving target recovery in case of target undershoot based on internal models of the peripheral speech apparatus in the brain (Lœvenbruck & Perrier, 1997). In the line of the Adaptive Variability Theory we could also imagine that non-linguistic contextual information could be integrated to deal with cases where the acoustic regions defining the speech goals are not reached. This is still an unsolved question.

An important aspect of the representations of the speech production task was not treated in this paper: the representation of time. This is obviously an important drawback, since time is an essential component of speech. It carries phonemic and prosodic information that is at least as important as the configurational aspects that were considered in this paper, in the spatial and in the frequency domains. The time issue is even more complex since it addresses at the same time the cognitive issue of the representation and the perception of time in human beings, and the physical issue of the relation between time and dynamics in physical systems. Time in speech is the complex combination of both aspects. Addressing this issue, together with those of the internal representations of the peripheral speech apparatus in the brain, of the intentionality recovering, and of the potential role of mirror neurons in speech perception, could be a nice challenge for a tutorial during the next Lubmin Summerschool on *Cognitive and physical models of speech production and speech perception and of their interaction*…..

References

Atal, B.S., Chang, J.J., Mathews, M.V. & Tukey, J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America, 63,* 1535-1555.

Blumstein, S.E. (1986). On acoustic invariance in speech. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 178-193). Hillsdale N.J.: Lawrence Erlbaum.

Blumstein, S.E. & Stevens, K.N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America, 66 (4),* 1001-1017.

Fant, G. (1960). *Acoustic Theory of Speech Production*. he Hague: Mouton.

Ferrari, P. F., Gallese, V., Rizzolatti, G. & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17,*1703-1714.

Fowler, C.A (1979). "Perceptual centers in speech production and speech perception. *Perception and Psychophysics, 25,* 375-388.

Fowler, C.A. (1986). An event approach of the study of speech perception from a direct-realist perspective. *J. Phonetics, 14*, 3-28.

Fowler, C.A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America,89*, 2910-2915.

Fowler, C.A (1996). Listener do hear sounds no tongues. *Journal of the Acoustical Society of America, 99(3)*, 1730-1741.

Gibson, J.J. (1966). *The sense considered as perceptual systems*. Boston: Houghton Mifflin. (cited by Fowler, 1986)

Gomi, H. & Kawato, M. (1996). Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science, 272*,117-120.

Gribble, P.L., Ostry, D.J., Sanguineti, V. & Laboissière, R. (1998). Are complex control signals required for human arm movement? *Journal of Neurophysiology, 79*, 1409-1424.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychological Review, 102*, 594–62.

Guenther, F. H., Hampson, M. & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*, 611–633.

Hawkins, S. (2004). Puzzles and patterns in 50 years of research on speech perception. *Proceedings of the Conference "From Sound to Sense"* (CDROM) (pp. B-223 – B246). Cambridge, Massachusetts: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Jones, J.A. & Munhall, K.G. (2005). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the acoustical Society of America, 113(1)*, 532-543

Jordan, M.I. & Rumelhart, D.E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16*, 316-354.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17(2)*, 187-245.

Kawato, M., Furukawa, K. & Suzuki, R. (1987). A hierarchical neural network model for control and learning of voluntary movement. *Biological Cybernetics, 57*, 169-185.

Keating P.A. (1988). The window model of coarticulation: articulatory evidence. *UCLA Working Papers in Phonetics, 69*, 3-29. Los Angeles : University of California.

Laboissière, R., Ostry, D.J. & Feldman, A.G. (1996). Control of multi-muscle systems: human jaw and hyoid movements. *Biological Cybernetics, 74*, 373–384.

Liberman, A.M; (1996). *Speech: A special code*. Cambridge Massachusetts: MIT Press.

Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech production revised. *Cognition, 21*, 1-36. (Note: The page numbers referenced in the text correspond to the reproduction of the *Cognition* paper published in *Haskins Laboratories Status Report on Speech Research, SR-82/83*, pp. 63-93)

Liberman, A.M. & Mattingly, I.G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Liberman, A.M., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Lindblom, B. (1988). Phonetic invariance and the adaptive nature of speech. In Ben A.G. Elsendoom & H. Bouma (Eds.), *Working Models of Human Perception* (pp. 139-173). London, UK: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America, 99(3),* 1683-792.

Lœvenbruck, H. & Perrier, P. (1997). Motor control information recovering from the dynamics with the EP Hypothesis. *Proceedings ofEUROSPEECH'97* (Vol. 4, pp. 2035-2038). International Speech Communication Association.

McGowan R. S. & Faber A. (1996). Introduction to papers on speech recognition and perception from an articulatory point of view. *Journal of the Acoustical Society of America, 99(3),* 1680-1682.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices, *Nature, 264*, 746-748.

Perkell, J.S. & Klatt, D.H. (1986). *Invariance & Variability in speech processes.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Perkell, J., Matthies, M.L., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication, 22,* 227-250.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28,* 233-272.

Perrier, P. (2005). About speech motor control complexity. In J. Harrington & M. Tabain (eds), *Speech Production: Models, Phonetic Processes, and Techniques.* Psychology Press: Sydney, Australia (To appear)

Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America, 114,* 1582-1599.

Rizzolatti, G., Fogassi, L. & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review Neuroscience, 2* , 661-670

Rizzolatti, G. & Arbib, M.A. (1998). Language within our grasp. *Trends in Neuroscience, 21,* 188-194.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,*169-192.

Sabes, P.N. & Jordan, M.I. (1997). Obstacle avoidance and a perturbation sensitivity model for motor planning. *The Journal of Neurosciences, 17(18).*, 7119-7128

Savariaux, C., Perrier P. & Orliaguet, J.-P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: a study of the control space in speech production. *Journal of the Acoustical Society of America, 98,* 2428-2442.

Savariaux, C., Perrier, P., Orliaguet, J.P. & Schwartz, J.L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *Journal of the Acoustical Society of America, 106,* 381-393.

Shadle, C.H. (1989). Articulatory-acoustic relationships in fricative consonants. In W.J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 211-240). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Schwartz, J.L., Abry, C., Boë, L.J. & Cathiard, M. (2002). Phonology in a theory of perception-for-action-control. In J. Durand & B. Laks (eds.) *Phonology: from Phonetics to Cognition* (pp. 255-280). Oxford: Oxford University Press.

Sock, R. (2001). La théorie de la viabilité en production-perception de la parole. In D. Keller, J.-P. Durafour, J.-F. Bonnot & R. Sock (Eds.), *Percevoir : monde et langage* (pp. 285-316). Liège, Belgium: Mardaga (Psychologie et Sciences Humaines)

Soechting, J.F & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. *The Journal of Neurosciences, 1,* 710-720.

Stetson, R.H. (1928). Motor Phonetics: a study of speech movements in action. *Archives néerlandaises de phonétique expérimentale, 3,* 1-216. (2nd edition., 1951, North Holland: Amsterdam. 1988, by J.A.S. Kelso & K.G. Munhall, Boston).

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In Jr. E.E., David & P.B., Denes (Eds.) *Human Communication: A unified view* (pp. 51-66). New York: Mc Graw Hill.

Stevens, K.N. (1998). *Acoustic phonetics*. Cambridge, Massachusetts: MIT Press.

Stevens, K.N. & Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358-1368.

Stevens, K.N. (1989). On the quantal nature of speech. *J. Phonetics, 17,* 3-45.

Stevens, K.N. (1996). Critique: Articulatory-acoustic relations en their role in speech perception. *Journal of the Acoustical Society of America, 99(3),* 1693-1694

Tremblay, S., Shiller, D.M. & Ostry, D.J. (2003). Somatosensory basis of speech production. *Nature, 423,* 866-869.

Wolpert, D.M., Ghahramani, Z. & Jordan, M.I. (1995). Are arm trajectories planned in the kinematic or dynamic coordinates? An adaptation study. *Experimental Brain Research, 103,* 460-470.