



# Changes in speech production in response to formant perturbations: An overview of two decades of research,

Tiphaine Caudrelier, Amélie Rochet-Capellan

## ► To cite this version:

Tiphaine Caudrelier, Amélie Rochet-Capellan. Changes in speech production in response to formant perturbations: An overview of two decades of research,. Susanne Fuchs; Joanne Cleland; Amélie Rochet-Capellan. Speech production and perception: Learning and memory., 6, Peter Lang, pp.15-75, 2019, Speech Production and Perception, 978-3-631-79787-7. hal-02426327

**HAL Id: hal-02426327**

**<https://hal.science/hal-02426327>**

Submitted on 8 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tiphaine Caudrelier, Amélie Rochet-Capellan

## Changes in speech production in response to formant perturbations: An overview of two decades of research

**Abstract:** One way to investigate speech motor learning is to create artificial adaptation situations by perturbing speakers' auditory feedback in real time. Formant perturbations were introduced by Houde and Jordan (1998), providing the first evidence that speakers adapt their pronunciation to compensate for these perturbations. Twenty years later, this chapter provides an overview of the general impact of Houde and Jordan's work in speech research and beyond, as well as a more detailed review of studies that involve formant perturbations. The impact of Houde and Jordan's work appears to be cross-disciplinary. Although mainly related to speech production and perception, it has also been cited in the limb movement and even animal research, mainly as evidence of adaptive sensorimotor control. Formant perturbations research has expanded rapidly since 2006, spreading across the world and many research teams. We identified 77 experimental studies focused on formant perturbations which we then analyzed with regard to technical and theoretical issues. This analysis showed that various apparatuses and procedures were used to address important topics of speech research. A primary interest has been in feedback and feedforward control mechanisms in speech. These mechanisms were addressed in different populations, including adults and children with typical vs. atypical development, with behavioral or neurophysiological approaches, or both. Some formant perturbations studies more specifically focused on the integration of auditory and somatosensory feedback in speech production, while others explored the interaction between speech production and perception of phonemic contrasts. Some research questioned the processes and the nature of speech representations by investigating generalization of adaptation to formant perturbations. Finally, a few studies were interested in the effect of extraneous variables such as surface effects or speakers' general cognitive abilities. Altogether, these studies provide insights into speech motor control in general and into the understanding of sensorimotor interactions in particular. The field has developed recently and may still expand in the future, as it allows us to address fundamental topics in speech research such as perception-production links or abstract vs. exemplar representations. Future

research with formant perturbations may also further connect sensorimotor adaptation to linguistic and cognitive factors and in particular to working and long-term memory.

**Keywords:** perturbation, real-time auditory feedback, formants, speech units, learning

## 1. Introduction

As an “extraordinary feat of motor control” (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984, p. 812), speech production is a challenging research topic, highly influenced by movement sciences (Grimme, Fuchs, Perrier, & Schöner, 2011; Maas et al., 2008). Speech motor control indeed shares numerous features with other sensorimotor systems and in particular with limb motor control. Among these features, sensorimotor adaptability of speech is of particular interest to speech science as the basis of speech rehabilitation (Maas et al., 2008), and since it is ubiquitous in daily life. Common examples include, among others, changes in the way we speak according to our interlocutor or to the surroundings, such as speaking louder when talking with someone with a hearing impairment or in a noisy environment (Garnier, Henrich, & Dubois, 2010); or spontaneously imitating our interlocutor’s speech sounds (Pardo, 2006). Speech motor control also adapts throughout the lifespan to natural or accidental alterations of our sensory systems or vocal tract geometry, temporarily or more permanently (Jones & Munhall, 2003; Lane et al., 2007). These adaptations allow maintenance of some level of intelligibility despite vocal tract growth, hearing loss, orofacial surgery, or when wearing a dental apparatus, losing teeth, speaking while eating etc. Being essential to speech production, sensorimotor adaptation of speech is the topic of numerous studies. For the purpose of this chapter, we will focus on studies that involved specific perturbation of formants. Formants are frequencies corresponding to peaks of acoustical energy, the relative values of which characterize vowels. Research in this field, and especially Houde and Jordan’s work, was inspired by the study of visuomotor adaptation in the limb movement literature (Houde & Jordan, 1998).

Pioneering work on adaptation of different visuomotor activities appeared at the end of the 19th century (Held, 1965; Stratton, 1897). This

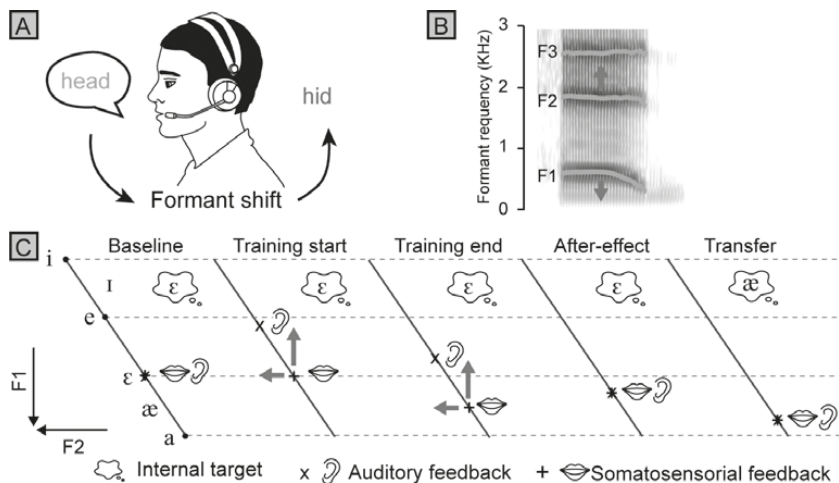
work introduced a now common approach to assessing visuomotor adaptation that consists of investigating changes in movement in response to a systematic distortion of visual feedback, such as prism adaptation. As an illustration, Stratton (1897) reported his own and extreme everyday life experience while wearing an apparatus for eight days that reversed the retinal image upside down and left to right. On the first day, “the entire scene appeared upside down”. He felt nauseous. His movements were “laborious”, “embarrassed”, “inappropriate” (p. 344), required a lot of attention and were “extremely fatiguing” (p. 344). By the start of the third day things were much better, with no sign of “nervous distress” (p. 349). At the end of the fourth day, he “preferred to keep the glasses on rather than sit blindfolded” (p. 351/352). When the apparatus was removed on day eight, it took him some time to go back to normal feelings and motions.

Later work on visuomotor adaptation focused on more specific activities, less dramatic and more local and short-term changes, with a focus on reaching movements performed with rotations of the visual field. In this context, it has been repetitively demonstrated that when movements are achieved while the visual field is shifted by a specific angle ( $\alpha$ ), participants first miss the target by the same angle  $\alpha$ . However, with repetition, they progressively learn to adapt their movements to the new feedback and reach the target accurately again. When they return to normal vision, after-effects and transfer effects are observed: participants miss the training target (*after-effects*) and/or a new target (*transfer*) by an angle more or less close to  $-\alpha$ . These effects vary as a function of the angular distance between the training and the testing targets (Krakauer, Pine, Ghilardi, & Ghez, 2000; Shadmehr & Mussa-Ivaldi, 1994). Sensorimotor adaptation has been attributed early on to feedforward control (i.e. predictive control based on learnt sensorimotor mappings) in contrast to forward closed-loop control (i.e. online processing of sensory inputs), visible in correction to unexpected perturbations (Golfinopoulos, Tourville, & Guenther, 2010; Houde & Chang, 2015). These notions are defined later in this chapter.

Twenty years ago, Houde and Jordan (1998) introduced an analogous procedure of visuomotor rotation adaptation to question feedforward control in speech, which used real-time alterations of formant frequencies in vowels. By altering the frequencies of the first and/or second formants (F1 and F2 respectively) it is possible to make a vowel sound like another

vowel. For example, by decreasing F1 and increasing F2, the vowel / $\epsilon$ / would sound closer to the vowel / $u$ /, as illustrated in Figure 1. This alteration *displaces* the auditory feedback, in the same way as prism vision displaces the visual position of the target. For example, the speaker says “head”, speaking into a microphone and wearing headphones (Figure 1.A). The signal is processed in real time so that F1 and F2 formants are moved towards “hid” (Figure 1.B), and played-back into the headphones. The consequence for the speaker is a discrepancy between the auditory target expected from the planned movements (“head”) and the auditory target they actually got (~“hid”). In other words, similar to visuomotor adaptation, the speaker first *misses* the auditory target (Figure 1.C, “Training start”). With practice – repetition of shifted utterance(s) with the same perturbation – the speaker adapts to the perturbation (Figure 1.C, “Training end”): To *reach* the auditory target “head” again in the presence of the perturbation, they produce formants in the opposite direction to the perturbation. In our example, this corresponds to the production of an utterance closer to “had”. When the feedback is returned to normal or masked with a noise, for the same vs. different utterance(s) than the training one(s), after-effects vs. transfer effects are observed (Figure 1.C, column “After-effect” and “Transfer”). This suggests that the compensation is not only an online feedback control change but also affects auditory-motor mappings supporting feedforward control, in a more or less utterance or segment-specific way. The procedure was later adapted to address feedback control by investigating online compensation to unexpected perturbations (Purcell & Munhall, 2006b).

Adaptation to formant perturbations has been investigated per se, or used as a paradigm to address more general issues in speech science. The current chapter reviews research in formant perturbations by analyzing Houde and Jordan’s seminal study (Houde & Jordan, 1998, 2002) and the scientific literature that has referred to it. Using this approach (detailed in the first section of the chapter) we can see the cross-disciplinary impact of Houde and Jordan’s work and in particular, identify the main topics of the scientific literature that have cited this work (reported in the second part of the chapter). Among the collected papers, only a subsection corresponded to empirical studies involving formant perturbations. Based on the analysis of these studies, including review of their reference lists, the latter parts



**Figure 1:** The auditory prism adaptation. (A) The speaker speaks into a microphone; his feedback is altered such as when he produces “head” he is hearing a signal closer to “hid”; (B) To do so, F1 and F2 are changed in real time; (C) Before the introduction of the perturbation (Baseline) the auditory feedback is consistent with the target. The first exposure to the perturbation (Training start) induces a discrepancy (or an error) between the auditory feedback and the planned target. With repetitive exposure to the perturbation, the talker changes his production to compensate for the perturbation (Training end). When the perturbation is removed after-effects and/or transfer effects are observed.

of the chapter provide: (1) a description of the main apparatuses and paradigms used in formant perturbations studies; (2) an overview of the research topics addressed using these perturbations and the main reported results; and (3) some perspectives for future research.

## 2. Paper collection and analysis

As we were interested in the impact of Houde and Jordan’s work and also wanted to provide an analytical review of formant perturbations studies, we first analyzed the published work that referred to Houde and Jordan (1998 and/or 2002) from 1999 to 2018 (last update on July 6th 2018). This was performed using the “Cited by” function in Google Scholar. We choose this approach rather than keyword research, as we wanted to collect various sorts of publications, and because it appeared to be the

**Table 1:** Number of references in each category of the first level of selection (see text for details)

Formant shift		No formant shift		Not in English	Error ref.	Total
Rejected	Kept	Rejected	Kept			
26	72 (+ 2, Houde & Jordan 1998 and 2002)	140	287	35	22	584

most systematic way to collect publications in the field. To compensate for potential errors and omissions by Google Scholar, the results were then analyzed very closely.

An analysis by year of Google Scholar output resulted in a total of 584 references (including the two papers by Houde and Jordan, see Table 1). As a first step, we excluded documents that were not written in English or that corresponded to reference errors (57 in total, see Table 1). Among the 527 remaining references, we distinguished between those without vs. with an empirical study that included formant perturbations. In the former category (n=427, without formant perturbation), we kept only journal papers for a thematic analysis of Houde and Jordan’s broad impact (n=287). In the latter category (n=100, with formant perturbations), we first kept all the documents except PhD or Master theses, posters or abstracts to conferences (74 references kept, 26 rejected). Note that there were 11 PhD theses; most of them were associated with journal publications. For consistency in criteria, we did not include Frank (2011)’s PhD thesis, even though it is often cited by studies investigating linguistic effects on formants adaptation. Its results were never published in peer-reviewed papers.

Three more papers were added that included formant perturbations. One paper that did not cite Houde and Jordan was found in the reference list of the selected papers (Niziolek & Guenther, 2013); and two papers in course of publication at the time of writing that we were aware of (Caudrelier, Perrier, Schwartz, & Rochet-Capellan, 2018; Klein, Brunner, & Hoole, in this book). The general characteristics of the documents including formants perturbations are described in Table 2. Technical

**Table 2:** Number of papers considered for the analysis of formant perturbations according to source and type. Houde & Jordan (1998, 2002) are included.

	Journal papers	Proceedings papers	Reports/ chapters	Total
Google Scholar	55	17	2	74
Other sources	1	1	1	3

papers as well as papers investigating compensation to unexpected formant perturbations were included.

The full list of analyzed papers related to formant perturbation is available in Table 4, with their main related research topic indicated. As the paper collection is based mainly on the “cited by” function of Google Scholar some papers may be missing despite our careful attention. However, we believe our analysis provides an accurate picture of the field at the time it was run.

### 3. Overall impact of Houde and Jordan’s seminal work

The overall impact of Houde and Jordan (1998, 2002) is illustrated in Figure 2. We distinguished seven broad categories of research: (1) formant perturbations studies ( $n=77$ ); (2) studies that investigated speech compensation and/or adaptation to other auditory perturbations or equivalent situations ( $n=91$ ) or (3) to an alteration of the vocal tract ( $n=16$ ); (4) empirical or theoretical papers on speech production ( $n=61$ ) or (5) on speech perception ( $n=46$ ); (6) studies involving non-speech actions ( $n=25$ ); and (7) experimental or theoretical papers involving animals ( $n=43$ ). Five papers were not considered, as they were difficult to classify in these categories. We first analyzed the journal papers that did not empirically test formant perturbations. As described above, this involved 286 articles. Broad research topics were identified mainly from abstract reading. A subset of papers was selected and read in more detail to illustrate the different topics. The articles on formant perturbations will be reviewed in detail in the next sections. We will now briefly overview the research topics in the six other categories. References in the following section are illustrative.



### **3.1. Compensation/adaptation of speech production to various auditory perturbations**

Speech compensation and adaptation were investigated prior to the development of formant perturbation studies and used various methods. These methods continued to be used in some of the later work that cited Houde and Jordan. About half of the papers in this first category investigated speech modifications in reaction to either an unexpected or a predictable modification of F0 in different populations and conditions. A number of papers in this topic were published by Jones et al. (Jones & Munhall, 2000); Larson et al. (Burnett & Larson, 2002); or Hanjun et al. (Li et al., 2016). The other half of the studies investigated speech modifications in reaction to other types of auditory perturbations such as delayed auditory feedback (Chon, Kraft, Zhang, Loucks, & Ambrose, 2013); changes in intensity or noise level (Maas, Mailend, & Guenther, 2015); hearing loss (Palethorpe, Watson, & Barker, 2003); real or simulated use of cochlear implants (Casserly, 2015; Lane et al., 2007); or replacement of the auditory feedback by a stranger's voice (Hubl et al., 2014). Other work modified consonant features such as frication (Shiller, Sato, Gracco, & Baum, 2009) or voicing (Mitsuya, MacDonald, & Munhall, 2014). Self-regulation in adaptation to formant perturbations was also linked with interpersonal auditory-motor regularizations in speech such as phonetic convergence (Pardo, 2006).

### **3.2. Compensation/adaptation of speech production to perturbations of the vocal tract dynamics or geometry**

Research on compensation and adaptation to perturbations affecting the somatosensory feedback is another field closely connected to adaptation to formant perturbations. Houde and Jordan's work was thus cited by studies involving an alteration of the vocal tract geometry or dynamics. This includes dental prostheses (Jones & Munhall, 2003); lip tubes in children and adults (Ménard, Perrier, & Aubin, 2016); false palates (Thibeault, Ménard, Baum, Richard, & McFarland, 2011); mechanical forces applied to the jaw with a robot (Tremblay, Shiller, & Ostry, 2003); or more permanent changes such as those induced by oropharyngeal cancer treatments (de Bruijn et al. 2012).

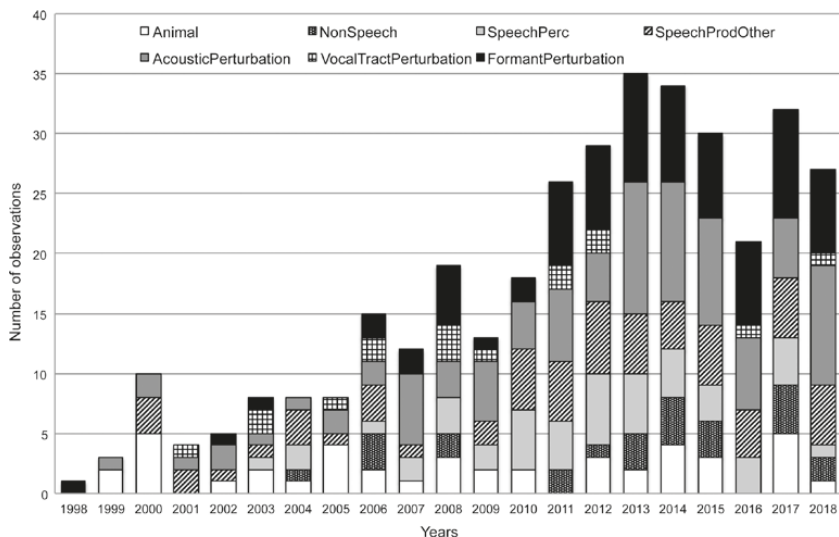


Figure 2: Overall impact: number of analyzed papers by year and categories.

### 3.3. Empirical or theoretical papers on speech production

Houde and Jordan's work is cited by empirical and theoretical research on speech production. For example, adaptation to formant perturbations is mentioned by studies providing further evidence of the role of auditory feedback in speech motor control, such as work linking auditory acuity to the production of speech contrasts (Perkell et al., 2004); auditory perceptual learning with improvement in production (Shiller, Rvachew, & Brosseau-Lapr , 2010); comparing overt and covert speech (Brumberg et al., 2016) or analyzing the neurophysiological activities of the auditory cortex during speech production (Curio, Neuloh, Numminen, Jousm ki, & Hari, 2000). Adaptation to formant perturbations provides support for neurocomputational models of speech production such as the Directions Into Velocity of Articulators model (DIVA, Golfnopoulos et al., 2010) or the State Feedback Control model (SFC, Houde & Chang, 2015), both models assuming a feedback and a feedforward control mechanism. Further information about these control mechanisms will be provided in the section describing formant perturbation studies related to this topic.

### **3.4. Empirical or theoretical papers on speech perception**

Adaptation to formant perturbations is also taken as evidence of sensorimotor integration in speech. As such, it is relevant for papers probing or discussing the role of the motor system in speech perception (Sato, Troille, Ménard, Cathiard, & Gracco, 2013) or in theoretical papers related to the dual-stream model of language processing. Basically, this model proposes a cortical ventral stream that maps speech sounds to concepts, and a dorsal stream for auditory-motor mapping. Adaptation to formant perturbations is then cited as an evidence that a dorsal auditory-motor integration path is still functional in adulthood (Hickok & Poeppel, 2004).

### **3.5. Non-speech movement studies**

Various non-speech studies cited Houde and Jordan's work to illustrate sensorimotor adaptation in humans. These studies focused on activities involving auditory feedback such as piano playing (Pfordresher & Palmer, 2006); or the learning of artificial auditory-arm movement maps (van Vugt & Ostry, 2018). Some papers were also interested in other kinds of sensorimotor adaptations such as swallowing (Wong, Domangue, Fels, & Ludlow, 2017), or visuomotor adaptation of limb movements (Wei et al., 2014). Note that as formant perturbations studies were inspired by visuomotor adaptation, they often referred to limb movement literature. The converse seems not necessarily true as our research suggests that few works on limb adaptation have cited Houde and Jordan's work. This result should be taken cautiously as limb movement research could cite other studies using formant perturbations to illustrate the adaptability of speech motor control, and we only collected papers that reference Houde and Jordan using "cited by" functionality of Google Scholar.

### **3.6. Animal studies**

Finally, animal studies have early, and regularly, cited Houde and Jordan's work (Figure 2), with a main focus on the role of auditory feedback in action control. Over half of these papers were dedicated to birdsong and published by Brainard et al. and/or Doupe et al. and/or Sober et al. Many of these papers include studies of birdsong production or learning using auditory perturbations with behavioral and/or neurophysiologic recordings, as

well as interspecies comparative reviews about the processing of auditory feedback of self-produced sounds (Brainard & Doupe, 2000; Doupe & Kuhl, 1999; Sober & Brainard, 2009). Analogous works were done in bats (Smotherman, Zhang, & Metzner, 2003) and primates (Eliades & Miller, 2017).

To summarize, this non-exhaustive analysis of the overall impact of Houde and Jordan's seminal work suggests that it is (as expected) cited by papers investigating speech compensation and adaptation to other types of sensory perturbations. Most of the scientific questions in this first set of papers overlap with the research topics we will review based on the more detailed analysis of formant perturbations studies in the related section of this chapter. In a broad context, adaptation to formant perturbations is often interpreted as evidence for sensorimotor integration and sensorimotor plasticity in speech production and perception. It is cited to illustrate auditory feedback and feedforward control mechanisms in speech production, as explained below, and taken as an example of such mechanisms (and their plasticity) in studies investigating animal vocalizations, singing, music playing, but also inter-personal convergence or coordination of movements.

Note that more research topics related to formant perturbation studies may be found by including "2nd order" connections to Houde and Jordan's work (i.e. references that cite any of the studies on formant perturbations).

## 4. Methods in formant perturbation studies

In this section, we provide an overview of the apparatuses used to apply real-time formant perturbation and a description of the main procedures identified in the collected papers.

### 4.1. Real-time formant perturbation

The systems used to shift formants in the collected papers are summarized in Table 3. Paper details can be found in Table 4. With regards to formant perturbation, it is important to emphasize that in order to preserve the best quality of self-perception, the real-time modification of formants in speakers' auditory feedback should meet some requirements, specifically:

- (1) The signal should be processed and played back fast enough for the speaker not to perceive any delay (less than 30ms, see Yates, 1963). Specific digital signal processing boards (DSP), including systems from the music industry were used, especially in earlier work. Nowadays, this can be achieved at a software level, on a PC with appropriate sound card and software to analyze and change formants. For the same code, the achieved delay can vary depending on the operating system and hardware.
- (2) The parameters of the signal processor should be adapted to the speaker and/or to the vowel. This parameterization improves the formant detection and the reliability of the perturbation.
- (3) Perception of unperturbed feedback (bone conduction and air conduction outside the headphones) should be reduced as much as possible. Different approaches were used to achieve this aim, such as:
  - Using whispered speech (Houde & Jordan, 1998, 2002) although subsequent studies were run with normal speech;
  - Using closed headphones or insert earphones to reduce the perception of the air-conducted signal. The occlusion effect of the headphones on adaptation was recently investigated with no significant difference in the magnitude of F1 adaptation between the use of the closed Sennheiser “HD 265” and the insert Etymotic Research ER2 (Mitsuya & Purcell, 2016);
  - Increasing the level of the feedback in the headphones, up to 87dB SPL (Villacorta et al., 2007);
  - And/or using a masking noise mixed with the played back signal to mask bone-conducted speech.
- (4) The shifted vowel should have clearly distinguishable F1 and/or F2 values, and the shift should be consistent with these values. For this reason, the vowel /*ɛ*/ is chosen in most of the studies as shifting more extreme front or back vowels could be limited by overlap in F1–F2 or F0–F1 frequencies (Mitsuya, MacDonald, Munhall, & Purcell, 2015), and this vowel allows upward and downward perturbations.

Different research groups have developed their own formant perturbation systems (Table 3) with four main categories: (1) The two systems developed by Houde described with more details in Houde’s PhD (Houde, 1997) for whispered speech (1.a), and then in Katseff, Houde, & Johnson

(2012) for voiced speech (1.b); (2) The system developed and used by Munhall, Purcell and collaborators that used a specific hardware; (3) The system used by Perkell and Guenther's teams that first included specific hardware (Villacorta et al., 2007) and was then adapted as a free software for Matlab. It supports various auditory perturbations, including changes in F1 and/or F2, but also more complex ones such as formant trajectory perturbations (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Tourville, Cai, & Guenther, 2013). The last version is called "Audapter" and can be download on github.com ([https://github.com/shanqing-cai/audapter\\_matlab](https://github.com/shanqing-cai/audapter_matlab), this link was retrieved July, 6, 2018); (4) The last system was developed in parallel by three teams: Max et al., Ostry et al., and Shiller et al. It uses a device from the music industry (VoiceOne, TC Helicon) that by default allows shifting of all the formants while preserving F0. This system was used as a way to alter all formants in the same direction (Max & Maffett, 2015) or, with supplementary signal processing steps, including filtering and mixing, as a way to perturb F1 only (Rochet-Capellan & Ostry, 2011). A few papers were dedicated to the presentation and first

**Table 3:** Main signal processing systems used in the literature to perturb formants in real time (references indicate the publication describing the system) and number of papers using the system.

	System 1	System 2	System 3	System 4	
References	Houde (1997); Katseff et al. (2012)	<i>Purcell &amp; Munhall (2006ab)</i>	<i>Villacorta et al. (2007); Cai et al. (2008); Tourville &amp; al. (2013)</i>	Feng et al. (2011); Rochet-Capellan & Ostry (2011); Shum et al. (2011)	Others
Signal processing	1.a. Whispered speech: Analysis-synthesis process, DSP- 96 board, Ariel, Inc. 1.b.Voiced speech: "Feedback Alteration Device" – Sinewave synthesis	National Instruments PXI-8176 embedded controller	Texas Instruments C6701 Evaluation Module DSP board <b>then</b> C-extension Mex for Matlab, opened access – Audapter	Electronic speech processor from music industry VoiceOne; TC Helicon + filters	Other software or hardware solutions –
Number of papers	10	23	2 <b>then</b> 20	19	3

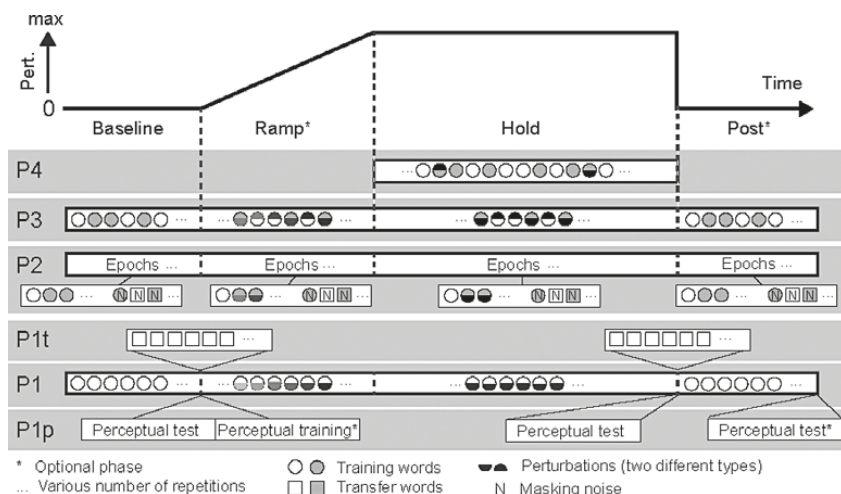
evaluation of these different perturbation systems. This was the case with Cai et al. (2008) and Tourville et al. (2013) and with the preliminary work by Shih, Suemitsu, & Akagi (2011). Two papers also presented a method to perturb formants in populations in which speech acoustics have deteriorated, by coupling articulatory synthesis with Audapter (Berry, North, & Johnson, 2014; Berry, North, Meyers, & Johnson, 2013).

As displayed in Table 4, most of the studies involved native speakers of English, mainly from North America. Other languages were investigated in a few comparative studies or in relation to other research questions as described in the next section. Potential generalization of these findings to other languages and populations should therefore be taken with caution.

#### **4.2. Main procedures in formant perturbation studies and related concepts**

The main procedures identified in the collected papers about formant perturbations are summarized in Figure 3. These procedures will be referred to in relation to the research topics detailed in the next section. Two main approaches can be distinguished:

- (1) Unexpected formant perturbation during the production of prolonged utterances: This first approach was used in only a few of the collected papers ( $n=11$ ,  $\sim 14\%$  of the papers with formant perturbations, see Table 4). The perturbation is only applied to a small proportion of utterances so that talkers cannot anticipate the perturbation. Moreover, the utterances are produced with long vowel duration (steady-state vowels) so that corrective answers result from online processing of the auditory feedback (cf. Figure 3, procedure P4). This correction is called *compensation*.
- (2) Systematic and constant perturbation over a number of utterances: This second approach was used in the majority of the papers ( $n=66$ ,  $\sim 86\%$ , Table 4). The basic procedure is represented in Figure 3, procedure P1. It involves the production of utterances with “natural” duration, in general. After a baseline with unaltered auditory feedback, the perturbation is introduced either gradually or abruptly, and then systematically applied at a constant level. Depending on the research group, changes in formant production



**Figure 3:** Overview of procedures used in formant perturbations studies. Duration of experimental phases and perturbations were variable across studies. P1 is the basic procedure to study auditory-motor adaptation, used in Munhall et al.'s studies. It was adapted to investigate the transfer of adaptation (P1t) (MacDonald, Pile, Dajani, & Munhall, 2008; Rochet-Capellan, Richer, & Ostry, 2012) and the effect of auditory motor adaptation on perception (P1p) (Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014) or the effect of perceptual training on sensorimotor adaptation (Lametti, Krol, Shiller, & Ostry, 2014). P2 is the procedure used in Houde & Jordan (1998) and then by Perkell et al. (Villacorta, Perkell, & Guenther, 2007). It is structured in epochs with training words produced with feedback followed by training words and generalization words produced with a masking noise. P3 is the multiple perturbation procedure developed in Rochet-Capellan & Ostry (2011), during which words are produced in random order with specific perturbation associated with each word. P4 is the compensation procedure to unpredictable perturbations. In this last case, long steady-state vowels are produced and the perturbation is introduced randomly for a small proportion of utterances to assess online correction (Purcell & Munhall, 2006b). Grey scale gradient in the ramp phase represents the progressive introduction of the shift.

at the end of the training phase are referred to as *compensation* (cf. Houde & Jordan, 1998; Purcell & Munhall, 2006b) or *adaptation* (cf. Rochet-Capellan, Richer & Ostry, 2012, Martin et al., 2018), and residual changes when the feedback is returned to normal after training are referred to as *adaptation* or *after-effect*, respectively.



This procedure was also used to assess *generalization* (or *transfer*) of adaptation to untrained utterances, either in the course of the training phase (Figure 3, procedure P2) or after the training (Figure 3, procedure P1t), as presented in the next section.

Hereafter, *adaptation* will refer to changes observed at the end of the training phase in response to a systematic perturbation. *Compensation* will mainly refer to changes in response to unpredictable perturbations but will also be used to qualify the direction of adaptive responses (by contrast with following responses that go in the same direction as the perturbation).

## 5. Research topics tackled with formant perturbations

In this section, we provide a thematic review of the collected papers that included an empirical study of formant perturbation. As much as possible, we chose to associate each paper with a main topic but obviously a paper could be related to more than one topic. Table 4 provides a list of all the cited references and their main associated research topics.

### 5.1. Properties of feedback and feedforward control

Many studies involving formant perturbations are related to the role of auditory feedback in speech motor control and distinguish between feedback and feedforward control mechanisms. Feedback control is a closed-loop system that involves the sensory consequences of the current motion. It is regarded as too slow to account for rapid control and rapid adjustments observed in fast coordinated actions. Rapidity and adaptability of motion were identified early on as evidence of a feedforward control mechanism by researchers in visuomotor adaptation. The core idea is that the brain makes predictions of the sensory consequences of its actions based on an efference copy of the motor command (Houde & Jordan, 2002). These predictions involve mappings between motor and sensory representations also called *internal models* (Purcell & Munhall, 2006a) or sensorimotor memories (see Perrier, 2012, for a discussion of the nature of internal models in speech). The DIVA (Golfinopoulos et al., 2010) or the SFC (Houde & Chang, 2015) neurocomputational models of speech production assume the existence of both feedback and feedforward control networks that involve auditory and somatosensory systems. When

the prediction based on *internal models* does not match the actual sensory input, the internal representations are changed to reduce this prediction “error” so that future movements performed in similar conditions will be accurate. This mechanism is claimed to underlie sensorimotor adaptation.

In this context, a first subset of studies with formant perturbations was designed to “Investigate the nature, level of details, and use of internal models in speech production” (Max, Wallace, & Vincent, 2003, p. 1053) and to “begin to parameterize the formant feedback system” (MacDonald, Goldberg, & Munhall, 2010 p. 1060). The main contribution of these studies is to describe the role of auditory feedback in the control of formant production, and the adaptability of this control. In these papers, adaptability is mainly explained or taken as an evidence for feedforward internal models.

To address the properties of adaptation to formant perturbations, Houde and Jordan (2002) analyzed in more detail the adaptation phenomenon introduced in Houde and Jordan (1998). The results highlight some properties of feedback and feedforward control that were subsequently discussed and investigated in later work, involving various types of formant perturbations and procedures.

The first observation of Houde and Jordan was that the changes in F1 and F2 production in talkers’ speech were compensatory responses, in the opposite direction to the perturbation. This result has been reproduced consistently in later work when between-speaker data are aggregated. Individual data suggests that some speakers follow the shift, however. For example, in a meta-analysis of their own studies of adaptation to formant perturbations, MacDonald et al. (2011) found that 26 out of 116 female speakers followed F1 or F2 shifts when their production of “head” was perturbed toward “had”. A possible explanation is that non-adapted speakers may not be able to dissociate their own production from the auditory feedback (Vaughn & Nasir, 2015). Following the formant shift rather than compensating for it was actually the most frequent behaviour observed in a preliminary study investigating compensation in Japanese speakers to unexpected perturbations of F1, F2 and F3 (Shih et al., 2011). Aside from this study, all other published work on formant perturbations observed significant compensatory adaptation in acoustic analyses, whereas preliminary analyses of articulatory correlates of adaptation are

**Table 4:** List of all the studies related to formant perturbation included in the present review. The first column provides the reference of the article. The 2nd column gives the language of participants (Du: Dutch, En: English, Fr: French, Ge: German, Ja: Japanese, Ko: Korean, Ma: Mandarin, Ru: Russian, Sp: Spanish). Column 3 is related to the perturbation systems, which are described in Table 3 (briefly, 1.a: Houde & Jordan (1998), 1.b. Katseff et al. (2012); 2: Purcell & Munhall, (2006a); 3: Audapter and its previous versions; 4: VoiceOne, TC Helicon, 5: Others) and column 4 indicates whether an article is mainly dedicated to the description of a perturbation system. Each study has been classified into either compensation (to unpredictable perturbations, column 5) or adaptation (to sustained perturbations). Columns 7 to 14 show whether the article is related to each of the main research topics presented in the present review. A cross indicates that the article is cited in the corresponding subsection, while a (X) indicates it is not although it is related to the topic.

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Alsius, Mitsuya, Latif, & Munhall, 2017	En	2			X		(X)						X
Berry, Jaeger, Wiedenhoeft, Bernal, & Johnson, 2014	En	3			X	X			X				
Berry, North, & Johnson, 2014	En	3	X										
Berry, North, Meyers, & Johnson, 2013	En	3	X										
Bourguignon, Baum, & Shiller, 2014	En	4			X			X					
Bourguignon, Baum, & Shiller, 2015	En	4			X			X					
Bourguignon, Baum, & Shiller, 2016	En	4			X			X					
Cai, Beal, Ghosh, Tiede, Guenther, & Perkell, 2012	En	3			X		(X)			X			

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Cai, Boucek, Ghosh, Guenther, & Perkell, 2008	Ma	3	X		X								
Cai, Ghosh, Guenther, & Perkell, 2010	Ma	3			X	X		X					
Cai, Ghosh, Guenther, & Perkell, 2011	En	3		X	X								
Caudrelrier, Perrier, Schwartz, & Rochet-Capellan, 2016	Fr	3			X			(X)	X				
Caudrelrier, Perrier, Schwartz, & Rochet-Capellan, 2018	Fr	3			X			(X)					X
Caudrelrier, Schwartz, Perrier, Gerber, & Rochet-Capellan, 2018	Fr	3			X			(X)	X				
Daliri, Wieland, Cai, Guenther, & Chang, 2018	En	3			X		(X)		X				
Lametti, Krol, Shiller, & Ostry, 2014	En	4			X			X					
Lametti, Nasir, & Ostry, 2012	En	4			X		X						
Lametti, Smith, Freidin, & Watkins, 2018	En	4			X						X		
Demopoulos et al., 2018	En	1b		X	X				X		(X)		
Deroche, Nguyen, & Gracco, 2017	En	4			X		(X)			X			
Dimov, Katseff, & Johnson, 2012	En	1b			X								X
Eckey & MacDonald, 2015	Ge	5		X			X						
Feng, Gracco, & Max, 2011	En	4			X		X						
Houde & Jordan, 1998	En	1a			X	X			X				

(continued on next page)

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Houde & Jordan, 2002	En	1a			X	X							
Ito, Coppola, & Ostry, 2016	En	4			X		(X)				X		
Katseff & Houde, 2008	En	1b			X		(X)						
Katseff, Houde, & Johnson, 2012	En	1b			X		X						
Klein, Eugen; Brunner, Jana; Hoole, Phil (sous press)	Ru	3			X				X				(X)
Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014	En	4			X			X					
MacDonald & Munhall, 2012	En	2			X		X						
MacDonald, Goldberg, & Munhall, 2010	En	2			X	X	X						
MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012	En	2			X							X	
MacDonald, Pile, Dajani, & Munhall, 2008	En	2			X				X				
MacDonald, Purcell, & Munhall, 2011	En	2		X		X							
Martin et al., 2018	Sp	1b			X		X						
Max & Maffett, 2015	En	4			X	X							
Max, Wallace, & Vincent, 2003	En	5			X	X							
Mitsuya & Purcell, 2016	En	2			X	X							
Mitsuya, MacDonald, Munhall, & Purcell, 2015	En	2			X		X						
Mitsuya, MacDonald, Purcell, & Munhall, 2011	En	2			X			X					
Mitsuya, Munhall, & Purcell, 2017	En	2				X							

Table 4: Continued

References	Language	Perturbation System	System description	Compensation	Adaptation	Properties of feedback and feedforward control	Perception acuity and sensory integration	Perceptual & phonological categories	Transfer/Specificity and speech units	Pathology affecting speech production	Neural basis of speech motor learning	Development	Surface effects & speakers' characteristics
Mitsuya, Samson, Ménard, & Munhall, 2013	Fr	2			X			X					
Mollaei, Shiller, & Gracco, 2013	En	4			X					X			
Mollaei, Shiller, Baum, & Gracco, 2016	En	4		X			(X)			X			
Munhall, MacDonald, Byrne, & Johnsrude, 2009	En	2			X	X							(X)
Neufeld, Purcell, & Van Lieshout, 2013	Ko	2			X	X							
Niziolek & Guenther, 2013	En	3		X				X					
Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017	En	1b		X	X					X			
Pile, Dajani, Purcell, & Munhall, 2007	En	2			X				X				
Purcell & Munhall, 2006a	En	2			X	X							
Purcell & Munhall, 2006b	En	2		X		X							
Purcell & Munhall, 2008	En	2			X	X	X						
Reilly & Dougherty, 2013	En	3		X			X	(X)					
Reilly & Pettibone, 2017	En	3			X				X				
Rochet-Capellan & Ostry, 2011	En	4			X				X				
Rochet-Capellan, Richer, & Ostry, 2012	En	4			X				X				
Sato & Shiller, 2018	Fr	3			X		(X)				X		X
Schuerman, Nagarajan, & Houde, 2015	En	1b			X			X					
Schuerman, Nagarajan, McQueen, & Houde, 2017	En	1b			X			X					

(continued on next page)

[illegible]

less clear. Max et al. (2003) analyzed acoustic changes to perturbation of all formants in the same direction in relation to jaw and tongue movement during adaptation. No consistent behaviour were observed in articulatory kinematics. Similar results were obtained in a pilot study in one Korean speaker with an F2 shift (Neufeld, Purcell, & Van Lieshout, 2013), while clearer tongue compensation movements were reported in speakers with blindness (Trudeau-Fisette, Tiede, & Ménard, 2017). On the other hand, while the majority of studies on adaptation to formant perturbations found significant compensatory responses, it was also shown that adaptation vanishes when perturbed feedback is delayed by more than 100ms (Max & Maffett, 2015), or is at least largely reduced (Mitsuya, Munhall, & Purcell, 2017).

Houde and Jordan also reported that maximal changes at the end of training did not fully compensate for the perturbation. This result was systematically reproduced in later studies. As an illustration, in Purcell & Munhall (2006a), the maximal adaptation to a 200Hz upward vs. downward shift of F1 compensated for about 30 % of the perturbation, regardless of the number of repetitions during the hold phase. This also suggests that adaptation is a fast process, in agreement with Max et al. (2003)'s observation that compensatory responses occurred after only a few repetitions. However, a F1 perturbation of at least 60Hz (80Hz on average across conditions) was required in Purcell & Munhall (2006a) to initiate the compensatory response. Similar thresholds were reported in later work, regardless of the delay in the auditory feedback (Mitsuya et al., 2017) and the occlusion of the headphones (Mitsuya & Purcell, 2016). Furthermore, MacDonald et al. (2010) highlighted a linear relationship between the magnitude of the perturbation and the magnitude of changes in speakers' utterances for perturbation magnitudes up to +200Hz in F1 and -250Hz in F2, compensating for 25 % of the perturbation in F1 and 30 % in F2. With larger perturbations, there was no improvement, and a decrease even appeared in response to perturbations larger than 300Hz in F1 and larger than 400Hz in F2. Similar limits were observed by Katseff and colleagues (Katseff & Houde, 2008; Katseff et al., 2012), as discussed in the next section. Comparable adaptations were reported in the meta-analysis provided by MacDonald et al. (2011), with an average of 26.5 % for F1 and 23.2 % for F2. Moreover, in this last analysis, changes in F1 in



speakers' production weakly correlated with changes in F2, suggesting a specific control of the two parameters and the existence of speaker-specific strategies. The magnitude of the response was also found to vary according to the vowel in *pet*, *bus* and *law* utterances in Max et al. (2003). Further work addressing this last point with regard to more specific research topics is presented in the next section.

Houde and Jordan also noticed that inter-speaker variability was not related to a speaker's awareness of the auditory shift. When interviewed after the study, talkers reported they were unaware of the perturbation or of any change in their production. By contrast, Purcell & Munhall (2006a) reported that 40 % of their participants indicated awareness of "some kind of change in the auditory feedback over the course of the experiment", with only 8 % noticing that the perturbation transformed the vowel into a different one. However, the magnitude of adaptation did not seem to be related to the responses in this interview. This difference to Houde and Jordan might be related to the abrupt suppression of the perturbation after training in Purcell & Munhall (2006a) (Procedure P1, Figure 3) that was probably perceived by the speakers, while Houde and Jordan assessed how adaptation was sustained using catch trials with masking noise (Procedure P2, Figure 3). Munhall, MacDonald, Byrne, & Johnsrude (2009) then confirmed that the awareness of the perturbation does not influence adaptive behavior, as discussed later in the "*Surface effects & speakers' characteristics*" subsection.

Another important result in Houde and Jordan was that changes for perturbed utterances were larger than changes for utterances produced with a masking noise. The authors discussed this result as evidence that "vowel production could be partly under immediate auditory feedback control" (Houde & Jordan, 2002, p. 307). By contrast, in their preliminary study of adaptation to a shift of all formants in the same direction, Max et al., (2003) argued that the modifications in talkers' production should be considered as adaptive responses rather than reactive changes, as they already occur at vowel onset, and have been observed for sustained vowels as well as vowels with shorter duration. The variability of changes in formants according to the vowel's parts were not systematically investigated in adaptation studies as most of the studies used a single steady-state value, often around the middle of the vowel. However, in their preliminary

work, Berry, Jaeger, Wiedenhoeft, Bernal, & Johnson (2014) suggested that this single value might not be the most appropriate, depending on consonant context and coarticulatory effects. Vaughn and Nasir (2015) also provided evidence that full trajectory analysis might better capture adaptation phenomena. The relationship between formant values in consecutive trials (as measured with one-lag cross correlation analyses), in the absence of any perturbation, may also be predictive of adaptation magnitude (Purcell & Munhall, 2006a). Altogether, these results suggest that changes observed over the course of adaptation to a perturbation result probably from a mix of feedback and feedforward control.

Houde and Jordan (2002) suggested investigating compensation to formant perturbations in steady-state vowels to determine the role of online feedback in formant control. Studies focusing on compensation to an unexpected formant perturbation in sustained vowels usually analyzed changes at different points of the vowel. For instance, in Purcell and Munhall (2006b) upward vs. downward perturbations of F1 were applied randomly in five utterances of “head” over 100 utterances of different CVC words. Results show partial compensation, with on average, 16.3 % vs. 10.6 % of the upward vs. downward shifts, but with high variability for the same talker between utterances and between talkers. However, this study was not designed to measure the delay in compensatory response. This delay was found in later studies to be around 160ms, at least when F1 is shifted upward (e.g. Tourville, Reilly, & Guenther, 2008), and when more complex spatial or temporal perturbations of formants trajectories are applied during the production of short sentences (Cai, Ghosh, Guenther, & Perkell, 2011). The smaller compensation of perturbation observed in studies involving unexpected perturbation compared to studies involving systematic perturbation, as well as the delay required to observe a compensatory response, confirm the idea that responses produced in the presence of the perturbation in adaptation studies are at least partially adaptive.

One of the most intriguing outcomes of Houde and Jordan (2002) was that the modification in formants was still present when talkers came back a month later to run a control study evaluating changes in production without perturbation. This long-term effect was attributed by the authors to implicit memory of the task or specific control mechanisms for

whispered speech. Although not reproduced in later work – as there was no study with equivalent long-term assessment in our review at least – Purcell and Munhall, (2006a) showed that 115 repetitions without perturbation after the training phase were not enough to fully return to the baseline state. The explanation introduced by Houde and Jordan echoes the idea that auditory-motor learning could be specific to some situations or ways of speaking, as discussed in generalization studies. The ability to memorize specific ways of speaking according to the situation could be a way to support fast speech adaptability in known situations. This idea could be further investigated by means of transfer of adaptation from one context to another as discussed below.

Finally, large inter-speaker variability was also pointed out in Houde and Jordan (2002) and then observed in all the subsequent studies. MacDonald et al. (2010, 2011) suggested that this variability is not clearly related to the variability in baseline production, nor to the size of the vowel space. Inter-speaker variability, as well as partial compensation, in formant adaptation studies was often discussed in terms of a tradeoff between auditory and somatosensory feedback. For example, Purcell and Munhall (2006a) suggested that “Some [speakers] may rely more on kinesthetic feedback and thus are not influenced as much by acoustic feedback” (p. 975), while Houde and Jordan (2002) suggested “it may be that there are differences across participants as to the degree to which they rely on auditory feedback” (p. 308). The tradeoff between auditory and somatosensory feedback, as well as the role of sensory acuity in adaptation was then explored in several papers, as described in the next section.

## 5.2. Perception acuity and sensory integration

Formant perturbations’ paradigms involve modifying the auditory feedback, i.e. sensory input of speech control system, and measuring the outcome in terms of speech production, or motor control. Hence these paradigms are by nature relevant to the question of the relationship between perception and production. Several aspects of this relationship have been investigated over the past two decades.

First, adaptation to auditory perturbations may be influenced by speakers’ sensory acuity. Auditory acuity has been positively correlated with adaptation magnitude in two studies (Martin et al., 2018;

Villacorta et al., 2007) involving 13 and 31 subjects respectively. Auditory acuity measurements were based on discrimination tasks in both cases. Villacorta et al. focused on acuity for F1 while Martin et al. measured acuity based on pitch and loudness, as well as melody discrimination tasks. A possible interpretation of the relation between adaptation magnitude and auditory acuity is that better acuity could lead speakers to have smaller goal regions for their production, resulting in higher adaptation (Villacorta et al., 2007). However, auditory feedback may not be the only feedback used to control speech production. Feng et al. (2011) investigated the relationship between the adaptation magnitude of F1 and the auditory acuity for F1, as well as somatosensory acuity for jaw position. They did not find a reliable correlation. However, fewer subjects were involved in this study than in previously cited ones (8 subjects vs. 13 and 31).

Feng et al. also combined a somatosensory perturbation induced by a robotic device pulling the jaw, with an auditory shift on F1. Using this procedure, they found that speakers mainly compensated for the auditory perturbation. They suggested that auditory feedback may be dominant over somatosensory input, but that their relative weight could evolve with speech experience. Using similar methods, Lametti, Nasir, and Ostry (2012) found that all speakers adapted for at least one of the two perturbations. The group who adapted to the somatosensory perturbation (half of the participants) did not significantly compensate for the auditory perturbation while the group that did not adapt to the jaw perturbation significantly compensated for the F1 shift. This observation suggests a speaker-specific sensory preference for either auditory or somatosensory inputs. In addition, the weights attributed to auditory and somatosensory feedback may vary according to the articulator (i.e. vocal folds, tongue or jaw) to control. Indeed, no correlation has been found in the magnitude of adaptation in F0, F1 and F2 across speakers while altering them simultaneously or separately (Eckey & MacDonald, 2015; MacDonald & Munhall, 2012). Interestingly, Trudeau-Fisette et al. (2017) showed that speakers with blindness adapted more to an F2 shift than control speakers, independently of their auditory acuity, and that they also produced larger articulatory changes in response to the auditory shift. Speakers with blindness may rely more on auditory feedback than control speakers, who

may have more precise somatosensory goals, probably built and supported by visual perception of speech.

However, sensory preference in the control of speech, which can be modeled by different weights attributed to each kind of sensory feedback, may also evolve with experience. Most studies on auditory-motor adaptation report a partial compensation for the auditory perturbation as already mentioned in the previous section. Some studies showed that the percentage of compensation relative to the magnitude of the perturbation decreases when the magnitude of the perturbation increases, reaching an asymptote, and can even tend to decrease for larger perturbations (Katseff & Houde, 2008; Katseff et al., 2012; MacDonald et al., 2010). Katseff et al. (2012) interpreted this phenomenon as evidence that the weights attributed to auditory and somatosensory feedback may vary according to experience: “For small discrepancies between auditory and somatosensory feedback, auditory feedback takes precedence, and for large discrepancies between auditory and somatosensory feedback, somatosensory feedback takes precedence” p. 307. Thus, a high-amplitude shift may lead the speech system to consider auditory feedback as unreliable and therefore give more weight to somatosensory feedback. In addition, the relative importance of sensory input may depend on the specific sounds produced. Several studies observed less compensation in closed vowels than in open vowels (Mitsuya et al., 2015; Purcell & Munhall, 2008; Reilly & Dougherty, 2013). This could be explained by better-specified somatosensory information in the former than in the latter case (Mitsuya et al., 2015). Another possible explanation is that the importance of F1 as an acoustic cue in perception may depend upon the vowel (Reilly & Dougherty, 2013).

### 5.3. Perceptual and phonological categories

Speech perceptual space is structured by phonological categories, which are delimited by perceptual boundaries. Niziolek and Guenther (2013) showed an effect of perceptual boundaries on the magnitude of compensation to unpredictable auditory perturbations. They observed that if the auditory signal resulting from the perturbation is near a boundary, the compensation, as well as the cortical activation, is higher than when it

is far from a boundary, the magnitude of the shift being equal. In addition, various studies have investigated the relation between perceptual boundary and adaptation to sustained auditory perturbations.

The influence of perceptual boundaries on adaptation can be investigated using perceptual learning on the perceptual contrast that is at stake in the adaptation paradigm. For instance, Shiller, Lametti, & Ostry (2013) manipulated speakers' perceptual boundaries between "head" and "had" through perceptual training preceding auditory-motor adaptation to a perturbation consisting of altering "head" into "had" (see Procedure P1p on Figure 3). The group whose boundary was shifted towards "head" (i.e. who was more likely to classify ambiguous stimuli as "had") adapted more to the auditory perturbation than the group whose boundary was shifted towards "had" by the perceptual training. Similarly, children adapted more to a perturbation transforming /beb/ into /bab/ after a perceptual training manipulating /ɛ/-/æ/ boundary towards /ɛ/ than before training (Shiller & Rochon, 2014). They also adapted more than children having undergone a perceptual training on an unrelated contrast. Furthermore, Lametti, Krol, et al. (2014) observed in adults that the amount of adaptation to auditory-feedback perturbation was correlated with the position of the perceptual boundary obtained through perceptual training.

Instead of using perceptual training, changes in perceptual boundaries were obtained by manipulating the pitch and formant of the carrier phrase "please say what this word is..." (Bourguignon, Baum, & Shiller, 2015, 2016). In this study, the group exposed to high carrier-phrase (high pitch and formants) had the boundary between 'bit' and 'bet' shifted toward 'bet'. They adapted more to an auditory feedback alteration transforming /ɛ/ into /ɪ/ than the speakers exposed to low carrier-phrase (low pitch and formants). This finding suggests that "context-dependent plasticity in speech perception may also transfer to production" (Bourguignon et al., 2016, p. 1040). Interestingly, Bourguignon, Baum, and Shiller (2014) also showed an effect of the lexical status that can be interpreted in terms of perceptual boundaries. In their study, a group of speakers produced pseudo-words that resulted in real word when auditory perturbation was applied (e.g. "kess" changed into "kiss"). Another group produced real words that were transformed into pseudo-words by the same formant shift

(e.g. “less” changed into “liss”). The first group showed greater adaptation than the second group, indicating a lexical effect on auditory-motor adaptation.

The influence of phoneme categories on speech motor adaptation was also highlighted in cross-language studies. Mitsuya, MacDonald, Purcell, and Munhall (2011) contrasted the adaptation to upward and downward shifts in F1 in three groups: English speakers pronouncing “head”, Japanese speakers producing the Japanese word /he/ and Japanese speakers learning English, producing “head”. The magnitude of adaptation was equivalent in all groups in response to the downward shift, but the adaptation was smaller in Japanese than in English speakers in response to the upward shift. This difference is evidence for the influence of the phonological system in adaptation. Mitsuya, Samson, Ménard, and Munhall (2013) also showed differences between English speakers and French speakers in the adaptive response to the same auditory perturbation. In this study, a perception test suggested that this language effect on adaptation was mediated by a difference in perceptual boundaries: larger adaptation in French speakers was related to greater sensitivity to some phonetic contrasts.

Reciprocally, the influence of adaptation on perceptual boundaries has also been investigated. Lametti et al. (2014) incorporated perceptual tests in a classic auditory-motor procedure (Figure 3, procedure P1p), before and after the training phase – during which adaptation occurs – as well as after the after-effect phase, used here as a wash-out of adaptation. They observed that auditory-motor adaptation resulted in a shift of a perceptual boundary in the phonetic range of what speakers produced but not what speakers heard. For instance, speakers who produced “head” and heard an auditory feedback shifted toward “had”, compensated by producing an utterance closer to “hid”. Their perceptual boundary between “head” and “hid” was shifted toward “head”, that is, speakers became more likely to report hearing ‘hid’ in the perceptual test, while there was no effect on the perceived boundary between “head” and “had”. This result suggested that the change in perception was specifically driven by speech motor adaptation and not by the auditory input during learning. The interpretation of these results, together with the results of other studies on the effect of auditory-motor adaptation on categorical perception, was recently specified in a Bayesian modeling framework, suggesting

that speech motor adaptation results both in speech sound remapping and changes in phoneme categories (Patri, Perrier, Schwartz, & Diard, 2018). Yet, using a similar paradigm to that of Lametti et al. (2014), Schuerman, Meyer, and McQueen (2017) did not find significant influence of auditory-motor adaptation on related perceptual boundaries. It should be noted that this experiment had fewer subjects than Lametti et al. (2014); was run with speakers of Dutch as opposed to English; and used a continuum with isolated vowels rather than a continuum between words, during the perceptual test. However, this last study also recorded EEG signals during initial vs. final perception tests. The analysis of ERPs to the stimuli of the / $\epsilon$ /-/ $u$ / continuum revealed changes in N1 and P2 components for ambiguous stimuli, which correlated with the magnitude of adaptation as measured by F1. The effect on both N1 and P2 suggest that auditory-motor adaptation influences both early perception and late perceptual decisions. Interestingly, Schuerman, Nagarajan, and Houde (2015) and Schuerman et al. (2017) showed that the adaptation to an auditory perturbation of F2 shifting the front vowel / $i$ / towards the back-vowel / $u$ / resulted in a shift in the perceptual boundary between “see” and “she”. More specifically, the shift in perceptual boundaries depended on the behavior of speakers during the adaptation task: speakers who followed the auditory perturbation had their perceptual boundary shifted in the opposite direction to that of speakers who compensated for the auditory feedback. This last group was more likely to categorize ambiguous stimuli as “see” than “she”, the place of articulation of the consonant / $s$ / being more anterior than / $ʃ$ /. These findings are in agreement with the idea that some transfer of adaptation may occur between vowels and consonants articulated with a similar tongue position.

While this impact of a change in production on the perception of another contrast is actually a transfer from production to perception, the term *transfer* is typically investigated in speech production itself, from one utterance to another.

#### 5.4. Transfer/specificity and speech units

In the limb movement literature, generalization of motor learning is the “ability to correctly extrapolate to contexts that are different from our limited experience” (Krakauer, Mazzoni, Ghazizadeh, Ravindran, &



Shadmehr, 2006, p. 1798). This extrapolation could be the result of an interpolation of previous experiences (Mattar & Ostry, 2007). Generalization has been extensively investigated in motor learning research, and in speech, in particular to address the specificity of motor adaptation and the underlying representations (Tremblay, Houle, & Ostry, 2008). Transfer of adaptation is usually defined as a positive generalization, as opposed to interference (Krakauer et al., 2006). However, we will use generalization or transfer to designate changes observed in untrained utterances after adaptive training, going in the same direction as adaptation. When no significant transfer is observed, changes related to adaptation are considered to be specific to the training utterance.

The investigation of generalization or transfer of adaptation relied on two different motivations. The first set of work focused on generalization as a way to assess the global vs. specific nature of auditory-motor mapping. This approach is derived from limb movement studies that analyzed generalization of visuomotor adaptation to address the global vs. specific nature of visuomotor mapping. The second set of work, that is sometimes an extension of the first one, considered generalization of auditory-motor learning as a way to assess the nature of speech production units, by questioning the linguistic level of auditory-motor mapping. This second approach was introduced by Houde and Jordan (1998) and is consistent with earlier work on transfer of perceptual learning to assess speech perception units (e.g. Chambers et al., 2010).

Different procedures were used to investigate generalization of auditory-motor adaptation. The first one is structured in “epochs” (Figure 3, P2). Each epoch includes utterances with feedback on and utterances with a masking noise, which can be either the training utterances or different utterances. Transfer is evaluated at the end of the training phase, when the perturbation is maximal by measuring changes in transfer (or test) utterances as compared with their baseline. Using this procedure, Houde and Jordan (1998) found significant transfer from the training words sharing the same vowel / $\epsilon$ / (“pep”, “peb”, “bep”, and “beb”), shifted toward / $i$ / or / $\ae$ / to the various test words (same vowel as training words – “gep”, “peg”, “teg”, or different vowels – “pip”, and “pap”). The amount of transfer was variable depending on the test word, but not statistically different. Consistent results were reported in Villacorta et al. (2007),

where adaptation on the vowel / $\epsilon$ / for nine CVC words to an F1 perturbation, significantly generalized to the same vowel in different CVC or to the vowels in “pit”, “pat”, and “pot”. Results were less consistent for “put” and “pete” and seemed to depend on the direction of the perturbation. Still with a similar procedure, but with perturbation of F1 trajectory in speakers of Mandarin, Cai et al., (2008) and Cai et al. (2010) found gradients of generalization that depended on the similarity in formant trajectory between a training triphthong and the tested utterances. Finally, Reilly and Pettibone (2017) tested generalization from the vowels /i/ vs. / $\epsilon$ / (embedded in a set of CVC utterances) to /i/, / $\epsilon$ / and / $\epsilon$ / (also in CVC) produced with a masking noise. In both training conditions, / $\epsilon$ / was the “near” vowel in test utterances, while /i/ and / $\epsilon$ / were either the same as the training vowel or the “far” vowels, depending on the training condition. Adapted speakers exhibited significant generalization to all vowels, regardless the training vowel. However, correlation between adaptation and generalization were unclear suggesting that generalization may depend on multiple factors and may be sensitive to inter-speaker variability.

Similar procedures, mixing training and transfer trials, were used in limb movement studies. However, the approach was later criticized. In particular, with this procedure, “the patterns of generalization observed are difficult to interpret, as transfer could reflect an averaging that takes place when subjects experience several training conditions simultaneously” (Rochet-Capellan et al., 2012 p. 1711). For this reason, other studies tested transfer after the training phase, when the feedback is turned-back to normal (Procedure P1t on Figure 3). In preliminary work, MacDonald et al. (2008) compared transfer tested in the course of training vs. after training. In both cases, speakers were trained on “head” shifted towards “had” and transfer was tested on the production of “hid” with unaltered feedback. When the transfer utterance “hid” was inserted during training, changes in “hid” were observed at the beginning of the training phase, but then its production came back to baseline. When tested after training, no change was observed at all in “hid”. Overall this suggests that adaptation is specific to the trained vowel, although it slightly depends on the training conditions. Pile, Dajani, Purcell, and Munhall (2007) then observed similar adaptation and lack of generalization toward “hid” or “hayed” (i.e. /hed/). Both studies were published in proceedings and were preliminary,

with restricted analyses. Later work by Rochet-Capellan et al. (2012) evaluated how adaptation to a perturbation of F1 in /pen/, /ben/, /ken/, /gen/, /ten/, /den/, /pan/, /pin/ then affect the production of /pen/ produced without perturbation. Results were consistent with previous work that tested generalization with a mixed procedure (Figure 3, P2): generalization was variable according to the training word and seemed to depend on the acoustical proximity between the training and the testing utterance. Another important result of this work was that the after-effect, assessed on the training utterance after the transfer phase, was still significant, suggesting that the production of the transfer utterance with normal feedback did not wash out adaptation. This last result is consistent with the idea that learning is related to the training experience, and at least to some extent specific to this experience.

Another way to assess specificity of adaptation is to evaluate how speakers can specifically compensate for several perturbations in the same training session. This approach is inspired by limb movement studies and in particular Osu, Hiraï, Yoshioka, and Kawato (2004). In Rochet-Capellan and Ostry (2011), speakers produced “head” and “had” in random order with F1 shifted downward in “head” and upward in “had” and conversely (Procedure P3 in Figure 3). On average, speakers were able to change F1 frequency in opposite directions for “head” and “had”, suggesting that auditory-motor mapping is specific to each vowel. To assess whether auditory-motor mapping could be specific to a word, the authors then evaluated multiple adaptations for “head” and “bed” shifted in opposite directions and “ted” un-shifted. Again, on average, specific adaptation in opposite directions were observed for “head” and “bed” while F1 in “ted” remained unchanged, suggesting that different auditory-motor mappings could be built for a same vowel in different words. Similar results were obtained recently by Klein, Brunner and Hoole (in this book) with a Russian vowel in /d/ vs. /g/ CV syllables and a perturbation of F2. The authors also provided analysis of speakers’ data showing symmetrical vs. asymmetrical profiles of adaptation.

Altogether, these results suggest that generalization of auditory motor adaptation occurs in a way that depends on the similarity between the training and the testing utterance and that specific control can be achieved, at least under specific conditions. The results were interpreted as an

indication of global control for vowel production vs. specific control. Furthermore, generalization from a vowel to the same vowel in different contexts suggests that auditory-motor mapping could occur at the level of the phoneme. It is thus a way to question the structure of feedforward mapping, and the nature of its underlying representations (Houde & Jordan, 1998). The fact that transfer is in general smaller than after-effect suggests that word context may play a role. The idea that multiple representations may coexist in auditory-motor mapping of speech was directly assessed in recent papers by Caudrelier et al. (Caudrelier et al., 2016; Caudrelier et al., 2018). In this work, several linguistic levels were contrasted by assessing transfer on test utterances that shared either the same vowel, and/or the same syllable or was the same word as the training utterance. Transfer was smaller (although significant) at the vowel level than transfer to the same syllable, which was lower than after-effect in the same word, suggesting that these three levels – words, syllables, phonemes – could coexist in parallel in the structure of the speech sound map. This conclusion is consistent with multiple traces connectionist models of long-term memory (Ans, Carbonnel, & Valdois, 1998; Carbonnel, Charnallet, & Moreaud, 2010) in the sense that multiple units could emerge as common information of multiple experiences (Goldinger, 1998; Hintzman, 1986). Specific production of the vowel to the syllable or word context also questions the role of coarticulation in adaptation and transfer of adaptation, a topic introduced in a preliminary paper by Berry et al. (2014).

In addition to the theoretical insights mentioned above, a better understanding of generalization in speech may have clinical implications in speech rehabilitation (e.g. after stroke), since transfer from training with a speech therapist to daily life is essential (Aichert & Ziegler, 2013). Other clinical applications are described in the next section.

### 5.5. Pathology affecting speech production

Auditory feedback perturbation paradigms may be instrumental in the understanding of mechanisms underlying disorders related to or affecting speech production. In particular, low compensation or adaptation observed in patients with a given pathology is regarded as evidence for a lack of sensorimotor integration or as an impairment of feedforward control mechanisms.

Stuttering is suspected to be driven by abnormal integration of sensory input in speech motor control, and has been an early target for auditory perturbation studies, and more recently for studies using formant perturbations. Cai et al. (2012) observed smaller compensation to unpredictable perturbation of formants in persons who stutter compared to control participants. The latency of compensation was however found to be equivalent in both groups. According to the authors, this suggests impairment of the inverse model responsible for translating auditory error detection into proper correction in motor commands. Reduced responses to formant perturbations were also observed in adaptation studies, with systematic perturbations. Sengupta, Shah, Gore, Loucks, & Nasir (2016) found smaller adaptation in adults who stutter as compared with control speakers that was also related to anomalous EEG phase coherence. This hints at a miscommunication between speech sensory and motor areas, which confirms a potential deficit in sensorimotor integration in people who stutter. A recent study, Daliri, Wieland, Cai, Guenther, & Chang (2018) also found reduced adaptation in adults who stutter compared to control speakers. However, the difference was not observed in children who stutter as compared with their aged-match controls. These results suggest that reduced adaptation observed in adults may be a consequence of compensatory strategies induced by the pathology rather than a root cause.

Terband, Van Brenk, and van Doornik-van der Zee (2014) used a similar adaptation paradigm as Daliri et al. (2018) with children with CAS (Childhood Apraxia of Speech). CAS was described as “a disordered development of the functional synergies/coordinative structures that underlie speech motor coordination causing impairment of the forward model leading to poor feedforward control” (Terband et al., 2014, p. 66). In agreement with this description, children with CAS were shown to follow the auditory perturbation on average, while their aged-match controls adapted to the perturbation by compensating for it.

Van den Bunt et al. (2017) used formant adaptation to assess the nature of the phonological deficit observed in dyslexia, known as a “difficulty in acquiring fluent word-decoding skills” (p. 1). Adults with dyslexia showed greater adaptation and after-effects than control speakers to a formant feedback perturbation that doesn’t cross a phonemic boundary (i.e. an

allophonic perturbation). Moreover, a negative correlation was observed between reading skills and the magnitude of adaptation: the worse the reading score, the larger the adaptation. This result could be interpreted as a weaker perceptual magnet effect (Kuhl et al., 2008) in speakers with dyslexia and supports theories claiming that dyslexia is associated with a greater distinction between allophones, which may lead phoneme categories to be less prominent. However, a condition with a perturbation crossing the phonetic boundary is required to further support this hypothesis.

Compensation or adaptation to formant perturbations were also investigated in populations with neurogenetic or neurodegenerative diseases. Demopoulos et al. (2018) used adaptation to formant perturbation to address the origin of the speech production deficit observed in young individuals with a subtype of autism (due to a 16p11.2 deletion). The adaptation was reduced in this population as compared with age-matched controls while compensation to unexpected perturbation of F0 was larger. According to the authors, this suggests that feedforward models could be altered in people with 16p11.2 deletion, leading to an over-reliance on feedback control. A comparable profile of larger compensation to unexpected perturbation of F0 was observed in patients with Parkinson Disease (PD). However, both compensation and adaptation to unexpected vs. constant formant perturbation were reduced in speakers with PD as compared with age-matched control speakers (Mollaei, Shiller, Baum, & Gracco, 2016; Mollaei, Shiller, & Gracco, 2013). The authors interpreted the difference in pitch and formant compensation in terms of somatosensory and muscle activation deficits of the larynx and oral cavity. This dissociation between compensation to F0 vs. formant perturbations calls into question the conclusion of Demopoulos et al. (2018): as feedback control was only assessed with F0 in speakers with 16p11.2 deletion, it remains unclear whether they indeed rely more on feedback control in general or if the effect was specific to F0 control. Finally, Parrell, Agnew, Nagarajan, Houde, & Ivry (2017) found that speakers with cerebellum degeneration compensate for unexpected formant perturbations more than their age-matched controls, while they show weaker adaptation to sustained perturbation. This suggests that the cerebellum plays an important role in feedforward control, and probably less in feedback control. The involvement of the cerebellum in feedback control is discussed in the next section.

## 5.6. Neural basis of speech motor learning

The neural correlates of speech motor control and learning have been investigated through a variety of techniques, including EEG, fMRI, rTMS and tDCS.

fMRI is not suitable to observe changes in the timeframe of adaptation to sustained perturbation because it could be confounded with low-frequency noise observed in fMRI (Zheng et al., (2013). However, it is feasible to investigate the neural networks involved in feedback control using unpredictable perturbations. In Tourville et al. (2008), trials under altered auditory feedback (as opposed to normal feedback) were associated with increased bilateral activation in posterior auditory cortex (including posterior Superior Temporal Gyrus, pSTG, and Planum Temporale, PT). This observation is regarded as evidence for the existence of auditory error cells, dedicated to detect errors in auditory feedback. The increased activation in right pSTG was observed to be enhanced when auditory perturbation outcomes were close to a perceptual boundary. In addition, Tourville et al., (2008) found increased right activation in ventral Motor and Premotor Cortex (vMC and vPMC, respectively) and anterior medial cerebellum (amCB). This suggests that feedback control involves mainly the right hemisphere whereas the left hemisphere, which is known to be dominant in speech production, would be mainly associated with feedforward control. Zheng et al., (2013) conducted further fMRI investigation. Their experimental procedure consisted of production trials with normal feedback, altered feedback (with F1 shift) and feedback with masking noise. Speakers then passively listened to every signal corresponding to their auditory feedback in the production session. Combining fMRI with an analysis of neural pattern similarity analysis enabled differentiation of three functional networks: an error signal network (including right AG, right SMA, and bilateral cerebellum), a passive listening network, and a network responding to both production and passive listening conditions, that may correspond to sensorimotor integration, located in bilateral Inferior Frontal Gyrus (IFG).

The Inferior Parietal Lobe (IPL), which comprises Supramarginal Gyrus (SMG) and Angular Gyrus (AG) may be involved in multisensory integration. An rTMS stimulation applied over the SMG just before the auditory-motor adaptation procedure reduced adaptation responses in

comparison with a sham stimulated group (Shum et al., 2011). Similarly, a tDCS stimulation applied over IPL affected auditory-motor adaptation (Deroche, Nguyen, & Gracco, 2017). More specifically, anodal stimulation aiming at facilitating neuronal excitability resulted in stronger adaptation magnitude whereas cathodal stimulation, which has an inhibitory effect, prevents auditory-motor adaptation to predictable perturbations.

Lametti, Smith, Freidin, and Watkins (2018) investigated the specific role of two areas involved in motor control, the cerebellum and the premotor cortex. In this experiment, anodal tDCS was applied during the baseline phase and the training. The auditory perturbation consists of an F1 shift making the training words “bed”, “head” and “dead” sound more like “bad”, “had” or “dad”, respectively. Stimulations over either motor cortex or cerebellum were both found to lead to higher adaptation and/or after-effect than in the sham-stimulated group. Interestingly, stimulation over the cerebellum increased error compensation on F1, while stimulation of the motor cortex also led to adaptation in F2. Adaptation in F2 when altering F1 only has been reported for the front vowel / $\varepsilon$ / with variable size-effects (MacDonald et al., 2011; Rochet-Capellan & Ostry, 2011; Villacorta et al., 2007). Changing F2 in answer to a perturbation of F1 may be a strategy to reach an appropriate phoneme auditory category, as F1 and F2 vary at the same time in the contrast of front vowels. Thus, the cerebellum is suggested to contribute to error correction only, while motor cortex may lead to more general adaptation, possibly related to previously learnt movements.

While rTMS and tDCS can reveal the functional role of a specific brain area, neuronal oscillations as observed in EEG combined with phase coherence analysis may provide insights into the communication between brain areas as proposed by Sengupta and Nasir (2015). Phase coherence over a specific brain area can also represent a measure of this area’s engagement. In this study, a redistribution of phase coherence in specific frequency bands (theta and gamma bands) occurred at the end of the training phase and was related to the amount of speakers’ adaptation. This phenomenon was interpreted as a sign of the establishment of a new feedforward map (i.e. associating an auditory target to a motor gesture that enables the speaker to reach it) together with increased engagement of sensorimotor areas. Sengupta and Nasir (2016) then found that by late training, power in specific frequency bands during speech planning and



speech production was related to whether speakers were adapting to the auditory perturbation or not. Finally, Sato and Shiller (2018) analyzed event-related potentials (ERPs) during adaptation to an increase of F1. They observed that electro-cortical potentials at certain temporal windows (N1, P2) amplitude mirrors adaptation, as larger adaptation magnitude correlated with smaller N1/P2 amplitude. This larger speaking-induced suppression with learning was interpreted as an indication of auditory prediction during speaking.

### 5.7. Speech development

Auditory perturbation is an artificial way to generate speech learning, which otherwise occurs in *natural* situations: learning a new language, as well as during the development of speech. Studying adaptation to perturbations in typical adult speakers might help understand potential mechanisms occurring in these natural situations. It also questions the way children learn speech sounds. Daliri et al. (2018) and Terband et al. (2014) studied adaptation in atypical development, as reported in the “Pathology” section. Shiller and Rochon (2014) investigated the relation between adaptation on perceptual boundaries in children, as reported in the “*Perceptual and phonological categories*” section. MacDonald, Johnson, Forsythe, Plante, and Munhall (2012) and Terband and Van Brenk (2015) focused on adaptation in typically developing children at different ages. Terband and Van Brenk (2015) found greater adaptation in 4 to 9-year-old children than in adults, although the magnitude of adaptation did not correlate with age in the group of children, and the proportion of children exhibiting a consistent compensatory response was lower than in adults. MacDonald et al. (2012) showed that 4-year-old children adapted to a sustained perturbation with a similar magnitude of adaptation as adults, whereas 2-year-old toddlers did not adapt at all. This could suggest that toddlers ignore their own auditory feedback to focus on external stimulation or have an immature feedforward control. According to Messum and Howard (2012), this observation contradicts the widely held view that children learn speech sounds by imitation, which would require them to listen to what they produce and try to make it match what they want to imitate. Instead, it supports the idea that a child learns to speak thanks to a tutor: “Mothers reflect (or mirror) what

their children say, but such imitation generally takes the form of reformulation into well-formed sounds of the ambient language, rather than simple mimicry” (Messum & Howard, 2012, p. 160). Thus, plasticity observed in adults in the situation of adaptation to auditory perturbations may be different in nature to what occurs in the early speech development.

### 5.8. Surface effects and speakers’ characteristics

Other effects related to speakers or context, like the characteristics of the prompt during the adaptation procedure, may influence speech adaptation. Alsius, Mitsuya, Latif, and Munhall (2017) investigated the influence of the stimulus used to prompt the training word “head” by contrasting visual and auditory modalities as well as linguistic vs non-linguistic prompts. No effect of the sensory modality was found on the magnitude of adaptation but linguistic prompts (“head” as a spoken or written word) were found to induce more adaptation than non-linguistic prompts (a cross or a tune). Similarly Sato and Shiller, (2018) found no difference in the magnitude of adaptation between visual and auditory modalities. In addition, Caudrelier et al. (2018) investigated whether naming a picture or reading a word aloud would make a difference in adaptation and in transfer. Although no effect was found in the adaptation response, the pattern of generalization was influenced by the prompt used during the transfer phase, regardless of the training prompt, hinting at possible surface effects.

With regards to speakers’ abilities, Martin et al. (2018) found no correlation between general executive control and adaptation magnitude. In a preliminary study, Dimov, Katseff, and Johnson, (2012) investigated the influence of speakers’ characteristics including some social and personal aspects. In particular, less empowered subjects were found to adapt more than more empowered ones. Finally, Munhall et al. (2009) reported equivalent adaptation in naïve speakers and in speakers who were informed of the shift and who were asked to compensate or not. These results suggest that auditory-motor recalibration is at least in part an automatic process. More work is required to better understand the complexity of adaptive profiles that might be determined by numerous factors, as discussed in the next section.

## 6. Research outlook on formant perturbations

In this section, we identify some perspectives for future studies in adaptation to formant perturbations, in relation to methodological aspects as well as to some of the reviewed research questions.

### 6.1. Toward standards to investigate and report adaptation to formant perturbations

Various interests have motivated adaptation to formant perturbations studies in various teams. This induced the use of different methods to alter formants but also different procedures and analyses. These methodological differences often make studies difficult to compare directly. Therefore some standards should be developed, in particular to facilitate meta-analyses of formants perturbations studies, at least with regards to the way to report the methods and the results. Munhall, Purcell and collaborators studies are very interesting in this regard, as they have involved a significant number of speakers and have used similar methods to alter formants to run the adaptation and to analyze the data. A number of questions should be taken into consideration when designing and reporting studies. Some of them may also require further methodological studies, in line with Munhall and collaborators work. For instance:

- Should the participants be only females or males? What is the effect of mixing vs. not mixing gender on adaptation?
- This first question could be crossed with the effect of the type of perturbation: should the perturbation be absolute vs. relative, formant values being clearly different across gender? What is the effect of shifting only F1 vs. F1 and F2 in opposite direction?
- Whether participants are monolingual or multilingual should be controlled and reported, and as far as possible kept available for meta-analysis. Indeed, adaptation seems influenced by perceptual categories, which are related to phonological systems of languages. One of the best ways to address the question would be to be able to compare large datasets recorded around the world in the different research topics.
- What is the real effect of bone conduction on adaptation? This question has not been addressed systematically, although it has been considered in the conception of apparatuses to shift formants. Most studies used

quite high sound intensity of feedback and/or mixed the signal with noise. The effect of the feedback level, the signal to noise ratio as well as the type of noise on adaptation were not systematically reported.

- What is the real effect of the perturbation on the signal heard by speakers? This question is rarely investigated in papers, while the obtained perturbation can be far from the expected one (Mitsuya et al., 2015). In particular, when using existing packages such as Audapter, delay in feedback should be checked, as it could depend on the properties of the OS and computer hardware. The evaluation of formants provided by the tool, especially when applying unusual shifts, should be verified as there is no guarantee that the system will be able to track and shift the formants in the expected way. This is true for all the systems and could be easily verified by comparing the obtained formant values with corresponding spectrograms or with values assessed by an independent formant assessment software. This approach was used in Reilly and Pettibone (2017).
- Due to the high variability in adaptation magnitude between participants, apparent differences on some parameters of adaptation between conditions are often found to be non-significant. Some effects and, in particular, surface effects such as visual vs. audio prompts might exist but may require testing a large number of speakers to reach significance. This could also be the case for effects related to the direction of the perturbation or to the number of trials during the hold phase as well as to the way the perturbation is introduced. At the very least, non-significant results between different groups of speakers should be interpreted carefully, in relation to this large variability.

These examples suggest that methodological aspects should be directly addressed and clearly reported to help teams working in the field share standards and enable the constitution of large databases. Large between-subjects variability suggests that adaptation to formant perturbations is a complex phenomenon, influenced by different factors. Multifactorial analyses such as introduced in Dimov et al. (2012) could be run on large datasets, but this requires – at the very least – recording of systematic information about the participants and reporting clear information about the perturbation and its real effect.

## 6.2. Topics which will benefit from further investigation

Due to the broad range of research topics addressed by formant perturbation studies, more studies are still required to reproduce or better understand some results. This is particularly the case for the effect of adaptation on categorical perception, as results between studies have been sometimes inconsistent. Only a few studies were published on the effect of adaptation to formant perturbations on categorical perception of speech (Lametti et al., 2014; Schuerman et al., 2015; Schuerman et al., 2017; Schuerman, Nagarajan, et al., 2017) with some inconsistent findings between Lametti et al. (2014) and Schuerman et al. (2017). The two studies were run with speakers of different languages (English vs. Dutch) and with different types of continua for the perceptual test (words vs. vowels). It would be useful to gain more awareness of other attempts with non-significant or inconsistent profiles of perceptual changes following adaptation if any exist. This will avoid a publication bias towards significant-only results that seems to be a sensitive topic for this research question, in particular as the effects of speech production on changes of categorical boundaries may be sensitive to numerous variables, including the number of speakers, their gender, regional accent, languages skills etc. Replication is also required as the involvement of the motor system in perception is an important challenge for speech research more generally.

Investigating the development of feedback and feedforward control systems and their potential interaction in typically developing children is also an important topic to further develop using formant perturbations paradigm. Moreover, using compensation to unpredictable perturbations in conjunction with sustained perturbations in atypical speakers may shed light on the root causes of some pathologies affecting speech production. For instance, van den Bunt et al. (2017) provides a rather convincing explanation about the sensorimotor bases of dyslexia, which could be further investigated in children. As adaptation has been shown to interact with phoneme categories, it allows investigating the development of phonological categories in both typical children and children with phonological disorders.

An important topic also under-investigated so far is the influence of extraneous factors (i.e. not directly related to language or speech) on

auditory-motor adaptation. First results by Munhall et al. (2009) suggested that the magnitude of the compensation is relatively independent of the awareness of the experimental aim and that speakers compensate even when asked not to compensate. This suggests that adaptation is quite independent from higher cognitive functions such as attention. Martin et al. (2018) also found no significant contribution of general executive control skills on adaptation. However, the preliminary work by Dimov et al. (2012) suggests that variables related to speakers' social status may play a role. Further investigations linking working memory abilities, attention levels etc. to formant adaptation will help tackle the mainstream issue of the link between cognitive and sensorimotor functions. This topic, as a number of others, has already been investigated in adaptation or compensation to F0 perturbation (Guo et al., 2017; Hu et al., 2015; Scheerer, Tumber, & Jones, 2015). Last but not least, results by MacDonald et al. (2012) showing that toddlers do not adapt and the associated discussion of this result by Messum and Howard (2012) suggest that the communicative context may also influence adaptation. The question was investigated in birds by Sakata and Brainard (2009) suggesting larger adaptation when the song is produced in presence of another bird but also in humans with other type of perturbations such as speech in noise (Garnier et al., 2010). Social context might thus be relevant to question the real nature of speech targets.

An important topic not developed in this chapter is a systematic analysis of the results of formant perturbation studies in relation to current models of speech production. A joint analysis with the results of other auditory and somatosensory perturbation studies could improve our understanding of feedback and feedforward controls.

Finally, as it is relevant to the link between learning and memory, we would like to emphasize that transfer of adaptation was under-studied so far, despite its potential to bring insight into the nature of speech representations. As already introduced in Houde and Jordan (1998), transfer of learning is an empirical tool to question the nature of speech production units. This approach should be better connected to the equivalent approach developed for perceptual learning (e.g. Chambers et al., 2010). As noted by Cai et al. (2010) patterns of transfer question the way models of speech production represent sensorimotor mapping: both significant

generalization effect, as well as gradient effects should be explained. These models should also be adapted to integrate results from transfer or multiple adaptation studies suggesting that the mapping between auditory and articulatory domains could occur at different linguistic levels and be related in some way to the training word. But more generally, adaptation might be related to the episode of learning, as also discussed in Houde and Jordan (2002) when explaining the long-term effects of adaptation by implicit memory. We strongly believe that understanding the link between sensorimotor learning and memory would be a fruitful path towards understanding of embodied cognition and the links between language and speech. In any case, identifying the condition of specificity vs. generalization of adaptation will clearly contribute to the debate on the nature of speech production representation and to the debate on the nature of internal models and their relation to sensorimotor memories.

## 7. Conclusion

Twenty years ago, Houde and Jordan introduced formant perturbations in auditory feedback as a new paradigm to explore speech production. This seminal study is cited by papers in various domains: speech production and perception in general, studies using other kind of perturbations related to speech (e.g. pitch alteration, vocal tract perturbation), motor control as well as vocalizations in animals. Moreover, it has inspired a whole research field which is still in expansion. In this review, we scanned all studies citing Houde and Jordan (1998, 2002) and selected 77 articles focused on formant perturbations. The perturbation systems designed for this purpose are reported and described in the review. The main research topics addressed in these studies are also explained, along with their main findings.

The formant perturbation paradigm proved to be insightful in exploring the relationship between speech production and perception. First, the observation of responses to auditory perturbations has shed light on the role of auditory feedback in speech production, and the mechanisms that control it. Experimental findings have been incorporated in speech production models, although some results still need to be modeled. Altering both auditory and somatosensory feedback showed that both modalities

are integrated in the control of speech in a manner that may be specific to the speaker and/or to the task (e.g. the vowel to produce). Associating perceptual categorization tasks and training with formant perturbations revealed a close relationship and mutual influence between speech motor control and phonological categories, mediated by categorical perception.

This relationship between motor control and linguistic units (e.g. phonemes or syllables) has also been explored by observing the generalization of auditory-motor adaptation. Generalization, or transfer, has been observed from a vowel to the same vowel in different words, suggesting the existence of an underlying phoneme representation. While transfer may occur from one vowel to another, supporting the idea of broad generalization in speech learning, the magnitude of transfer seems to depend on some similarity relationship between the training and the transfer utterances. Moreover, simultaneous adaptation to opposite perturbations has been observed in two different vowels and even in the same vowel in different words. This apparent contradiction may represent a challenge for speech production models, as it requires much flexibility in the translation of auditory goals into articulatory gestures, and questions the nature of mental representations interfacing with speech articulation.

Studies in cognitive neurosciences have pinpointed neural correlates of sensory integration and motor control in speech production, in terms of brain regions as well as communication networks and frequency bands. While studying patients with cerebellar degeneration also contributes to this purpose, research in other pathologies, including stuttering, Parkinson's disease, dyslexia, developmental speech disorders, and some autism subtypes, have benefited from formant perturbation experiments in understanding of the main causes and mechanisms underlying these specific disorders. Finally, studying compensation and adaptation in children gives insights in the development of sensorimotor processes at stake in speech production. Effects of communicative situation or social context may also be explored, as it has proven influential in some speech motor control characteristics in adults. Further investigations in children in various communicative contexts could eventually shed light on one of the most intriguing questions in our research field: how does a child learn to speak?



Beyond these core topics associated with Houde and Jordan's paradigm, other questions have emerged in relation to speakers' cognitive functions and social characteristics, as well as learning context. The prompt has been suggested to influence adaptation and transfer pattern. Moreover, Houde and Jordan had already noticed that adaptation was still there when speakers were tested one month later with normal feedback. This observation may suggest that learning is to some extent specific to the context in which it occurs, the testing room for instance. This is consistent with multiple-trace memory models or exemplar-based views (Goldinger, 1998; Hintzman, 1986), according to which each event is recorded in the brain in the form of a trace combining multiples elements from sensory inputs. Being confronted with one of these elements may activate all the traces containing it, and therefore the other elements associated with it. Thus, the specific context of the testing room may reactivate the adaptation that had washed out in other contexts. Investigating retention of adaptation in various time ranges and contexts may pave the way to fruitful research exploring the relationship between speech, learning and memory.

## Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007–2013 Grant Agreement no. 339152).

## References

- Aichert, I., & Ziegler, W. (2013). Segments and syllables in the treatment of apraxia of speech: An investigation of learning and transfer effects. *Aphasiology*, 27(10), 1180–1199.
- Alsius, A., Mitsuya, T., Latif, N., & Munhall, K. G. (2017). Linguistic initiation signals increase auditory feedback error correction. *The Journal of the Acoustical Society of America*, 142(2), 838–845.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, 105(4), 678–723.
- Berry, J.J., Jaeger, I.V., Wiedenhoef, M., Bernal, B.A., & Johnson, M.T. (2014). Consonant context effects on vowel sensorimotor adaptation.

- In: ISCA (eds.): *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, (pp. 2006–2010).
- Berry, J.J., North, C., & Johnson, M.T. (2014). Sensorimotor adaptation of speech using real-time articulatory resynthesis. In *IEEE Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 3196–3200).
- Berry, J.J., North, C., Meyers, B., & Johnson, M.T. (2013). Speech sensorimotor learning through a virtual vocal tract. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060099). ASA.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, 8, 208.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2015). Extrinsic talker normalization alters self-perception during speech. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (Eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Bourguignon, N.J., Baum, S.R., & Shiller, D.M. (2016). Please say what this word is—Vowel-extrinsic normalization in the sensorimotor control of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 42(7), 1039–1047.
- Brainard, M.S., & Doupe, A.J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, 1(1), 31–40.
- Brumberg, J.S., Krusienski, D.J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A.L., & Schalk, G. (2016). Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *Plos One*, 11(11), e0166872.
- Burnett, T.A., & Larson, C.R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *The Journal of the Acoustical Society of America*, 112(3), 1058–1063.
- Cai, S., Beal, D.S., Ghosh, S.S., Tiede, M.K., Guenther, F.H., & Perkell, J.S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation. *Plos One*, 7(7), e41830.

- Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong/iau/. *Proceedings of the 8th International Seminar on Speech Production (ISSP)*, (pp 65–68).
- Cai, S., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong/iau/and its pattern of generalization. *The Journal of the Acoustical Society of America*, 128(4), 2033–2048.
- Cai, S., Ghosh, S.S., Guenther, F.H., & Perkell, J.S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience*, 31(45), 16483–16490.
- Carbonnel, S., Charnallet, A., & Moreaud, O. (2010). Organisation des connaissances sémantiques: des modèles classiques aux modèles non abstraits. *Revue de Neuropsychologie*, 2(1), 22–30.
- Cassery, E.D. (2015). Effects of real-time cochlear implant simulation on speech production. *The Journal of the Acoustical Society of America*, 137(5), 2791–2800.
- Caudrelier, T., Perrier, P., Schwartz, J., Rochet-Capellan, A. (2018) Picture naming or word reading: Does the modality affect speech motor adaptation and its transfer? *Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH 2018)*, (pp. 956–960).
- Caudrelier, T., Perrier, P., Schwartz, J.-L., & Rochet-Capellan, A. (2016). Does auditory-motor learning of speech transfer from the CV syllable to the CVCV word? *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)*, (pp. 2095–2099).
- Caudrelier, T., Schwartz, J.-L., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does it tell us about speech production units? *Journal of Speech, Language, and Hearing Research*, 61(7), 1613–1625.
- Chambers, K.E., Onishi, K.H., & Fisher, C. (2010). A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 821–828.

- Chon, H., Kraft, S.J., Zhang, J., Loucks, T., & Ambrose, N.G. (2013). Individual variability in delayed auditory feedback effects on speech fluency and rate in normally fluent adults. *Journal of Speech, Language, and Hearing Research*, 56(2), 489–504.
- Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human brain mapping*, 9(4), 183–191.
- Daliri, A., Wieland, E.A., Cai, S., Guenther, F.H., & Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental science*, 21(2), e12521.
- de Bruijn, M.J., ten Bosch, L., Kuik, D.J., Witte, B.I., Langendijk, J.A., Leemans, C.R., & Verdonck-de Leeuw, I.M. (2012). Acoustic-phonetic and artificial neural network feature analysis to assess speech quality of stop consonants produced by patients treated for oral or oropharyngeal cancer. *Speech Communication*, 54(5), 632–640.
- Demopoulos, C., Kothare, H., Mizuiri, D., Henderson-Sabes, J., Fregeau, B., Tjernagel, J., Houde, J.F., Sherr, E.H., & Nagarajan, S. S. (2018). Abnormal speech motor control in individuals with 16p11.2 deletions. *Scientific Reports*, 8(1), 1274.
- Deroche, M.L., Nguyen, D., & Gracco, V.L. (2017). Modulation of speech motor learning with transcranial direct current stimulation of the inferior parietal lobe. *Frontiers in Integrative Neuroscience*, 11, 35.
- Dimov, S., Katseff, S., & Johnson, K. (2012). Social and personality variables in compensation for altered auditory feedback. *UC Berkeley PhonLab Annual Report*, (6)6, pp. 259–282.
- Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22(1), 567–631.
- Eckey, A., & MacDonald, E. (2015). Compensations of F0 and formant frequencies in a real-time pitch-perturbation paradigm. *Fortschritte der Akustik DAGA'15*, (pp. 1444–1447).
- Eliades, S.J., & Miller, C.T. (2017). Marmoset vocal communication: Behavior and neurobiology. *Developmental Neurobiology*, 77(3), 286–299.

- Feng, Y., Gracco, V.L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, 106(2), 667–679.
- Frank, A.F. (2011). *Integrating linguistic, motor, and perceptual information in language production*. Doctoral dissertation, University of Rochester.
- Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, 53(3), 588–608.
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Golfinopoulos, E., Tourville, J.A., & Guenther, F.H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*, 52(3), 862–874.
- Grimme, B., Fuchs, S., Perrier, P., & Schöner, G. (2011). Limb versus speech motor control: A conceptual review. *Motor Control*, 15(1), 5–33.
- Guo, Z., Wu, X., Li, W., Jones, J.A., Yan, N., Sheft, S., Liu, P., & Liu, H. (2017). Top-down modulation of auditory-motor integration during speech production: The role of working memory. *Journal of Neuroscience*, 37(43), 10323–10333.
- Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, 213(5), 84–97.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1), 67–99.
- Hintzman, D. L. (1986). «Schema abstraction» in a multiple-trace memory model. *Psychological Review*, 93(4), 411–428.
- Houde, J.F. (1997). *Sensorimotor adaptation in speech production*. Doctoral dissertation, MIT.
- Houde, J.F., & Chang, E.F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, 33, 174–181.
- Houde, J.F., & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216.

- Houde, J.F., & Jordan, M.I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310.
- Hu, H., Liu, Y., Guo, Z., Li, W., Liu, P., Chen, S., & Liu, H. (2015). Attention modulates cortical processing of pitch feedback errors in voice control. *Scientific Reports*, 5, 7812.
- Hubl, D., Schneider, R. C., Kottlow, M., Kindler, J., Strik, W., Dierks, T., & Koenig, T. (2014). Agency and ownership are independent components of ‘sensing the self’ in the auditory-verbal domain. *Brain topography*, 27(5), 672–682.
- Jones, J.A., & Munhall, K.G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251.
- Jones, J.A., & Munhall, K.G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, 113(1), 532–543.
- Katseff, S., & Houde, J. (2008). Partial compensation in speech adaptation. *UC Berkeley Phonology Lab Annual Reports*, 4(4), (pp. 445–461).
- Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A trade-off with somatosensory feedback? *Language and Speech*, 55(2), 295–308.
- Kelso, J.S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C.A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812–832.
- Klein, E., Brunner, J., & Hoole, P. (2019). Spatial and temporal variability of corrective speech movements as revealed by vowel formants during sensorimotor learning. In S. Fuchs, J. Cleland & A. Rochet-Capellan (eds.) *Speech production and perception: Learning and memory*. Peter Lang Publisher (*current book*).
- Krakauer, J.W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., & Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS biology*, 4(10), e316.

- Krakauer, J.W., Pine, Z.M., Ghilardi, M.-F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, 20(23), 8916–8924.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, 363, 979–1000.
- Lametti, D.R., Krol, S.A., Shiller, D.M., & Ostry, D.J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning: *Psychological Science*, 25(7), 1325–1336.
- Lametti, D.R., Nasir, S.M., & Ostry, D.J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27), 9351–9358.
- Lametti, D.R., Rochet-Capellan, A., Neufeld, E., Shiller, D.M., & Ostry, D.J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, 34(31), 10339–10346.
- Lametti, D.R., Smith, H.J., Freidin, P.F., & Watkins, K.E. (2018). Cortico-cerebellar networks drive sensorimotor learning in speech. *Journal of Cognitive Neuroscience*, 30(4), 540–551.
- Lane, H., Matthies, M.L., Guenther, F.H., Denny, M., Perkell, J.S., Stockmann, E., Tiede, M., & Zandipour, M. (2007). Effects of short-and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research*, 50(4), 913–927.
- Li, W., Chen, Z., Yan, N., Jones, J. A., Guo, Z., Huang, X., Cheng, S., Liu, P., & Liu, H. (2016). Temporal lobe epilepsy alters auditory-motor integration for voice control. *Scientific reports*, 6, 28909.
- Maas, E., Mailend, M.-L., & Guenther, F.H. (2015). Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language, and Hearing Research*, 58(2), 185–200.
- Maas, E., Robin, D.A., Hula, S.N.A., Freedman, S.E., Wulf, G., Ballard, K.J., & Schmidt, R.A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, 17(3), 277–298.

- MacDonald, E.N., Goldberg, R., & Munhall, K.G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, 127(2), 1059–1068.
- MacDonald, E.N., Johnson, E.K., Forsythe, J., Plante, P., & Munhall, K.G. (2012). Children’s development of self-regulation in speech production. *Current Biology*, 22(2), 113–117.
- MacDonald, E.N., & Munhall, K.G. (2012). A preliminary study of individual responses to real-time pitch and formant perturbations. *The Listening Talker: An interdisciplinary workshop on natural and synthetic modification of speech in response to listening conditions*. 2012, (pp. 32–35).
- MacDonald, E.N., Pile, E., Dajani, H., & Munhall, K.G. (2008). The specificity of adaptation to real-time formant shifting. *Proceedings of the International Seminar on Speech Production*, 2008, (pp. 397–400).
- MacDonald, E.N., Purcell, D.W., & Munhall, K.G. (2011). Probing the independence of formant control using altered auditory feedback. *The Journal of the Acoustical Society of America*, 129(2), 955–965.
- Martin, C.D., Niziolek, C.A., Duñabeitia, J.A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J.F. (2018). Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience*, 12, 91.
- Mattar, A.A., & Ostry, D.J. (2007). Modifiability of generalization in dynamics learning. *Journal of Neurophysiology*, 98(6), 3321–3329.
- Max, L., & Maffett, D.G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience letters*, 591, 25–29.
- Max, L., Wallace, M.E., & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. *Proceedings of the 15th International Congress of Phonetic Sciences*, Futurgraphic Barcelona, Spain, (pp. 1053–1056).
- Ménard, L., Perrier, P., & Aubin, J. (2016). Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. *The Journal of the Acoustical Society of America*, 139(5), 2514–2531.
- Messum, P., & Howard, I.S. (2012). Speech development: Toddlers don’t mind getting it wrong. *Current Biology*, 22(5), R160–R161.



- Mitsuya, T., MacDonald, E.N., & Munhall, K.G. (2014). Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, 135(5), 2986–2994.
- Mitsuya, T., MacDonald, E.N., Munhall, K.G., & Purcell, D.W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, 138(1), 413–424.
- Mitsuya, T., MacDonald, E.N., Purcell, D.W., & Munhall, K.G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America*, 130(5), 2978–2986.
- Mitsuya, T., Munhall, K.G., & Purcell, D.W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *The Journal of the Acoustical Society of America*, 141(4), 2758–2767.
- Mitsuya, T., & Purcell, D.W. (2016). Occlusion effect on compensatory formant production and voice amplitude in response to real-time perturbation. *The Journal of the Acoustical Society of America*, 140(6), 4017–4026.
- Mitsuya, T., Samson, F., Ménard, L., & Munhall, K.G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, 133(5), 2993–3003.
- Mollaei, F., Shiller, D.M., Baum, S.R., & Gracco, V.L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, 1646, 269–277.
- Mollaei, F., Shiller, D.M., & Gracco, V.L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, 28(12), 1668–1674.
- Munhall, K.G., MacDonald, E.N., Byrne, S.K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, 125(1), 384–390.
- Neufeld, C., Purcell, D., & Van Lieshout, P. (2013). Articulatory compensation to second formant perturbations. *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060097). ASA.
- Niziolek, C.A., & Guenther, F.H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, 33(29), 12090–12098.

- Osu, R., Hirai, S., Yoshioka, T., & Kawato, M. (2004). Random presentation enables subjects to adapt to two opposing forces on the hand. *Nature Neuroscience*, 7(2), 111–112.
- Paethorpe, S., Watson, C.I., & Barker, R. (2003). Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss. *The Journal of the Acoustical Society of America*, 114(2), 1055–1068.
- Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393.
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R.B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *Journal of Neuroscience*, 37(38), 9249–9258.
- Patri, J.-F., Perrier, P., Schwartz, J.-L., & Diard, J. (2018). What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. *PLoS Computational Biology*, 14(1), e1005942.
- Parkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4), 2338–2344.
- Perrier, P. (2012). Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control. In S. Fuchs, M. Weirich, D. Pape & P. Perrier (eds.). *Speech Planning and Dynamics*, Peter Lang Publishers, pp.191–238.
- Pfordresher, P.Q., & Palmer, C. (2006). Effects of hearing the past, present, or future during music performance. *Attention, Perception, & Psychophysics*, 68(3), 362–376.
- Pile, E.J.S., Dajani, H.R., Purcell, D.W., & Munhall, K.G. (2007). Talking under conditions of altered auditory feedback: does adaptation of one vowel generalize to other vowels. *Proceedings of the International Congress of Phonetic Sciences* (pp. 645–648).
- Purcell, D.W., & Munhall, K.G. (2008). Weighting of auditory feedback across the English vowel space. *Proceedings of the International Seminar on Speech Production* (Vol. 8, p. 389–392).

- Purcell, D.W., & Munhall, K.G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2), 966–977.
- Purcell, D.W., & Munhall, K.G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297.
- Reilly, K.J., & Dougherty, K.E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*, 134(2), 1314–1323.
- Reilly, K.J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology*, 118(5), 2925–2934.
- Rochet-Capellan, A., & Ostry, D.J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *Journal of Neuroscience*, 31(7), 2657–2662.
- Rochet-Capellan, A., Richer, L., & Ostry, D.J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717.
- Sakata, J.T., & Brainard, M.S. (2009). Social context rapidly modulates the influence of auditory feedback on avian vocal motor control. *Journal of Neurophysiology*, 102(4), 2485–2497.
- Sato, M., & Shiller, D.M. (2018). Auditory prediction during speaking and listening. *Brain and Language*, 187, 92–103.
- Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., & Gracco, V.L. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, 227(2), 275–288.
- Scheerer, N.E., Tumber, A.K., & Jones, J.A. (2015). Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *Journal of Neurophysiology*, 115(2), 826–832.
- Schuerman, W.L., Meyer, A.S., & McQueen, J.M. (2017). Mapping the speech code: Cortical responses linking the perception and production of vowels. *Frontiers in Human Neuroscience*, 11, 161.
- Schuerman, W.L., Nagarajan, S., & Houde, J. (2015). Changes in consonant perception driven by adaptation of vowel production to

- altered auditory feedback. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Schuerman, W.L., Nagarajan, S., McQueen, J.M., & Houde, J. (2017). Sensorimotor adaptation affects perceptual compensation for coarticulation. *The Journal of the Acoustical Society of America*, 141(4), 2693–2704.
- Sengupta, R., & Nasir, S.M. (2015). Redistribution of neural phase coherence reflects establishment of feedforward map in speech motor adaptation. *Journal of Neurophysiology*, 113(7), 2471–2479.
- Sengupta, R., & Nasir, S.M. (2016). The predictive roles of neural oscillations in speech motor adaptability. *Journal of Neurophysiology*, 115(5), 2519–2528.
- Sengupta, R., Shah, S., Gore, K., Loucks, T., & Nasir, S.M. (2016). Anomaly in neural phase coherence accompanies reduced sensorimotor integration in adults who stutter. *Neuropsychologia*, 93, 242–250.
- Shadmehr, R., & Mussa-Ivaldi, F.A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, 14(5), 3208–3224.
- Shih, T., Suemitsu, A., & Akagi, M. (2011). Influences of transformed auditory feedback with first three formant frequencies. *International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11)*.
- Shiller, D.M., Lametti, D., & Ostry, D.J. (2013). Auditory plasticity and sensorimotor learning in speech production. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060150). ASA.
- Shiller, D.M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1308–1315.
- Shiller, D.M., Rvachew, S., & Brosseau-Lapr  , F. (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech-Language Pathology & Audiology*, 34(3), 181–192.
- Shiller, D.M., Sato, M., Gracco, V.L., & Baum, S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103–1113.

- Shum, M., Shiller, D.M., Baum, S.R., & Gracco, V.L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience*, 34(11), 1817–1822.
- Smotherman, M., Zhang, S., & Metzner, W. (2003). A neural basis for auditory feedback control of vocal pitch. *Journal of Neuroscience*, 23(4), 1464–1477.
- Sober, S.J., & Brainard, M.S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience*, 12(7), 927–931.
- Stratton, G.M. (1897). Vision without inversion of the retinal image. *Psychological Review*, 4(4), 341–360.
- Terband, H., & Van Brenk, F. (2015). Compensatory and adaptive responses to real-time formant shifts in adults and children. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith (eds.), *Proceedings of the 18th International Congresses of Phonetic Sciences (ICPhS 2015)*. London: International Phonetic Association.
- Terband, H., Van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, 51, 64–77.
- Thibeault, M., Ménard, L., Baum, S.R., Richard, G., & McFarland, D.H. (2011). Articulatory and acoustic adaptation to palatal perturbation. *The Journal of the Acoustical Society of America*, 129(4), 2112–2120.
- Tourville, J.A., Cai, S., & Guenther, F.H. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060180). ASA.
- Tourville, J.A., Reilly, K.J., & Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443.
- Tremblay, S., Houle, G., & Ostry, D.J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, 28(10), 2426–2434.
- Tremblay, S., Shiller, D.M., & Ostry, D.J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866–869.
- Trudeau-Fisette, P., Tiede, M., & Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: Acoustic and articulatory data. *Plos One*, 12(7), e0180300.

- van den Bunt, M.R., Groen, M.A., Ito, T., Francisco, A.A., Gracco, V.L., Pugh, K.R., & Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit. *Journal of Speech, Language, and Hearing Research*, 60(3), 654–667.
- Van Vugt, F. T., & Ostry, D. J. (2018). The structure and acquisition of sensorimotor maps. *Journal of Cognitive Neuroscience*, 30(3), 290–306.
- Vaughn, C., & Nasir, S.M. (2015). Precise feedback control underlies sensorimotor learning in speech. *Journal of Neurophysiology*, 113(3), 950–955.
- Villacorta, V.M., Perkell, J.S., & Guenther, F.H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319.
- Wei, K., Yan, X., Kong, G., Yin, C., Zhang, F., Wang, Q., & Kording, K.P. (2014). Computer use changes generalization of movement learning. *Current Biology*, 24(1), 82–85.
- Wong, S.M., Domangue, R.J., Fels, S., & Ludlow, C.L. (2017). Evidence that an internal schema adapts swallowing to upper airway requirements. *The Journal of Physiology*, 595(5), 1793–1814.
- Yates, A.J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3), 213–232.
- Zheng, Z.Z., Vicente-Grabovetsky, A., MacDonald, E.N., Munhall, K.G., Cusack, R., & Johnsrude, I.S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *Journal of Neuroscience*, 33(10), 4339–4348.