Speech-breathing-limb interaction and coordination during the production andthe memorization of verbal information

Hélène Serré

To cite this version:

Hélène Serré. Speech-breathing-limb interaction and coordination during the production andthe memorization of verbal information. Cognitive Sciences. Université Grenoble Alpes [2020-..], 2022. English. ⟨NNT : 2022GRALS032⟩. ⟨tel-04003752⟩

HAL Id: tel-04003752
https://theses.hal.science/tel-04003752v1

Submitted on 24 Feb 2023

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.

Autorisation HAL

HAL Authorization

THÈSE

Pour obtenir le grade de

DOCTEURE DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : PCN - Sciences cognitives, psychologie et neurocognition
Unité de recherche : Grenoble Images Parole Signal Automatique

Interaction et coordination parole-respiration-membres pendant la production et la mémorisation d'informations verbales

Speech-breathing-limb interaction and coordination during the production and the memorization of verbal information

Présentée par :

Hélène SERRÉ

Direction de thèse :

AMÉLIE ROCHET-CAPELLAN Chargée de recherche HDR, CNRS DELEGATION ALPES	Directrice de thèse
MARION DOHEN Maître de conférences, GRENOBLE-INP	Co-encadrante de thèse
SUSANNE FUCHS Directrice de recherche, LEIBNIZ-CENTRE GENERAL LINGUISTICS	Co-encadrante de thèse

Rapporteurs :

BENOÎT BARDY
Professeur des Universités, UNIVERSITE DE MONTPELLIER
JÜRGEN TROUVAIN
Chargé de recherche, UNIVERSITÄT DES SAARLANDES

Thèse soutenue publiquement le **20 octobre 2022**, devant le jury composé de :

AMÉLIE ROCHET-CAPELLAN Chargée de recherche HDR, CNRS DELEGATION ALPES	Directrice de thèse
BENOÎT BARDY Professeur des Universités, UNIVERSITE DE MONTPELLIER	Rapporteur
JÜRGEN TROUVAIN Chargé de recherche, UNIVERSITÄT DES SAARLANDES	Rapporteur
CÉCILE FOUGERON Directrice de recherche, CNRS ILE-DE- FRANCE VILLEJUIF	Examinatrice
WIM POUW Chargé de recherche, MAX PLANCK INSTITUTE	Examinateur
SONIA KANDEL Professeure des Universités, UNIVERSITE GRENOBLE ALPES	Présidente

Invitées :

SUSANNE FUCHS
Directrice de recherche, LEIBNIZ-CENTRE GENERAL LINGUISTICS
MARION DOHEN
Maître de conférences, GRENOBLE INP

Financement

Cette thèse a été financée par l'Agence Nationale de la Recherche (ANR) sous la subvention ANR-17-FRAL-0005 en partenariat avec la Fondation Allemande pour la Recherche (DFG) sous la subvention DFG FU791/8-1 dans le cadre du projet SALAMMBO (Spoken Language in Motions: Learning and Adaptation of Speech Communication in the context of Body Motions).

Abstract

A growing body of behavioral and neurophysiological research shows the implications of limb movements in cognitive activities. For instance, manual gestures facilitate learning abstract knowledge such as mathematical concepts (Perry et al., 1988; Goldin-Meadow and Wagner, 2005); language understanding involves motor areas in the brain (e.g., Pulvermüller, 2005) and motor systems interact with language comprehension (Glenberg and Kaschak, 2002). In the grounded cognition framework, these results are integrated as evidence that most of our cognitive activities are a 'reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind' (Barsalou, 2010 p. 618). These approaches change the perspectives on cognitive abilities, and language in particular (Lakoff, 2012). To improve our understanding of spoken language, research should embrace body movements and the situations in which they occur. In this context, limb motion could be considered as a potential body-ground for spoken language. As such, and as suggested by previous work, limb motion may specifically influence online spoken language and may play a significant role in the recall of verbal information. These mutual influences may occur at least partially through a third body-actor: the respiratory system, an inherent and strong constraint shared by the speech and limb motor systems. Breathing is actually a core topic of limb motion, speech and more recently cognition research. These links between motor control of the limbs, physiology and spoken language create new challenges both at methodological and theoretical levels. In this broad framework, this thesis aims at assessing the body-ground of spoken language by investigating the interactions between speech, breathing and limb movements through methodologies coming from the research fields of physiology, phonetics and motor control. It first provides an overview of embodied cognition and previous works linking speech, arm movements and breathing. It then details the recording of the dataset and of the analysis methods used to address the relation between speech, limb motion and breathing. Narrative speech produced by native speakers of German was collected in different conditions of limb movements. Speech acoustics as well as breathing and limb motions were recorded synchronously on different sessions consisting in remembering short movies in four different conditions: arms free, arms blocked, biking with the arms and biking with the legs. Participants were invited to repeat the task on three different days (day 1, the next day, and ten days after) over two weeks. In this thesis we investigate: (1) the effects of limb movement conditions on speech through speech breathing: speech structure within the breath group, and speaker-specific character of speech breathing, across the different limb movement conditions and over days are analysed; (2) the effects of speech on limbs motion; (3) the coupling between limb motion and breathing over time and according to limb movement condition. Our results show changes in speech parameters over days only but poor impact of limb motion on speech. Lower limb motion yet seems to have a greater impact on speech intensity than arm mo-

tion (Serré et al.). Speech impacts cycling motion by increasing the variability of the cycle duration and disturbing motor respiratory coupling. Speech breathing also appears to be speaker-specific, this specificity resists over days and limb movements (Serré et al., 2021). These results are discussed regarding the strengths and limits of methodological aspects and in the framework of the current theories on joint control of speech and limb movements, coming from fields such as experimental psychology (for hand gestures) and motor control (for cycling motion). We also highlighted the key role of breathing and replaced our data driven approach in the current scientific stakes.

Remerciements

La thèse est un voyage, un voyage en train à grande vitesse, avec son chauffeur peu expérimenté, ses aiguilleurs, ses contrôleurs, ses grandes fenêtres d'où contempler le paysage scientifique, et ses aléas. Ce train a parfois pris du retard en raison d'incidents sur les voies, de défaillance des caténaires, mais a toujours redémarré, grâce à ses aiguilleurs aiguisés et aux équipes de maintenance qui ont su réparer les dommages avec beaucoup d'amour. Le bar en voiture 4 et les toilettes ont dû fermer temporairement, mais les passagers de ce train ont fait preuve d'une patience sans borne face à ces désagréments momentanés. Merci pour leur bienveillance. Ce train n'est peut-être pas arrivé à l'heure mais il est arrivé tout de même, et c'est déjà beaucoup.

Je tiens tout d'abord à remercier l'ANR et la DFG pour avoir financé cette thèse. Je remercie mon comité de suivi individuel, Olivier Perrotin et Marie-Line Bosse, pour avoir veillé au bon déroulement de la thèse avec bienveillance. Je remercie les membres du jury, Benoît Bardy, Jürgen Trouvain, Cécile Fougeron, Wim Pouw et Sonia Kandel, pour avoir pris le temps d'évaluer mon travail avec intérêt.

Je remercie bien sûr mes chères encadrantes, ces femmes scientifiques qui ont toute mon admiration, Amélie Rochet-Capellan, Marion Dohen et Susanne Fuchs, pour leur soutien sans faille, leur humanité, leur écoute. J'ai tant appris à vos côtés, que ce soit de la science ou de la vie. Vous avez façonné pas à pas la chercheuse et la femme que je suis aujourd'hui, et je ne suis pas peu fière du résultat. Merci aussi pour ces liens d'amitié qui se sont créés au fil des ces quatre années. Merci à Susanne pour son accueil à Berlin, et pour le temps passé à faire en sorte que je sois rémunérée sur les derniers mois.

Je remercie Heather, Nadège et Alexandra pour avoir collaboré avec moi, pour nos échanges joyeux et passionnants, aussi bien scientifiques que personnels.

Je remercie les membres de l'équipe PCMD et du département parole et cognition (Maëva, Thomas, Olivier, Frédéric B., Frédéric E., Jean-Luc, Pierre, Pascal, Nathalie, Coriandre, Sonia, Gérard) pour toujours s'être enquis de mon état et de l'avancée de ma thèse, pour leurs précieux conseils, et pour leur convivialité. Les réunions d'équipe et les journées du DPC ont toujours été de chouettes moments scientifiques et humains.

Je remercie les personnes du service informatique et du service plateforme (Mikael, Christian, Olivier, Silvain), pour m'avoir apporté leur aide toutes les fois où j'en ai eu besoin.

Je remercie Jörg pour son aide précieuse sur le set-up expérimental.

Je remercie toutes les personnes des ressources humaines (Allison, Virginie, Sophia, Akila, Lucia, Martine, Christine, Caroline) pour leur efficacité administrative, leur disponibilité, leur patience et leur convivialité.

Je remercie Caroline Bazzoli, Frédérique Letué, Vincent Brault et Chantal Bedecarrax pour m'avoir donné l'opportunité d'enseigner, pour avoir pris le temps de m'apporter de précieux conseils pour la suite, et pour leur aide sur mes candidatures ATER.

Je remercie Silvain Gerber pour ces longues heures passées ensembles à analyser mes données souvent coriaces, toujours dans le rire et la bonne humeur. Merci d'avoir pris le temps de me transmettre toutes ces précieuses connaissances. Tu as été une des raisons pour lesquelles je me levais avec entrain le matin. Travailler avec toi a été un très grand plaisir, de ce plaisir j'en ai tiré un ami, et j'espère croiser d'autre collaborateurs comme toi sur la route.

Je remercie Marc-Antoine Georges pour avoir été le meilleur des co-bureaux, pour toutes ces sucreries ingérées en refaisant le monde, en riant de nos galères de doctorants, et surtout en réfléchissant à la manière dont on pourrait gagner beaucoup d'argent vite et bien. Il ne me reste plus qu'à lire le best-seller 'l'autoroute du millionnaire'.

Je remercie dans le même sillon tous les co-bureaux que j'ai eu la chance de cotoyer, Hamid, Loriane, Alexandre, François.

Je remercie Omar, Lou, et Lise pour m'avoir épaulée dans toutes les étapes de la thèse, et avoir su me redonner espoir et motivation dans les creux.

Je remercie Iván, cette personne merveilleuse avec qui j'ai eu la chance de faire un bout de chemin, pour avoir été à mes côtés presque jusqu'au bout, et m'avoir soutenue avec un amour sans borne durant tout ce temps.

Je remercie les doctorants du groupe Dis-9 (Yigit, Yanis, Léa, Roza, Cyril, Charles, Juliette, Olivier, Clément, Fati, et tant d'autres) pour tous les joyeux repas au RU, pour la dynamique avec laquelle on veille les uns sur les autres, pour avoir été bon public à mes débuts dans le rap et le cinéma. J'espère avoir de vos nouvelles !

Je remercie l'équipe du Gipsa-Doc, sur les 4 années, pour avoir été au top sur l'organisation des événements et l'accueil des nouveaux arrivants !

Je remercie mes parents pour leur transmission, leur joie de vivre, leur amour et leur soutien inconditionnels, sans lesquels je n'aurai jamais pu prendre la liberté de faire ce que bon me semble, et traverser ces étapes avec l'assurance d'être rattrappée en cas de chute.

Je remercie les copains qui ne sont pas tout près (Camille, Virgile, Marion, Nico, Gaëtan, Elsa, Jérémie) pour avoir pris régulièrement des nouvelles, pour leur mots doux tout au long du chemin, et pour nos retrouvailles toujours au top.

Je remercie Julie pour son soutien constant et dénué de tout jugement, son réconfort et sa présence essentielle dans les coups durs.

Je remercie tous les copains tout près (Dawood, Jibé et sa bande de joyeux lurons, Lara, Alessio, les copains du théâtre et de la danse) pour leur présence joyeuse et lumineuse dans mon quotidien, présence sans laquelle les moments durs auraient été autrement plus rudes, et pour avoir fait de cette thèse un moment de vie privilégié, un moment de vie qui me fait dire que Grenoble, j'y reviens dès que je peux pour ne plus en repartir ;)

Je remercie Sébastien pour cette parenthèse inattendue dans ma rédaction et ses bons petits plats, qui ont fait la diff, pour citer ce cher Pierre-Louis.

Pour conclure et boucler la boucle, je remercie Ghilès Mostafaoui, la génèse de ce parcours scientifique en construction, la première personne à m'avoir transmis le goût de la recherche, sans qui je n'en serais pas là aujourd'hui.

Contents

List of Figures

List of Tables

Introduction

Speech can be produced and perceived while doing something else. This could be part of the reason why speakers of spoken languages favour the oral channel over the manual one (Corballis, 2009; Arbib, 2012). We can indeed speak while achieving a range of activities. Every day, we speak while doing crafts, cooking, or holding the handlebars of our bikes. And we do so even though speech production puts specific demands on resources such as attention or breathing, which are also involved in the co-occurring activity. In these situations, speech and motion appear as two competing tasks for cognitive or physiological resources. Modular views of cognition and conceptions of human language would argue in that direction (Hauser et al., 2002). As an illustration, the pressure put on breathing by physical effort can make it impossible to speak (Recalde et al., 2002). The interest for the interaction between speech and limb movements could have ended there. But one can also consider the co-occurrence of speech and limb movements as a recurrent human experience. Our brain faces this speech - limb motion duality so early and frequently in the course of language and motor skill acquisition that it seems reasonable to assume close connections between the two systems, in particular with regards to the control of breathing. It is indeed surprising that competition has been the most prominent view of the relation between non-communicative motion and speech while most of the time we manage to speak and move at the same time without even noticing we are actually doing it. Our understanding of speech-limb and breathing cooperation may be limited by the fact that the triadic relation has been poorly investigated as such. In particular, 2-level relationships between speech and limb motion; speech and breathing; and limb motion and breathing are specific research fields that poorly interact with each other. In order to better understand the “body ground” of spoken communication, this thesis aims to investigate the way speech-breathing and limb movements interact in a narrative task and to describe changes in these adaptations over time.

For this purpose, we collected a dataset of narrative speech produced in different conditions of limb movements. Spontaneous speech as well as breathing and limb motions were recorded in different sessions consisting in remembering short movies in four different conditions: hands free, hands blocked, biking with the hands and biking with the feet. Participants were invited to come three times (day 1, the next day, and ten days after) over two weeks. In this way, we were able to study the effect of each context on verbal learning of narrative content as well as the relationship between speech, breathing and limb movements. 25 German speakers were involved in this experiment.

In this manuscript, Chapter 1 will review the effect of body movements on cognition and its relationship with breathing. Chapter 2 will describe the 2 by 2 level relationships between speech, limb movements and breathing, as well as studies assessing the speech-limb-breathing triad. This thesis is at the intersection of several disciplines with an extensive literature. For this reason, the state of the art is not exhaustive, but gives an overall picture of the connection between speech, movement and breathing that motivated the collection of the dataset. Chapter 3 will detail the methods used to collect the data. This corpus was then used to tackle different research questions, addressed through different levels of analysis that constitute the following chapters of the thesis. Chapter 4 will be dedicated to the characterisation of the coordination between speech and limb movements when considered as continuous signals and the stability of this coordination over time. Chapter 5 will introduce breathing in the interaction between speech and limb motion and focus more specifically on disentangling the effect of limb motion vs. speech on breathing control. Finally, these analyses and results will be discussed with regards to strengths and limits of methodological aspects and in the framework of the current theories on joint control of speech and limb movements.

CHAPTER 1

General introduction on the relevance of studying movement for cognition

This first chapter presents developmental, neurocognitive, behavioural and physiological arguments in favor of a distributed view of cognition, and introduces the relevancy of studying breathing at the intersection between spoken language and motion. Speech is characterized by both cognitive and motor activity. Understanding the effects of limb movements on cognition provides insight into the effects that limb movements can have on speech. The objective of this chapter is to present why studying cognition in motion is relevant. The first section presents the theoretical framework in which motion is most likely to be considered as an actor of cognition: embodied cognition. The second section explains the role body motion can play in learning, thinking and language. The last part introduce breathing as a key element of the effect of motion on cognition and spoken language. The literature review in this chapter is not exhaustive, because the topic is interdisciplinary and there is a tremendous amount of publications. The arguments which are developed here are illustrated by the chosen studies.

1.1 Cognition is likely to be based on interactions between the body and the environment

The theoretical framework of embodied cognition consider sensorimotor processes as being part of cognitive processes. The conception of cognition as embodied runs counter to a hierarchical view of the central nervous system controlling its effectors. In this hierarchical view, the body is a mere executor, sending back information via sensory channels, which the central nervous system uses to adapt motor commands (in case of error for example). In the heterarchical vision of embodied cognition, sensorimotor processes are an integral part of cognitive processes, and participate pro-actively in the executive functions involved in memory, learning or language. This vision of cognition enables to overcome the dichotomy of mind versus body.

1.1.1 What is embodied cognition? A broad definition

Embodied cognition is a theoretical background in which the mind, the body and the environment are considered as equal actors of cognition interacting with each other. This theoretical background comes in opposition with computational views of cognition, in which the mindset is the central unit sending commands to the body as a mere executor. To define embodied cognition, Iverson and Thelen write: 'cognition depends crucially on having a body with particular perceptual and motor capabilities and the types of experiences that such a body affords. In other words, cognition is a product of the body and the ways in which it moves through and interacts with the world.' (Iverson and Thelen, 1999, p.31). Shaun Gallagher (2006) defines embodied, embedded, enactive and extended cognition as follows: 'the unit of explanation is not only neuronal processes in the brain but the whole complex of brain, body, environment' (see also his interview at the 2017 Summer Institute on Buddhism and Science: https://www.youtube.com/watch?v=M7ghXdujfLE).

Embodied cognition refers to how the body contributes to the actual cognitive processes (Varela et al., 2016). As Gallagher mentions (2006), in an evolutionary perspective, if we did not have hands, our brain would be really different. The mind is connected and coupled to the body, and has co-evolved with it. A growing body of behavioural studies observe the effect of bodily actions on the mindset. For instance, nodding while listening increases the probability to agree with the heard statements (Wells and Petty, 1980); activating smiling musculature while watching a cartoon increases the rate of how funny the cartoon is judged (Strack et al., 1988). Congruent motor rotations facilitate mental rotations (Wexler et al., 1998), while incongruent motor rotations suppress mental imagery underpinning the performance, leading to an increase in measured task difficulty.

Not only the body but the environment contributes to cognitive processes. Individuals are connected to their physical, cultural, and social environment and this environment offers possibilities for them to act. For instance, while seeing an object intended to be grasped, like a cup, the neural circuit of grasping fires in the motor cortex (Tucker and Ellis, 1998; Ellis and Tucker, 2000). The perception of the environment is shaped by possible actions in this environment. Information from this environment is constantly sampled through various motor routines like occular saccades, or sniffing. Not only does perception guide action, but actions like active sensing also shape perception, creating a dynamic loop between action and perception, and interactions between the subject and its environment.

Even further, the tools available in the environment are also part of cognition. For instance, writing down a to-do list is an extension of memory (Clark and Chalmers, 1998). Currently, smartphones constitutes some of the most sophisticated extensions of the mind: they enable to communicate, find paths, or access a huge bank of data through key-word research on browsers.

1.1.2 The crucial role of mental imagery and motor simulations

This perception-action reciprocity has a crucial role in cognition and mental representations of the world: the ability to represent and manipulate absent or abstract information is achieved through the activation of mental imagery and sensorimotor mechanisms. Mental imagery is shaped and constrained by the physical interactions between the self and the world. In order for a simulation to be (re)activated, it has to be experienced first, and what is not experienced is more difficult to conceive. For instance, because of one's experience of gravity and liquid behaviour such as water, one can easily determine when water will flow out of a glass if one imagines the glass tilting from its upright position. Making the same prediction when the glass tilts upwards from a horizontal position is much more difficult (Schwartz and Black, 1999). In the framework of embodied cognition, most of cognitive activities are conceived as the result of these simulations. For Barsalou (2010), bodily states do play a role in cognition, but are not necessary for cognition to happen. The environment can also be considered as part of cognitive processes. For these two reasons, he uses the term grounded cognition rather than embodied cognition.

1.1.3 The modal representation and storage of common concepts

The concept of grounded cognition lies in simulations, which are 'the reenactment of perceptual, motor and introspective states acquired during experience with the world, body and mind.' (Barsalou, 2010, p.168). Storage in memory depends on the different modalities by which the agent perceives its environment, through mental imagery and motor simulations. For instance, the concept of a cat does not come to mind devoid of sensorimotor sensations, but invokes memories such as smell, touch, sound that one perceived when experiencing an interaction with a cat. One of the main implications is that memory also works through consolidation of sensorimotor traces. Symbols are meaningful only through sensorimotor experience. Neural studies tend to prove the multimodality of memory. Wheeler and colleagues showed that when vividly retrieving pictures and sounds encoded respectively through the visual or auditive modality, specific areas of the sensory cortex respectively related to auditory and vision processes were activated (Wheeler et al., 2000). In a similar vein, to maintain the memories of an absent stimulus in working memory, the neural populations activated in the frontal lobe are different whether the stimulus to maintain is an object, or a location, or human motion (Pasternak and Greenlee, 2005). Dreyer and colleagues (2020) found that when naming a tool, the neural circuits related to the action accomplished with this tool are activated.

Researchers such as Fodor (1983) claimed that language is a specific and encapsulated cognitive 'device' in the brain, present from birth, and independent of other cognitive abilities. This view of language influenced studies that observed language isolated from other cognitive or motor activities. As Dreyer and colleagues explain in their introduction (Dreyer et al., 2020), classic cognitive approaches consider that the circuits processing concepts and semantics are completely independent and isolated from modules of perception and action. This way, the semantic of an object would be apart from the mental imagery and its 3D representation (Caramazza et al., 1990). Such a view of semantics as amodal circuits has been disproved by behavioral and neuroimaging studies (see Dreyer et al., 2020).

1.1.4 Learning in an embodied cognition framework

Cognition being based on simulations and mental imagery means that life experience has a crucial role in cognition. In their paper on embodied learning, Kontra and colleagues (2012) review findings that show how experience can shape thinking 'as a function of a person's unique experience throughout the lifetime' (p.2). They take the study of Casile and Giese (2006) as an example. This study shows that two dancing movements performed with different angles between arms and legs were better visually discriminated by participants trained to perform these movements (being blindfolded) between pre and post-tests than participants not trained to these movements. These results suggest that non-visual motor experience can transfer to visual perception, and support the idea that the acquisition of sensorimotor patterns during an action can be reused by simulation for other tasks.

1.2 Movements are actively involved in learning, thinking and language

Kontra and colleagues (2012) also mention how the role of gestures on cognition and reasoning constitute another evidence of embodied learning at work. How can motion such as gestures, locomotion or mere rhythmic movements shape learning and thinking ? On one hand, gestures and other motions congruent with the ongoing cognitive activity improve learning in certain conditions. At a neural level, semantic processing interact with motor activation. Mere rhythmic movements enhance attention processes. On the other hand, dual task design experiments rather focus on the cognitive interference created by simultaneous non congruent motor activity. These research domains are completely different and apart from each other, but all describe the interactions between motion and cognition.

1.2.1 Producing gestures supports different types of learning

In the framework of this thesis, gestures are defined as bodily movements, especially hand movements. They often go along with or complete speech. The role of gestures on learning and thinking is manifold. Manual gestures can shape reasoning, support cognitive load, or be a marker of the learning process.

1.2.1.1 Gestures are markers of learning

Gestures are not always consistent with ongoing speech. For instance, gestures can mismatch with speech in the discourse of children that are learning either a mathematical concept (Goldin-Meadow and Singer, 2003) or the concept of conservation (Church and Goldin-Meadow, 1986). Mismatches occur when speech and gesture convey different information (not necessarily contradictory) while explaining the same thing. Starting to understand a concept is first reflected in the gestures of children, but not yet in their speech. Perry and colleagues (1998) observed the spontaneous gestures of a group of children between 7 and 10 years old during the acquisition of mathematical concepts such as equivalence. They find that inconsistencies between what children say and what they gesture reflect the transition of knowledge from unknown to acquired, that is, gestures are indicative of the learning process. Goldin-Meadow and Singer (2005) showed that children whose gestures mismatch with speech it accompanies during instructions and pre-tests were more likely to acquire the knowledge correctly than children for whom speech and gestures were consistent or children who did not gesture. According to Goldin-Meadow and Wagner (2005), gestures would reflect what learners cannot yet say clearly, but are beginning to understand.

1.2.1.2 Gestures help learning through supporting implicit knowledge

Gesturing while learning can also enhance the use of implicit knowledge and open new paths of thoughts. Gestures help resolve problems and be more receptive to new instructions and learning (Broaders et al., 2007). Goldin-Meadow and colleagues (2009) compared consistent co-speech gestures to mere hand-waving in a learning task. They found a positive effect of co-verbal gestures on learning, but no effect of hand-waving. The motor or rhythmic aspect of hand waving did not have any positive effect on the task. Gestures can be expression of implicit ideas not verbalised in speech and this implicit knowledge can be exploited by eliciting gestures (Broaders et al., 2007).

1.2.1.3 Gestures help learning through supporting cognitive load

In a similar vein, gestures help encode information units and access them during retrieval. Gesturing either when encoding content (Cook et al., 2010) or when retrieving it (Stevanoni and Salmon, 2005) facilitates access to memory when recalling items or events. When children gesture while recalling a cartoon (Stevanoni and Salmon, 2005), they recall more events and details than those who do not gesture. When adults gesture while memorizing actions graphically presented on cards (Cook et al., 2010), they also recall more events and details than those who do not gesture.

Gestures also release the working memory from its cognitive load, to allocate more cognitive resources to another task (Goldin-Meadow and Wagner, 2005; Goldin-Meadow, 2011). In a dual task paradigm in which participants had to solve a math problem and explain it after memorizing a set of letters, Goldin-Meadow and colleagues (2001) showed that participants gesturing consistently with speech content while explaining the math problem were better at recalling the letters than participants who did not gesture. The authors suggest that gestures unload working memory so that working memory can be used for a concurrent task (Goldin-Meadow et al., 2001). In the same paradigm, Cook and colleagues (2012) and Wagner et al. (2004) showed that these results were true only for co-speech gestures, not mere rhythmic movement like hand-waving. The fact that gestures support cognitive and working memory load is particularly true for communicating spatial information and solving mental problems (Chu and Kita, 2011; Alibali, 2005). Participants required to mentally rotate 3D objects performed better when encouraged to gesture. The production of gestures is also negatively correlated with visual and spatial working memory scores. The less people are able to solve mental problems, the more they need to gesture (Chu et al., 2014; Marstaller and Burianová, 2013). Chu and Kita (2011) noticed that the frequency of gestures increases with the difficulty of the mental rotation of objects, and decreases with practicing the resolution of mental rotations. This correlation appeared to be also true for verbal memory (Gillespie et al., 2014). Hence, co-thought and co-speech gestures seem to reduce cognitive load when cognitive difficulty increases.

1.2.1.4 Gestures are produced while learning common language concepts

Iverson (2010) details how the development of motor skills and language development are intertwined during early infancy. Iverson takes three actions (rhythmic arm movement, construction, and recognitory gestures) to show how these actions provide new sensory inputs to the infant and how she can use these inputs to interact with the surroundings.

Developments of object displacements in play are related to first words and the vocabulary spurt: the way infants interact with an object provides information on what they know about this object. At the motor level, through object manipulation, children learn how to use their hands independently from one another, to grip, and fine eye-hand coordination. The infants first separate objects, then assemble them. When they can do both, they can start to notice some objects are contents, others are contained. Once children understand the affordance of an object, they develop recognitory gestures, like taking the phone, putting it to the ear and putting it down quickly. Across the evolution of the recognitory gestures, infants also show that they understand that an action is not necessarily self-directed (putting the phone to the ear of a doll for example), and that a common action can be applied to different objects (taking a banana instead of a real phone). Children see that meanings are dependent on the context, and that different objects can have the same meaning. This is the bedrock for understanding that the same word can refer to many referents. In other words, recognitory gestures enhance naming ability. Iverson concludes by warning the reader that motor development is neither necessary nor sufficient to language development, but is ‘normally participatory’: motor milestones contribute to language-learning in a normative development.

To sum-up, as long as hand gestures do not interfere with the cognitive content of the task, they benefit to cognitive performances and play a crucial role in comprehension and reasoning (Pouw et al., 2014). Gestures also play a crucial role in early infancy, when acquiring language concepts.

1.2.2 Gestures are not the only movements to support learning

1.2.2.1 Hand gestures can be generalized to the body: the enactment effect

Hand gestures are not the only movements that contribute to improve learning. For instance, Ruiter and collaborators (2015) showed that for children, learning two-digit numbers by walking following a ruler on the wall improve their score compared to marking on a ruler drawn on a sheet of paper. Cherdieu and colleagues (2017) observed that learning human anatomy through moving these body parts improved retention compared to just seeing the movements. In his doctoral dissertation, Nyberg introduces the enactment effect as the ’superior memory performance of enacted events over non-enacted events’ (Nyberg, 1993, p.1). Enacted events refer to the fact that during encoding, participants perform the task to be remembered (self-performed-task, or SPT) instead of just encoding the sentence describing the task. More generally, body movements can benefit to educational settings: physically enacting the content to be learned may improve the acquisition of this content better than other modalities (Mavilidi et al., 2015).

1.2.2.2 Walking facilitates language acquisition

Walking also provides the infant with new opportunities to interact with the surrounding environment. Displacements are faster, mobility of head and trunk is increased, offering new directions for the eye gaze, and facilitating interactions like joint attention. Walle and Campos (2014) found that walking infants had significantly larger vocabularies than their age-matched peers who were not walking yet. Several studies observed that the language growth trajectory is significantly accelerated just after the onset of walking. Although not having a direct impact on language development, walking seems to facilitate language acquisition.

1.2.3 Interactions between cognitive activities and co-occurent motor tasks

Motor performances can be characterized by variability in space or time of movement execution. Cognitive performances can be linked to the reaction time or error rate on a specific task. A dual-task paradigm in which both cognitive and motor performances are required simultaneously, and in a non congruent way, can lead to either interferences or facilitation effects (Plummer et al., 2013; Leone et al., 2017). Interferences lead to a decrease in cognitive performances or in motor performances. Facilitation effects positively impact either motor or cognitive performances. One prototypical example about the impact of cognitive load on motor task is the effect of speaking on the phone while walking towards a particular target. With divided attention, participants show strong deviations from the path to the target (Lamberg and Muratori, 2012).

Walking is a task often achieved with a cognitive concurrent task, and the risk of falling increases for elderly people. A growing body of research investigates the response of either healthy adults, adults with neurological pathologies, or elderly people, in a cognitive-motor dual-task paradigm involving walking (see Al-Yahya et al., 2011 for a review).

Malcolm and colleagues (2015) found that older adults' movement accuracy decreased significantly when performing a go and no-go task task while walking. In general, in healthy participants, walking speed and regularity decrease with the increase of age, and the increase of cognitive load (Al-Yahya et al., 2011). Regarding patients after stroke, Plummer and collaborators (2013) found that overall, participants exhibit a decrease in gait accuracy only, or in both cognitive and motor performance. Srygley et al (2009) tested the effect of walking on cognitive performances for young and older adults. The cognitive tasks were to recitate serial subtractions out loud, and to listen to a story and count the occurrences of certain words. For older adults, cognitive performances decreased during walking for all the cognitive tasks. For younger participants, cognitive performances decreased during walking for all the cognitive tasks but one: it enhanced the content recall of the story. Li et al (2001) carried out an experiment in which healthy young adults and elderly had to walk while memorizing a list of words. They found that the performance gap between elderly and young adults was wider for the memory task than for the walking task. This suggests that older adults prioritized walking over memory. This is confirmed by the fact that when participants were proposed to have an external aid on one of the tasks, older adults chose to get help on walking, while younger adults optimized memory.

Most of the studies only investigated walking as the motor task. As Schmidt-Kassov and colleagues (2014) and Emhad and collaborators (2011) mention, although we could think of walking as a mere rhythmic motor activity, this complex movement involves higher executive functions necessary for navigation through space (Beurskens et al., 2014) or keeping attention focused on what could possibly come to disturb balance. This is one of the reasons it has been so widely studied: to answer the question whether increased cognitive load comes to disturb executive functions involved in walking (Al-Yahya et al., 2011). Most of the studies focus on the changes in gait, but few look at the cognitive performance.

Mandrick et al (2013) looked at performances during grasping contractions while doing an arithmetic task. They found that the performance of the cognitive task decreased significantly with the increase of force variability. Faulkner et al (2006) investigated how cognitive and walking performance are influenced when performed concurrently in older adults. They observed that push-button reaction-time was increased by 20% when walking. However, Schmidt-Kassov and colleagues (2014) carried out two experiments in which they observed that walking on a treadmill during learning foreign-language words improved retrieval compared to doing no movement.

From this quick overview of the literature on dual tasks, it emerges that concurrent motor tasks are detrimental for cognitive task implying reaction-time or problem solving, but can improve cognitive abilities related to language such as content recall or learning a foreign language. Neural links between action and semantics could partly explain these positive effects.

1.2.4 Interactions between action and semantics

Glenberg and Kaschak (2002) were among the first to empirically report the priming effect of language comprehension on related action performance. Participants had to listen to or read sentences describing actions that implied a movement toward the self or away from the self, and to judge whether the sentence made sense. Answering implied a motor action either matching with the movement of the sentence, or incongruent. The authors found that motor tasks matching with the action described were faster to be performed than incongruent motor tasks. Several studies replicating the same paradigm also found a priming effect of language comprehension on action (Zwaan and Taylor, 2006; Borreggine and Kaschak, 2010). But some studies found reverse effects (Buccino et al., 2005): actions implying the effector (hand, foot, etc.) involved in the sentence were slower to be performed than actions implying another effector (Boulenger et al., 2006). Timing seems to be the key factor playing on whether action-related semantics understanding have an interfering or facilitating effect on the related motor task (Borreggine and Kaschak, 2010; de Vega et al., 2013). One hypothesis related to viewing language sensorimotor simulations through mental imagery, called the resonance hypothesis (Pulvermüller, 2005), is that 'an action temporally overlapping a matching action verb interferes with the meaning-related motor simulation and thus impairs comprehension.' (de Vega et al., 2013, p.28). A wide set of neuroimaging studies suggest that cortical motor areas are implied in action recognition and understanding, resulting in a semantic somatotopy. When reading or listening to an action verb related to the arm, face or legs, the somatotopic area in the premotor cortex corresponding to the effectors of the action was active, as well as when seeing these actions (Pulvermüller, 2005; Hauk et al., 2004). Research using TMS to stimulate specific motor areas showed faster response-times when processing an action-verb involving the effector related to the motor region stimulated (Willems et al., 2011; Pulvermüller et al., 2005). Language comprehension also activates motor areas related to the tongue (Fadiga et al., 2002), the hand (Meister et al., 2003) and even the leg (Liuzzi et al., 2008). The activation of a motor area when processing an action verb increases with practicing the corresponding action (Lyons et al., 2010). The neural overlap between semantics and motor activities does not always facilitate cognition but can also interfere with a cognitive task. Shebani and Pulvermüller (2013) showed that working memory for arm and leg related action words was impaired by a complex motor task performed by the corresponding effector. In a more recent study (Shebani and Pulvermüller, 2018), they showed that a simple motor task such as finger tapping facilitated working memory for arm-related word.

1.2.5 The benefit of mere rhythmic movement on attention processes can enhance cognitive performances

According to the dynamic attending theory (Large and Jones, 1999), attention would not be constant over time, but evolve in cycles. External and periodic stimulus would entrain internal oscillators underpinning attention, supporting event processing. A rhythmic motor activity helps to couple fluctuations of attention with the timing of the event, improving sensory selection.

Indeed, our ability of moving enables us to actively sample the environment around us, for instance by sniffing, haptically explorating an object, or with occular saccades. However, when a movement induces sensory inputs, the nervous system needs to know whether these inputs were triggered by our movement or by an external agent. To do so, copies of movement commands are transmitted to sensory structures. These signals are called corollary discharges (Crapse and Sommer, 2008). They modulate the processing of sensory inputs, coupling sensory processing with motor sequences. On the other hand, neural oscillations responsible for attention can synchronize with external event (Jones et al., 2002). Perception accuracy increases when the peaks of the neural activity occur with the onset of the external event. This way, periodic stimuli attract attention and are processed faster and more accurately (Correa et al., 2006). Hence, an external rhythm enables attention to focus on anticipated points in time, enhancing a more efficient allocation of cognitive resources.

Through corollary discharges together with the tendency of attention to synchronize with a rhythmic stimuli, rhythmic motor activity helps to perceive the environment more accurately. Morillon and colleagues (2014) carried out an experiment in which participants had to track a beat while listening to auditory target tones inertwined with distractors. They found that a concurrent rhythmic motor activity improves sensitivity to target tones and thus enhances the discrimination of auditory information. Schmidt-Kassow et al. (2013) also found that compared to sitting and staying still, pedaling while listening to an external periodic auditory stimuli improved the anticipation of the stimuli, and that the cycling variability was correlated to the variability of the neural oscillations responsible for the attending process. More recently, Plancher et al. (2019) carried out a study in which participants had to encode a sentence verbally or by self-performing it, and retrieve it after a distraction task. They found that to verbally encode a sentence and to retrieve it after repetitively drawing the same number with the hands improved the retrieval scores. The authors proposed that this unexpected effect comes from the periodicity of the motor task: through the mechanism of coupling attention with a rhythm, this regular motor activity helped the attention process focus on the content in the working memory.

1.2.6 Effect of movement on cognition through breathing

Rhythmic motor behavior facilitates attention processes. It has been shown that attention and cognitive functions (Zelano et al., 2016) can be coupled to breathing (Melnychuk et al., 2021). Cognition is also enhanced during physical activity through physiological processes involving breathing. Being crucial for both motion and cognition, breathing is a key element in the interactions between motor and cognitive events.

1.2.6.1 Breathing benefit to cognition through physical activity

Breathing arousal is inherent to physical activity. Engaging in long-term aerobic physical activity triggers neural and physiological processes that come to enhance cognitive performances: new blood vessels are created and the blood flow increases, providing more nutrients; the number of synapses between neurons increases, as well as the number of neurons in the gyrus dentatus and the hippocampus. These two brain regions are widely affected in memory and learning. Exercise increases the production of molecules called brain-derived neurotrophic factor (BDNF) also involved in learning and memory (Kramer et al., 2006). People engaged in moderate to vigorous physical activity (compared to light activity) have larger hippocampal volume (Raichlen and Alexander, 2017).

Oppezzo and Schwartz (2014) observed that creativity is enhanced after walking. They discuss the cause of this enhancement. They eliminate an embodied explanation according to which moving the legs would stimulate brain areas related to creativity, since moving the legs while sitting on a chair does not have the same effect as walking. They also eliminate the factor of the environment (walking outside provides an external flow of stimulations), because walking inside, facing a white wall, has the same effect as walking outside. A possible explanation is that the creativity enhancement could emerge from the stimulation of physiological processes such as breathing, which in turn brings more oxygen to the brain, together with increasing bloodflow.

1.2.6.2 Attention and action are coupled with breathing

Neurological studies have recently highlighted an entrainment of the waves from hippocampus, cortical and subcortical areas by the respiratory frequencies through the olfactory bulb (Heck et al., 2019). Zelano et al. (2016) observed that retrieval of object pictures was better when encoded during inspiration phases. Park and colleagues (2020) showed that readiness potential of voluntary action was coupled to respiratory phases and modulated by the periodical breathing action.

The links between cognition and movement could be underpinned by breathing. This triad could play a key role in spoken language. The interactions between speech, breathing and limb movements have been poorly studied in other contexts than physical effort. Moreover, in this three-component interaction, one of the components is often used as a control variable and not analyzed. The following chapter reviews the interactions between speech and breathing, breathing and movement, speech and movement, and ends by discussing why and how to study the three components altogether.

CHAPTER 2

Interactions between speech, breathing and limb movements

The objective of this chapter is twofold: (1) to make a state of the art of the studies investigating the interactions between speech, breathing and motion; (2) to highlight the need of studying the interactions between speech, breathing and limb movements in an ecological context. The first section reviews the interactions between speech and limb movements depending on the type of limb motion: either a co-speech gesture, a concurrent motor task or a mere rhythmic movement. The second section briefly describes the breathing system in humans; the third section details the links between speech and breathing. The fourth part reviews the findings on motor respiratory coupling. The four first sections raise points that are still unclear and motivate the empirical work described in this thesis. The fourth section argues in favor of studying speech, breathing and limb movements and their interactions, and sets the context of this thesis, detailing the questions raised, the objectives and the paradigm chosen.

2.1 Interactions between speech and limb movements

Since the acquisition of speech motor control and motor control of other body actors are interleaved and occur at similar periods during early infancy, speech and body movements stay tightly linked and often synchronized in adulthood. This section reviews research about the interactions between speech and body movements.

2.1.1 Speech and co-speech gestures

Speech and limb movements have been explored altogether with a big emphasis on co-speech gestures, especially with the hands but also with the head, or the eyebrows (Wagner et al., 2014). Kendon (2004) defined a continuum to classify manual gestures that occur in a communication situation along 4 dimensions: co-occurrence with speech, linguistic properties, conventional and non-conventional gestures, synthetic and analytical. These dimensions enable to range gestures from gesticulations (co-verbal gestures) to sign language. In this thesis, we will focus on co-verbal gestures (gesticulations). These gestures accompany speech, are not conventional and are synthetic: they do not follow an analytical and linear structure like language, but deliver information all at once. Co-speech gestures are movements of the hands and the head (see Wagner et al., 2014 for a review), or even the eyebrows (Kim et al., 2014), that occur along with speech. Most of the time, these movements are synchronized to speech, or at least coordinated - gestures can emerge during disfluencies (Ragsdale and Fry Silvia, 1982) - and often convey meaning. Gestures are omnipresent in spoken communication. Corballis (2003) argues that manual gestures are the precursors of language in human evolution, and that they have been progressively accompanied by sounds. Some observations seem to corroborate this theory. For instance, Iverson and Goldin-Meadow (1997) showed that people who are blind from birth also produce gestures when they speak, suggesting that the speech-gesture link is not only learned by observing or imitating others but is intrinsic to speech production. Similarly, it has been famously reported that gesture-speech coordination was unaffected in a 'deafferented' patient who lost the sense of proprioception (Pouw et al., 2020c). This patient gestured normally without vision, at times not even knowing that he indeed was gesturing. McNeill and colleagues (1992) claimed that gestures are different from action, suggesting a specific link between language and the hand that would not necessitate a sensory feedback loop. Pouw and colleagues (2020c) reanalysed the data from this patient and suggest that he may have developed other sensorimotor and interoceptive feedback links that have to do with the biomechanical entanglement between speech and hand gestures via the respiratory vocal system. In any of these renderings, it is clear that gesture control is deemed special and can make unique use of perceptual feedback loops that have to do with the entanglement with speech. From these observations, how can the coupling of manual gestures to speech be conceived?

2.1.2 How can gesture-speech interactions be theoretically described?

To explain the underpinnings of the speech-gesture entanglement, three views stand out from the literature. All three consider speech and gestures as two channels, but differ in the way these two channels are processed. McNeill (1992; 2008) takes the view that speech and gestures come from the same internal representation and evolve as a whole from conception to execution, with bi-directional interactions between the two. He asserts that gestures are not a mere by-product of language, but an intrinsic component of it. Speaking would consist in two cognitive processes: mental imagery and analytic form. Mental imagery emerges from thinking, and is constantly present in the speaking process as a totality, whereas the analytic form allows the focus of attention to move from one component of the image to another (Blumenthal, 1970). This duality is packed altogether in what McNeill calls the growth point. The growth point is the 'smallest unit [...] that retains the essential properties of a whole, in our case the whole of an image and a linguistically-codified meaning category, such as we see in the speech-gesture window' (McNeill and Duncan, 2000, p.4). Speech and gesture chunks are conceived altogether through a growth point, to jointly express one unity of meaning. The growth point resists to perturbations such as delayed auditory feedback (McNeill, 1992; Pouw and Dixon, 2019), under which speech and gestures stay synchronized. McNeill and colleagues (1992) add that the conception of the growth point depends on the language spoken: a same meaning can be expressed differently depending on the language. This hypothesis is supported by the results of Kita and Özyürek (2003). They found that gestures were specific to the way an action was described in each language (Turkish, Japanese and English). Retelling the event 'the cat was rolling down the hill', English speakers can encode the manner (rolling) and trajectory (down) in the same clause, whereas Turkish and Japanese participants need two clauses to express it with two distinct verbs. These different encoding structures were reflected in gestures: while English speakers were more likely to imitate the rolling action with the arms going down in the same gesture, Japanese and Turkish speakers were more likely to imitate the action of rolling and going down in two different gestures. In line with these findings, Kita and Özyürek (2003) proposed the information packaging hypothesis: gestures and speech would come from the same designer; gestures would emerge from mental imagery while conceiving speech, and would provide speech with another manner of conveying information (spatial and synthetic). Gestures would be shaped both by the linguistic constraints that determine the way language is produced (by information package) and by the integration of non-linguistic information.

Still linked to the idea of the growth point, and rooted in theories of embodied cognition, Hostetter and Alibali (2008) conceive gesture as a simulated action: speech and gestures are produced from the same process of mental imagery, a process that manifests itself in the conception of a thought, and which activates the premotor cortex. If this activation exceeds a certain threshold (this threshold being defined by many factors such as context or life experience), then the activity of the premotor cortex spreads to the hand control area of the motor cortex (not only the co-articulation control) and triggers hand movements. The hands are particularly concerned because the hand and mouth motor areas are very close in the motor cortex, and their synchronisation suggests that activation of one causes activation of the other.

On the other side of the literature, researchers (Butterworth and Beattie, 1978; Levelt et al., 1985; Hadar, 1989) consider that speech and gestures are two distinct channels, with an unidirectionnal interaction from speech to gestures. Gestures would be a mere by-product of speech when speech is not available. In a less modular vision, Krauss and colleagues (1998) propose that gestures can impact speech (positively) only during lexical retrieval issues. The lexical recall hypothesis states that gestures would be conceived before language, based on mental images (Ruiter, 1998), and would emerge only with difficulties to remember a word, in which case the gesture would help this recall.

Box 1. Motor synergies and the framework of non-linear dynamics

The human body displays many degrees of freedom. One of the principal problematics of human movement studies is to understand how humans manage to control so many degrees of freedom at the same time, within a same task. One of the solutions proposed is that movements are organized into functional groups of muscles that act as a single unit (Bernstein, 1968; Kelso et al., 1986). These groups are called synergies. Research started to model synergies as complex systems composed of non-linear, limit-cycle oscillators synchronized to each other (Kelso et al., 1986). Synchronization phenomena were first introduced by Christiaan Huygens in 1665. Huygens described the synchronization between two pendulum clocks on a same surface. Synchronization phenomena are widely spread over biological systems, and present at many biological levels of interactions such as between heart and breathing paces (Schäfer et al., 1998), between limbs (Temprado et al., 2002), individuals (Richardson et al., 2007) or animals (Buck, 1938), implying two or more components. As diverse as these examples are, they can be described by the common framework of nonlinear dynamics (Pikovsky et al., 2003).

Synchronization can be understood as the "adjustment of rhythms of oscillating objects due to their weak interaction" (Pikovsky et al., 2003, p.8). For these interactions to take place, the oscillating objects have to exchange information. In the case of the pendulum clocks of Huygens, the oscillating pendula were hanged to the same beam. This beam constituted the common vector through which the pendulum clocks influenced the movement of each other.

In human behaviour, interpersonal synchronization can emerge from visual or auditory feedback. For instance, while walking next to another person, the two walking paces converge to the same frequency through exchange of perceptual (visual, auditory) information (Zivotofsky and Hausdorff, 2007). Naturally, agent's behaviors tend to converge to these synchronization patterns, called attractors, and stay within them, depending on the coupling strength and the stability of the attracting state. Synchronization states also emerge from intra-personal repetitive limb movements, such as bi-manual coordination (Kelso, 1984), or from the production of vocalizations along with finger tapping (Treffner and Peter, 2002).

In 1993, combining the growth point theory of McNeill and the findings of Kendon on temporal coupling between speech and gestures, Tuite (1993) proproses a model in which kinetic pulses would conjointly pace speech and gestures, and give them a certain regularity.

Coming from an internal clock, these pulses would be reflected in beat gestures or head and eyebrows movements, together with pitch accent in speech.

In line with the model of Tuite and based on the idea that speech and gestures are conceived through the same channel devoted to communication, Iverson and Thelen (1999) propose a model based on non-linear dynamics (see Box 1). Gesture-speech coupling would emerge from infant babbling in which, with practice, at least two oscillators (speech and gestures) would progressively fall into synchronisation, depending on the easiness of the behavior. They define two notions: the threshold at which a behaviour emerges and its activation. The threshold of a behaviour represents its ease of performance. They first describe how mouth and hands are coordinated in early infancy, with the babkin reflex, and how hands bring objects to the mouth. Then, rhythmical movements such as waving or banging emerge together with reduplicated syllables. Iverson and Thelen suggest that these two co-occurrent rhythmical activities influence each other, citing a study in which syllable length during babbling occuring along with hand movements were significantly longer than syllables produced alone (Ejiri and Masataka, 1999). These repetitive entrainments between hand movements and the vocal tract would set the stage for the development of the speech-gesture system, which develops in the last month of the first year with more controlled gestures such as pointing and a more directed communication. Words usually emerge a few weeks later. Iverson and Thelen hypothesized that gestures are easier to produce than speech, so they have a lower threshold and a higher activation. Gesture production would enable the emergence of speech, by the previous coupling of both that was set in early months. During word learning and speech acquisition, gesture production would tend to decrease to focus on the effortful activity of learning speech, and increase again once speech is acquired, both being tightly synchronized. Iverson and Thelen conclude by claiming that 'systems of movement for mouth and for hand cannot be separated from one another, and they are intimately linked in the production of language, the pinnacle of human cognition.' (p.35).

Across the literature, these models have been supported or disclaimed by behavioral results. What can speakers' behavior tell regarding the theories on gesture-speech interactions?

2.1.3 Behavioural evidence of gesture-speech interactions

2.1.3.1 Potential effects of constrained gestures on speech

A wide set of behavioral studies tend to confirm that hand gestures emerge with issues during lexical retrieval or conceptual load, and improve speech fluency (Ravizza, 2003). First studies observed gestures in the natural context of interview or conversation. Ragsdale and Sylvia (1982) reported that 81% of body movements during an interview were going along with vocal hesitation, and 83% of vocal hesitations were produced with movement. Pine and collaborators (2007) found that when children were free to gesture in a naming task, they gestured more when being in a tip of the tongue (ToT) state than when not, and resolved more ToT states when allowed to gesture than not. A ToT state occurs when the semantic information of a word is known but the phonological retrieval of this word is temporarily blocked (Brown, 1991).

To go deeper into the interaction between speech and hand gestures, researchers started to control either the speech difficulty or the gesturing condition: Rauscher and colleagues (1996) set-up an experiment in which people had to recall a cartoon, either allowed to gesture or not, in different conditions of speech difficulties (normal, use obscure words, don't use words with a certain letter) to trigger lexical research. They found that only during spatial content recall, speech fluency decreased when not being able to gesture (more pauses and longer pauses), and people gestured more when recalling spatial content. Increasing lexical access difficulties had the same effects, but not only on spatial content. In the experiment of Graham and colleagues (1975), people had to describe line drawings of two-dimensional shapes whether allowed to gesture or not. Not being able to gesture decreased semantic content about space and increased the length of the pauses, but not their number. Gestures also seem to have an effect on verbal content quantity and overall speech acoustics: more recently, Cravotta and colleagues (2019) asked participants to describe the content of comic strips, either being encouraged to gesture or not. When encouraged to gesture, the number of words of narratives increased, as well as maximum F0, and intensity mean and maximum. Other previous studies reported that not being able to move the hands while speaking can reduce vocabulary diversity and precision (Rimé et al., 1984; Hostetter et al., 2007b).

In a study implying the task 'explain to me how to tie shoelaces', Hostetter et al (2007a) assessed several variables linked to the different hypothesis on speech-gesture processing. When participants were prevented from gesturing, they observed that they produced more hesitations at the onset of an utterance and less rich verbs, suggesting that they need more conceptual load to decide how to pack their spatio-motor ideas. Indeed, according to the information packaging hypothesis of Kita and Ozyurek (2003), gestures help package thoughts into units for speaking. They did not replicate the results of Rauscher and colleagues (1996): the filled pauses did not increase with gesture prohibition. In a more recent study in which participants had to explain how to tie a tie, Hoetjes and collaborators (2014) did not find any effect of not being able to gesture on the duration of the instructions, on the number of words used, the speech rate, on the number of filled pauses used, or on the overall F0 level.

What comes out from the literature is that when speech difficulty or disfluency increases, hand gestures are more likely to occur, but not being able to gesture with the hands does not always have an impact on speech. Another component to consider in the gesture-speech event and that can afford answers on the role of gestures in speech is the temporal coordination between speech and gestures.

2.1.3.2 Speech-gesture coordination

As Parrell and colleagues (2014) review in their introduction, Kelso and colleagues (1983) were the first to investigate the synchronization patterns between vocalizations and finger tapping in a non-linear dynamic framework. Participants were asked to repetitively produce the word 'stock' and tap their finger with the right hand at the same time. They had to alternatively stress either speech or their finger tapping, and keep the magnitude of the other task constant. The authors noticed that participants kept stressing either the concurent task, unable to keep its magnitude constant. They suggested that speech and finger motion were coupled as a single coordinative structure. Parrell and colleagues (2014) generalized the results of Kelso and collaborators (1983) to controlled sentences. Zelic and colleagues (2015) examined the speech-gesture coupling during the production of silent and phonated non-communicative vocalization of syllables along with finger tapping. The synchronization between the two was maintained during both phonated and silent speech, suggesting a somatosensory basis of the coupling, not requiring any auditory feedback.

These findings are consistent with studies using different designs from non-linear dynamic paradigms. Krahmer and Swerts (2007) tested whether producing a beat gesture at a different position from the pitch accent within a sentence had an effect on speech acoustics. They found that the production of a beat gesture increases the duration and the formants (F2 and F3) of the co-occurring vocalization, even when it is not aligned with the intended pitch accent. Coordination was also found between manual pointing gestures and gestures of the vocal tract (Krivokapic et al., 2016; Rochet-Capellan et al., 2008; Roustan and Dohen, 2010). Interactions were also found between the production of a word concurrently with a grasping task: when an object is picked up while producing a syllable, the aperture of the lips and the first formant increase depending on the size of the grasped object (Gentilucci and Volta, 2008). Similarly, when a word is said along with a gesture, compared to no gesture, the second formant increases. Compared to no word, the gesture slows down when produced along with a word (Gentilucci et al., 2001).

Another way of testing the strength of the coordination between speech and gesture is to perturb one of them. Levelt and colleagues (1985), and more recently Chu and Hagoort (2014), perturbed either speech or gesture and observed if the other effector changed the dynamics of its motor execution to wait for the perturbed effector. Participants had to point and name the light that lit up. Their gesture was disrupted in random trials, by a weight of 1,600 grams, at the early or middle phase of gesture execution. In Levelt's experiment, the speech was delayed only in the early phase, not in the middle. They concluded in favor of a ballistic view in which no interactions occured during motor execution. In Chu's experiment, participants delayed the speech onset when the gesture was disrupted, and extended their gesture duration when the speech was disrupted, even when this disruption occurred after the execution onset of their action. These results suggest interactions between speech and gesture during motor execution. In line with this study, Pouw et al (2019) found a stronger gesture-speech synchrony under delayed auditory feedback compared to speaking and gesturing with no auditory feedback.

The emergence of a coordinative structure between speech and hand movements has been demonstrated, either in repetitive vocalization paradigms, or during controlled phonated words and sentences, and with different hand movements (pointing, finger tapping, beat gestures). As Parrell and colleagues (2014) mention, the way these synchronization patterns generalize to natural speech together with hand movements is not properly understood yet. Whether this coordination emerges from cognitive processes linked to spoken language or from purely motor levels and can be extended to non-communicative movement is not clear neither. It is not always possible to say whether the movements used in the previous studies have communicative purposes (and are entangled in a communicative coupling) or not. At first glance, finger tapping could be considered as non-communicative, but finger tapping movements are not completely absent from communicative gestures and could be easily fitted in. The question of speech and non-communicative limb movements acting as a coordinative structure remains. Answering this question could help understand whether speech-movement coordination necessitates communicative purposes and thus comes from higher cognitive levels, or if it is a general principle of motor behavior between different body effectors.

McNeill (1992) reported a pilot study examining speech-gesture synchrony under delayed auditory feedback in a natural spoken language paradigm, and observed that in the case of rehearsed speech and gestures (and contrary to extemporaneous speech), when speech or gesture is disturbed, the other does not adapt to this perturbation and gets out of synchrony. McNeill argues that gestures and speech are no longer tied by meaning, and that the speech gesture synchrony holds on the conceptual meaning ('normal dialectic of image and gestures in structuring thought', p.279). But speech-gesture synchronization also lies in the shared motor system between hand and mouth. Neurophysiological studies of Broca's area confirm the tight motor link between the hand and the mouth (Fadiga et al., 2009): some cytoarchitectonic properties are shared between BA44 area, which comprises Broca's area, and the adjacent area BA6 in the premotor cortex. Microstimulation and single-neuron studies showed that both hand and mouth movements are represented in area F5 of monkey's brains. Broca's area is active during human-made tool's manipulation and categorization, grasping (Binkofski et al., 2001; Gerardin et al., 2000) and observation of goal-directed gestures.

Speech often occurs while doing a non-related motor activity such as biking or cooking. If this motor activity involves forelimbs, the neural overlap caused by speech and hand movements within Broca's area should result in interferences in the co-occurent motor activity. The next section reviews the studies implying dual-task paradigms between speech and motor tasks.

2.1.4 Interactions between movements and speech in a dual-task paradigm

2.1.4.1 Speech disturbs kinematics of the limbs

The literature is consistent on the fact that speech affects co-occurent motor tasks. It is particularly true for older adults walking and talking at the same time (Verghese et al., 2007), or patients with a stroke: while talking, these populations lower the cadence of their movement (walking, finger tapping) and/or increase its variability. In a pilot study, Davie and colleagues (2012) investigated the effect of a verbal task on a continuous gait task within a group of women. The verbal task varied in its complexity (non-oral motor movements, spoken non-words, or spoken words). They found that all three conditions increased gait variability and were not statistically different between each other, suggesting a purely motoric interference between speech and gait. Studies on speech and co-occurent non-communicative forelimb activity are less common. Dayalu et al. (2013) investigated whether speech and manual fluency were coupled during a dual task. Participants had to draw a circle while producing and listening to fluent and disfluent speech. Significant decreases in manual fluency were found when participants read or listened to disfluent speech.

However, speech is not always affected by a dual motor task context. Verghese et al. (2007) did not find any effects on speech (saying alternate letters of the alphabet) but they found that prioritizing talking lowered velocity and cadence of walking. Dayalu et al (2013) did not find any effect of drawing on speech. Whitfield and collaborators (2019) tested whether dual-task situations enhanced speech motor deficit in individuals with Parkinson disease. Participants had to draw a circle counterclockwise during reading or extemporaneous speech. The authors found that for the control group, speech was not disturbed by the motor task. For patients, pause duration decreased in extemporaneous speech during the motor task. Controversy in the effects of motor task on speech may lie in the different speech tasks (saying alternate letters (Verghese et al., 2007) vs. spontaneous speech (Kemper et al., 2003) or different measures of speech: Whitfield et al (2019) measured speech rate, mean pause duration, F0 variability and formant space, and did not find that the manual task (drawing a circle) affected speech. However, Kemper et al

(2003) measured sentence length, grammatical complexity, and propositional content and found that walking or complex finger tapping affected these speech parameters (in the young group as well). The discrepancies of the effect of motor tasks on speech may also be due to the nature of the motor task. Walking involves gait performance and navigation in the space, while drawing a circle is a finer motor task and does not involve the whole body. While kinematic parameters of motor tasks are quiet obvious and easy to record (movement speech and variability, position, jerk etc.), speech is a more complex system using motor as well as cognitive resources. In his review on nonspeech oral movements, Ray Kent writes: ‘An implicit definition of speech generally is assumed, even though speech is not a monolithic behavior but rather subsumes a variety of sensory, motor, and cognitive skills that vary across behavioral tasks’(Kent, 2015, p.764). Hence, motor tasks may have effects on speech, but there is no consensus yet on how to measure these effects.

2.1.4.2 Limb movements impact speech acoustics

As seen with the particular case of beat gestures in section 2.1.3.2, limb movements can impact speech acoustics. McCaig et al. (2016) examined the effects of concurrent walking on speech intensity and speech rate during an extemporaneous speech task. The concurrent walking condition produced higher speech intensity in both the group with Parkinson disease and the control group compared to standing and sitting conditions. Recently, Pouw and colleagues (2019; 2020d; 2020b; 2020a) suggested a biomechanical link between upper limb movements and speech acoustics in relation to anatomical constraints between the arms and the breathing system. They indicated that abrupt accelerations or decelerations of the upper limbs have a cascading effect on the respiratory system. Abrupt accelerations or decelerations increase subglottal pressure, which may have consequences on acoustic parameters of speech. In particular, the amplitude envelope and fundamental frequency (F0) increase close to the deceleration peak of the arm movement. These effects were also measured for a vertical rhythmic flexion-extension movement of the wrist, but to a lesser extent than arm movement. The latter observations were made for syllable production (2019; 2020b) and spontaneous speech (2020a). This idea of a biomechanical entanglement is in line with findings on a developmental level. For infants, the motor milestone of sitting seems to accelerate vocabulary expansion (Oudgenoeg-Paz et al., 2012). Achievement of unsupported sitting is the forerunner of improvement in vocalization going from the production of simple and not timing-controlled vowels to more distinct speech patterns like CV units (Iverson, 2010). This may be due to biomechanical reasons: the sitting position allows more control of the ribcage and the upright head position changes the position of the spine and the vocal tract curve (including the tongue).

Exercises involving the lower limbs also affect speech acoustics: average speech intensity and average F0 increase while biking at different physical effort levels (Mohler, 1982; Fuchs et al., 2015; Weston et al., 2020). F0 also increases with effort during or after a treadmill task (Primov-Fever et al., 2013; Trouvain and Truong, 2015). Trouvain and Truong suggested that this increase comes from a higher subglottal pressure during physical effort. However, other studies highlight inter-speaker variability. Participants have different and sometimes even opposite behaviors regarding the evolution of their F0 while performing limb movements: some increase F0, some decrease it, and for others, F0 remains unchanged (Godin and Hansen, 2008; Godin and Hansen, 2015; Weston et al., 2020). Besides inter-speaker variability, the relationship between exertion level and F0 is not linear: Johannes and collaborators (2007) found that F0 is impacted by exercise only close to exhaustion. There may also be local effects of motion on F0 that might not appear when F0 is averaged over a trial. The local effect of motion on F0 may depend on the peak acceleration value of the movement (Pouw et al., 2019; Pouw et al., 2020b). Godin and Hansen (2011) found a larger effect of physical activity on nasals than plosives and fricatives while using a stepper. This effect may be related to breathing strategies (e.g., leaving the velar port open to inhale via the nostrils).

Interactions between spoken language and body movements are a complex object, depending on the co-ocurrent movement type (either gesture or non communicative motion) and implying cognitive, motor as well as biomechanical effects. This complexity comes from the fact that speech consists in the intricacy of various sensory, motor and cognitive processes. The interactions between movement and speech are no longer in question, but they have yet to be explained. A third, often forgotten, physiological actor that can play a fundamental role in speech-movement interactions is breathing, a resource shared by both speech and limb motion (see Chapter 1 section 2.2.4). The latter is involved both in the production of speech, for which a specific respiratory control must be developed, and in the control of limb movements, which also influences and depends on the oxygen needs of the body. These movement-breathing and speech-breathing relationships have been studied in different research areas.

2.2 Breathing: the neglected actor in the speech-limb movement interaction.

2.2.1 A brief description of the human respiratory system

The primary aim of respiration is to maintain the homeostasis of the body. Respiration supports gas exchange by means of a continuous flow of oxygen-rich air in the lungs and carbon dioxide–rich air out of the lungs. When breathing at rest, this mechanism results in fairly symmetrical inhalation and exhalation phases.

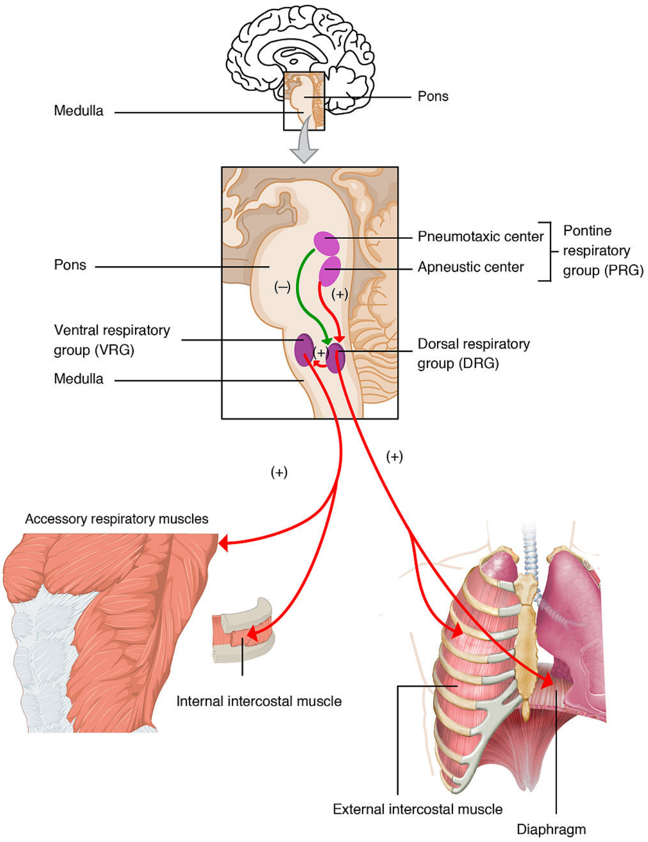


Figure 2.1: Breathing mechanisms from brain stem neural centers to the muscles (picture from wikimedia).

The breathing apparatus is situated in the thorax and the abdomen. The lungs itself do not contain muscles. They can't be active. They are surrounded by the pleura, which is embedded in the thorax. Inhalation involves the activity of the diaphragm, which drags the lungs down, increasing lung volume. Many muscles contribute to breathing. Inner and external intercostalis muscles are the most often discussed in the speech literature. The external intercostal muscles also increase the volume of the lungs between each rib. They are involved in inhalation while the inner intercostalis muscles are involved in exhalation. The quantity of air stays the same while the volume containing this quantity increases, resulting in lung pressure to decrease. Air particles always move in the direction of lower pressure. When this pressure is lower than the pressure outside the lungs, air flows inside the lungs. The passive exhalation is initiated by the relaxation of the diaphragm which involves an upward motion. This motion reduces lung volume, and increases the pressure inside the lungs. When the pressure in the lungs is higher than the atmospheric air pressure, the air naturally goes out. During active exhalation, like when speaking, the abdominal muscles can be contracted to monitor the airflow. Another thing coming into play are the upper vocal tract structure. The glottis has to be open for the air to move out. The larynx and the upper vocal tract can also modulate the airflow during exhalation in speech. Breathing can be characterized by the respiratory rate, the airflow shape or the air volume inhaled and exhaled. This volume takes various amplitudes depending on the needs at stake. When breathing at rest, one does not need to use the whole lung capacity volume. The air volume inhaled and exhaled at rest is called the tidal volume. Behavioural control upon breathing can alter this tidal volume, and can also alter breathing rate, airflow shape or the activity of respiratory muscles (Shea, 1996).

The control of breathing muscles is shaped to meet both metabolic and behavioral needs, and depends on various structures of the nervous system (see (Shea, 1996) for a review). When not behaviorally controlled, breathing is automatic. Neurons responsible for automatic respiratory activity are situated in the brain stem, and are called the central pattern generators. As seen in Figure 2.1, they consist of three groups (Feher, 2017): the dorsal respiratory group controls inhalation, while the ventral respiratory group controls exhalation. Both receive information from the Pontine group, which manages the pace and patterns ventilation. The Pontine group is composed of the pneumotaxic and apneustic centers, which are antagonists, and influenced by connections coming from other brain structures such as the hypothalamus, chemoreceptors and mechanoreceptors informing about the level of oxygen in the blood. These connections enable the modification of the breathing rhythm depending on the context, the emotions or the situation. Primary respiratory circuits are also located in the spinal cord, with phrenic, intercostal, and abdominal motoneurons. These circuits receive inputs from the brainstem respiratory center to send them to the concerned respiratory muscles, from which they receive feedback. In case of dysfunction of the automatic control of breathing, the cortical areas can also ensure its maintenance.

2.2.2 Breathing is speaker specific

Respiration patterns are specific to each individual. The individual specificity of breathing also lies in the temporal patterns of the breathing cycles and the shape of its airflow. In the 60's, researchers started to investigate the speaker-specific dimension of the breathing patterns. Dejour et al. (1966) claimed that "... even under identical conditions, the characteristics of ventilation differ from participant to participant, which leads to the concept of a ventilatory 'personality.'" . The term "personality" is used in psychology and may have a different implication. For this reason we rather refer to individuality, or individual consistency. Individual consistency in breathing profiles refers to the fact that individuals stay closer to themselves/more consistent than to the others in terms of breathing patterns. Benchetrit et al. (1989) investigated individuality in tidal breathing of adults by comparing within- versus between-participant profiles at two time points separated by 4–5 years (1989). The authors report a significantly lower within- than between- participant variability on different parameters used to characterize the breathing cycle. Shea et al. (1987) reported large between-participant differences in tidal breathing profiles, but consistency for the same person recorded four times over 2 days and for breathing during deep sleep (Shea et al., 1990). Tidal breathing is also more similar in monozygotic twins than in random and dizygotic pairs (Shea et al., 1989; Kawakami et al., 1984). These findings suggest that basic breathing rhythms are to some extent determined by genetic and/or physiological factors.

Breathing is the only mechanism controlled by the vegetative nervous system that can also be controlled voluntarily and that can be modified to achieve a wide range of skills. Voluntary breathing, such as that used in speech, singing, or playing wind instruments, might be variable and requires an adaptation of breathing control to the specific skill. This leads to the question of the relationship and potential interaction between the individuality of breathing profiles observed for tidal breathing and behavioral breathing. The fact that breathing is highly variable between activities, as well as between participants for a given activity, suggests task-specific individual control of breathing. Eisele et al. (1992) observed no within-participant stability between breathing at rest and breathing during exercise, but a within-participant stability within the physical activity. More recently, Besleaga et al. (2016) observed the individuality of breathing profiles during moderate volitional hyperventilation, this individuality being unrelated to that observed in the breathing profiles at rest. Hence, breathing at rest, as well as volitional hyperventilation, are variable between participants and between tasks for a given participant but consistent for the same participant in a given task. Does this also apply to breathing in spoken communication, which is another highly specific facet of breathing? It would be interesting to study speech in different bodily contexts over time to answer this question.

2.2.3 Automatic breathing arousal through limb movements

Bodily movements such as locomotion can impact respiratory rhythm through neural connections between pontomedullary respiratory networks, the mesencephalic locomotor region and the spinal cord. Peripheral sensory feedback from body movements also modulates respiratory rhythm (for a review see Shevtsova et al., 2019).

The spinal cord encompasses lomocotor circuits (at the lumbar and cervical level). These circuits can generate and modulate locomotor paces (Guertin, 2013). They include neurons generating the locomotor rhythm, and motoneurons responsible for the activity of the limbs, from which they receive sensory feedback. They also receive inputs from the brain stem and higher levels.

At the onset of moderate exercise, respiration frequency and tidal volume increase quickly (Whipp et al., 1982). This rapid increase is partially explained by stimulations of the chemoreceptors related to the respiratory circuits, but cannot be explained only by metabolic stimulations. Researchers suggests that the respiratory central generator pattern receives inputs from the motor cortex, subthalamic and mesencephalic locomotor areas and muscle or joint afferents located in the spinal cord, resulting in hyperpnoea (Eldridge et al., 1981). Indeed, work on the neural link between respiratory and limb muscle neural structures suggest that the brainstem respiratory network can receive inputs from the lower limb muscles, passing through the spinal locomotor circuits or even bypassing it (Morin and Viala, 2002). One of the possible element of the respiratory network responsible for transmitting inputs related to limb movements is the parabrachial and Kolliker-Fuse nuclei (PBN) (Smith et al., 2009). The PBN is a medullary region that sends inputs to the pre-Bötzinger group. These inputs come from the spinal cord through an afferent travelling up to the PBN. The interactions between the locomotion and the respiratory networks at the spinal level are not fully understood yet. Literature suggest some afferent links from limbs to respiratory neural circuits, through spinal interneurons involved in both locomotion and respiration (Morin and Viala, 2002; Le Gal et al., 2016).

The control of respiration largely depends on a central pattern generator located in the brain stem, but is also influenced by other subcortical and cortical areas involved in the control of complex behaviors. Respiration is highly sensitive and adaptive: any sensorial stimulation, emotional reaction, and an increase in cognitive or physical activity have consequences on the breathing rhythm and may even reshape the breathing cycle (Del Negro et al., 2018; Ben-Tal et al., 2019). These recent findings suggest a clear interaction between breathing and other motor or cognitive activities. How does breathing interact with speech and limb movements?

2.2.4 Speaking while moving: a competition between behavioural and automatic control of breathing

Breathing is inherent to speech. Speech is a particular kind of volitional breathing, which demands a fine control of the airflow depending on the air needed for the upcoming sentence and its prosody. Since speech cannot take place without breathing, and has to adapt to speaker specific physiological constraints. In this vein, as MacIntyre mentions in her thesis (Macintyre, 2022), breathing may contribute to give its rhythm to speech, also being speaker specific. This breathing rhythm is perceived and processed by addressees. For instance, a speaking virtual avatar is judged more human when breathing is audible in its speech (Bernardet et al., 2019).

Breathing is also needed during body movements and physical activity, to meet physiological demands. It often occurs that during intense physical effort, people can no longer speak normally. As Shea and colleagues (1996) mention in their review, during speech, needs in air volume and mean flow increase (Bunn and Mead, 1971; Phillipson et al., 1978). If physical exercise simultaneously occurs, automatic control of breathing increase respiratory rate and magnitude. If the level of physical activity does not necessitate too much resource, the airflow can still be voluntary limited to maintain a certain speech quality, or increased to facilitate exercise. Individuals generally chose to maintain speech quality: during physical exercise, ventilation during phonation does not increase as much as when being quiet (Doust and Patrick, 1981). The maintenance of speech quality cannot hold above a certain physiological demand. A method widely used in the literature to determine this threshold is the talk test. The talk test consists in assessing speech production difficulty to control aerobic exercise intensity (see (Foster et al., 2018) for a review). It has been recently shown that the stage at which normal speech can no longer be maintained corresponds to the onset of a metabolic acidosis during increasing exercise intensity (De Lucca et al., 2021).

2.3 How does breathing interact with speech and limb movements ?

2.3.1 Breathing while speaking: a trade-off between cognitive demands and respiratory capacity

The main view on the interaction between speech and breathing is that breathing adapts to speech (Fuchs and Rochet-Capellan, 2021). Speech requires radical changes in breathing that strongly vary according to linguistic, cognitive, and interactive parameters (Fuchs and Rochet-Capellan, 2021; Conrad and Schoenle, 1979). The larynx is controlling the subglottal pressure in a fine tuned way and keeping it constant (Brandimore et al., 2015). Unlike automatic breathing managed by the brain stem, such as in sleeping (Shea et al., 1990), speech breathing is also handled by the forebrain, and requires motor planning of the coordination between breathing muscles, the larynx and the articulators. To some extent, speech is probably the behavior involving the most dramatic reshaping of ventilation (Wilder, 1974). Compared with non-speech breathing, inhalation phases are strongly reduced, while exhalation duration increases. This is particularly the case for read speech and, to a lesser extent, for spontaneous speech (Fuchs et al., 2013; Winkworth et al., 1994; Rochet-Capellan and Fuchs, 2013; Winkworth et al., 1995). Inhalation duration and depth are correlated to the upcoming utterance, especially in reading (Sperry and Klich, 1992; Winkworth et al., 1994; Fuchs et al., 2013), but also for extemporaneous speech (Huber, 2008; Rochet-Capellan and Fuchs, 2013), although less evidence were found (Winkworth et al., 1995). Investigating the declination of the fundamental frequency (f0) to the expiratory phase, Fuchs and colleagues (2015) found that the declination of f0 often observed at the end of a sentence in German is not related to physiological resources such as lack of air, but to a cognitive control during which prosody would be planned according to the length of the sentence. Thus, motor planning of speech breathing would depend on the upcoming utterance length. Speech breathing is influenced by changes in loudness or voice quality. Depending on the age of the speaker and the speech context, the speech breathing strategy is a trade-off between muscle activation and recoil pressure (Huber et al., 2005; Huber, 2007). The structure of speech breathing is constrained by syntactic boundaries such as punctuation or conjunction (Rochet-Capellan and Fuchs, 2013): in reading as well as in spontaneous speech (although to a lesser extent), most of the inhalation onsets occur at syntactic boundaries. Speech breathing is also constrained by interactive events, such as turn-taking in dialogue (McFarland, 2001; Rochet-Capellan and Fuchs, 2014; Wlodarczak and Heldner, 2020). The results in the literature suggest that breathing planning during speech depends on syntactic and acoustic targets.

As flexible as the respiratory system is, it has some limits. Since breathing is inherent to speech, speech is also, to some extent, constrained by breathing capacities (see also Chapter 2 section 2.2.4). When breathing capacities decline, with age for instance, speech changes (Huber, 2008; Hoit and Hixon, 1987; Sperry and Klich, 1992): older adults need higher lung volume to start speaking, but also higher volume per breath group and syllables. Consequently, elderly people produce shorter breath groups and have a lower speech rate than younger adults when reading or during conversation (Huber, 2008). Finally, the entanglement between speech and breathing also depends on the cognitive load induced by the speech formulation steps. For instance, Mitchell and colleagues (1996) reported that preparing a speech before giving it resulted in more syllables per breath group and increased speech rate, but also decreased volume excursion, compared to no preparation.

As seen in Chapter 2 Section 2.2.4, limb movements have an effect on ventilation that may affect speech breathing. This effect is particularly well known in everyday physical activities and described in studies investigating speech production during physical effort induced by leg motion (locomotion and cycling) (Baker et al., 2008; Doust and Patrick, 1981; Rotstein et al., 2004). During physical effort, some coordination patterns between breathing and limb movements emerge, and could interact with speech breathing. The next section reviews research on the motor-respiratory coupling.

2.3.2 Breathing while moving: the motor respiratory coupling

The underpinnings of motor respiratory coupling (MRC) are biomechanical, neurophysiological and informational (Bardy et al., 2015; Hoffmann et al., 2012). The biomechanical impact of the limb movements on the chest is particularly at work during the locomotor respiratory coupling (LRC) in birds and mammals (Lee and Banzett, 1997; Bramble and Carrier, 1983). The abdominal viscera move up and down (on the longitudinal axis) during jumping or walking, due to vertical impulses generated by the ground forces on the feet, acting as a piston on the diaphragm. This piston effect can be caused by a spinal flexion: the trunk is shortened by flexion of the lumbar, moving up the viscera and the diaphragm. The visceral piston hypothesis has been spotlighted by most of the first studies on LRC, especially for animals (Bramble and Carrier, 1983; Bramble and Jenkins Jr, 1993). Quadrupeds undergo a heavier constraint of locomotion on the chest, and generally run at a 1:1 ratio between breathing and stride frequency. The respiratory system is less constrained by bipedal locomotion, and the locomotor respiratory coupling usually occurs at 1:2, 1:3 or 1:4 ratios. Although upper limbs do not trigger such a cascade reaction, pectoral, intercostal, and abdominal muscles are used for both locomotion and breathing in case of upper limb locomotion. Upper limb MRC has been observed for some species of birds (Butler and Woakes, 1980; Thomas and Suthers, 1972), and in the human activity of rowing (Mahler et al., 1991b; Mahler et al., 1991a). Amazeen et al. (2001) investigated the motor respiratory coupling during wheelchair propulsion, and observed locomotor-respiratory coupling at small integer frequency ratios (2:1, 3:1, 4:1), along with the consistent occurrence of the inhalation onset just after the onset of the movement cycle. They also noticed an effect of expertise on MRC: experts were more likely to display synchronization locking than novices.

Partly because of these biomechanical constraints, one of the reasons for MRC to occur would be to optimize energy expenditure and oxygen resources and consumption during physical exertion. For instance, LRC is stronger while running then when walking (Bechbache and Duffin, 1977). Hoffmann and colleagues (2012) found that increasing the strength of LRC decreased oxygen consumption. Garlando and colleagues (1985) also observed a negative correlation between the strength of LRC and the oxygen uptake at a moderate workload, but this result is participant dependent. The relationship between MRC and energy consumption optimization is not clear: Bernasconi and Kohl (1993) found that for a given workload while running, LRC decreased oxygen uptake, but LRC was not stronger with increasing workload. Rassler and Kohl (1996) observed that although the strength of LRC increases with increasing stride frequency during walking at a constant workload, this synchronization does not reduce oxygen uptake. Villard and colleagues (2005) hypothesized that if LRC was occuring for ressource allocation reasons, LRC should decrease with the increase of oxygen supply, i.e when breathing abilities were improved. They did not find any effect of improved aerobic abilities on LRC, suggesting that reasons for LRC to occur are not purely physiological.

One more reason to think that MRC does not occur only for energy cost optimization is that MRC also emerges from situations not necessitating high metabolic resources. MRC occurs during forearm tracking movements (Ebert et al., 2000), finger tapping (Wilke et al., 1975), wirst movements (Temprado et al., 2002) and even some evidence of MRC with eye movements has been observed (Rassler and Raabe, 2003). Studies on animals suggest that LRC comes from the common control of breathing and locomotion through hypothalamic (Eldridge et al., 1981) and medullar (Romaniuk et al., 1994) commands. This motor dependent control of breathing still has to be confirmed for humans.

A third channel modulating MRC lies in processing sensory stimuli. For instance, during walking, the optical flow perceived by our eyes is used to control the balance and sway of the body (Bardy et al., 1999). The perceptual channel through which coordination patterns emerge has been highlighted in inter-personal coordination (Schmidt et al., 1990; Richardson et al., 2007): when two or more participants perform the same movement at the same time, like swinging their lower legs, seeing each other leads them to converge to the same movement frequency. Perception of external information can also help MRC: displaying upward-downward motion of a ball controlled by forward-backward arm movement of the participants, or displaying the inflation-deflation of a balloon controlled by their inhalation-exhalation, improved the stability of certain frequency ratios of MRC (Hessler et al., 2010); the presence of an auditory stimuli at the respiratory or cycling frequency improved the stability of MRC (Hoffmann et al., 2012; Bardy et al., 2015; Hoffmann and Bardy, 2015). As Bardy and colleagues (2015) mention, mechanical and neurophysiological interactions together with informative stimuli take part in the perception-action loop underpinning motor-respiratory coupling.

The MRC patterns are susceptible to change with time and experience. While rowing (Mahler et al., 1991b) and running (Bramble and Carrier, 1983), LRC is stronger for experts than novices. Amazeen et al (2001) compared trained participants and novices in wheel chair propulsion, and noticed the same difference. When practicing an activity and becoming an expert, learning is characterized by a progressive integration of the subsytems implied in the action into effective task-specific organizations (Bernstein, 1968). These transitions can result in a reduction of redundant dimensions (degrees of freedom) implied in the movement (Mitra et al., 1998), which lead to more efficient actions. This reduction of dimensionality could result in coordination patterns between certain end effectors. Following this idea, Huys et al. (2003) investigated the emergence of MRC during juggling over 20 daily sessions of one hour, but did not observe any consistent coupling between breathing and juggling. These findings suggest that MRC plays a role in learning only for activities demanding a physical effort, and in which breathing and movement biomechanically interacts.

The motor activity is considered to be the leading oscillator and to attract the oscillator referring to the respiratory system. This is also suggested by the neurophysiological afferent connections from motor to breathing subsystems. Rassler and colleagues (1999) tested whether movement frequency could adapt to breathing frequency by increasing C02 in the blood and provoke hypercapnia, so that the respiratory system prioritizes hypercapnia treatment before being synchronized to movement. They found that breathing still adapted to movement. However, Bardy and colleagues (2015) recently demonstrated that although asymmetrical, the motor respiratory coupling was bi-directionnal: shifting the respiratory frequency had an effect on the cycling frequency.

Box 2. The non-linear dynamic framework applied to the motor-repsiratory coupling

Breathing and rhythmic motor activities have been found to have some period of synchronization. For instance, over a certain period of time, participants consistently produce n movement cycles for m breathing cycles. Breathing and movement are coupled on a $\frac{n}{m}$ ratio. The motor respiratory coupling is generally characterized by a multi-ratio coordination: the synchronization ratio between breathing and movement frequencies is not constant, and can switch from one ratio to another. Regarding the directionality of the coupling, most of the neural and behavioral findings suggest that the motor activity leads the breathing mechanism, but recent investigations suggest that this exchange is bi-directional (Hoffmann and Bardy, 2015).
The paradigm usually used to detect synchronization within the motor respiratory structure is to control the pace of either the movement or the respiration, and to test whether the other component adapts to the first one in its frequency and/or phase. The relative phase is the difference between the phase of one component within its own cycle and the phase of the other component also within its own cycle, at each time point. If the relative phase is constant, the coordination is absolute. When the system switches from one attracting state to another (for instance when the breathing-motor synchronization switch from one ratio to another), the relative phase abruptly switches from one value to another, before stabilizing again at a new ratio. When the relative phase continuously drifts, there is no phase locking, hence no synchronization. Studying the profile of the relative phase over time and its fluctuations enables the evaluation of the motor respiratory coupling. The difficulty of evaluating the coupling of the motor respiratory system lies into the relativity of its coordination. As Amazeen et al. (2001) mention, the coupling strength may be too low to maintain phase locking, and the system goes through several stable states with different coupling ratios and alternations between synchronization and desynchronization phases over time. Indeed, limb movements and breathing have different natural frequencies and rarely produce steady oscillations, implying constant modifications in their interaction. More methodological details are given in Chapter 5.

Most of the studies analyzing MRC control the motor frequency by giving a pace to the participant with an external stimuli (see Box 2). Few of them assessed MRC at the spontaneous rhythms of the participants. Hill et al. (1988) recorded participants walking at their own pace, and observed an intermittent entrainment between motor and breathing effectors, lasting less than 10 breaths at a time. However, they compared the variability of phase coupling during spontaneous and intended synchronization, and found that they were similar, suggesting that at a comfortable rhythm, LRC does emerge, but in a more intermittent form.

Besides, MRC remains highly participant dependent. The MRC strategies seem to vary from one individual to another. For instance, inhalation onset were consistent within each rower but varied across rowers (Mahler et al., 1991b). Loring and colleagues (1990) also noticed that LRC was stronger for some participants than others. The specific behaviour of each participant should therefore not be overlooked.

2.3.3 Why study interactions between speech, movement and breathing?

The interdisciplinary review of this chapter reveals that limb movements and speech share cognitive and physiological processes, with manifold links and adaptations. These links and mutual adaptations are, however, still not well understood: (1) The links between speech and arm vs.leg movements were discussed separately, as well as gestures vs. non-communicative motions; (2) Speakers' specificities such as breathing properties in speech production have not been investigated; (3) Fine-grained kinematic studies of movements have rarely been integrated in detailed linguistic analyses; (4) Until recently, breathing has been a neglected actor in speech-limb studies. Most of the studies in the literature have so far analyzed it either in the context of limb motion or in the context of speech production. Developing an integrated approach to these different levels is now required to embrace the complexity of the links between spoken language and limb movements. Combining breathing to limb and speech studies could reveal important insights into how speech and limb movements work together. The aim of this thesis is to introduce breathing as a marker of the interactions between speech and limb movements, and bring new empirical insights regarding the implication of limb motion in spoken communication based on a fine-grained analysis of the speech-breathing-limb movements' link in a natural language task and over several days. The following questions will be addressed:

(1) To what extent do communicative and non-communicative movements have a different impact on speech? (2) How does natural speech impact non-communicative movements? (3) How does speech impact motor respiratory coupling? (4) Is speech breathing speaker-specific, and is this speaker specificity maintained across different limb movements? (5) Do the interactions mentionned in the previous questions change over time?

The interactions between speech, breathing and limb movements will be analyzed with methodologies coming from different fields, using tools enabling quantitative analyses on the three signals altogether.

To work on the questions raised above, as detailed in Chapter 3, we designed an experiment in which spontaneous speech (recalling short stories) co-occurs with four different limb activities: spontaneous co-speech gestures, no gesture, biking with the arms or biking with the legs on a mini-bike. Cycling motions were chosen to create a dual-task paradigm with speech and limb movements while avoiding speech-gesture coupling due to speech related cognitive load. They are relatively automatic, and do not require a large attention span, spatial navigation or visual control that could interfere with speech. The use of the mini-bike enabled the comparison between upper and lower limb activity and avoided as much as possible effects due to physiological reactions regarding physical effort. To minimize physical effort, participants sat on a chair.

Chapter 4 focuses on the interactions between speech and limb movements. The following points will be investigated : (1) the impact of task repetition on the frequency of hand gestures; (2) the effect of the different limb conditions on content recall and speech fluency; (3) the impact of speech on biking movements; (4) the impact of physical impulse of biking movements on speech acoustics.

Chapter 5 details the analyses of the breathing cycles related to the different speech and limb conditions. The methodologies to analyze breathing are taken from physiological studies. The following points will be addressed: (1) the impact of the different speech and limb conditions on breathing parameters; (2) The existence of a speech breathing individuality and its consistency over days and limb movements; (3) The motor respiratory coupling over the different speech and limb conditions and over time.

Chapter 6 discusses the results of chapter 4 and 5 by replacing them in the theoretical background presented in the two first chapters, and questioning the data-driven approach, and the assets of fitting studies about co-speech gestures and limb motor control together.

Chapter 3

Recording of the speech-breathing-limb motion dataset

One of the objectives of the SALAMMBO (Spoken lAnguage in motion: Learning and Adaptation of speech coMMunication in the context of BOdy motion) project was the constitution of a dataset from speakers of German and French in a narrative task under different movement conditions. The data collection consisted in recording limb movements, breathing and speech together over several days, in order to assess the stability of the profiles and their evolution with task familiarisation. In this context, an experimental method to build this dataset was developed. Speech, breathing and limb movements were recorded under different movement conditions: (a) still; (b) with free arm movement; (c) with hand pedalling movements; (d) with leg pedalling movements.

Speakers were recorded while retelling short stories that we created such as to elicit gestures. Creating the support for the narrative task and setting-up a recording procedure over several days were the first steps of the thesis, from October 2018 to March 2019. These steps are detailed below. One of the big technical stakes was to record speech, breathing and limb movements synchronized altogether. The Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) had an experimental platform already equipped for these challenging recordings. Hence, these recordings took place at the ZAS in Berlin, with 25 native speakers of German, between March and June 2019. As part of the workshop on recording and measuring motion of the articulators, limbs and the respiratory system, organized by the Linguistic research Infrastructure in Zurich, we recorded a tutorial on the Etisense plethysmography system (advantages and limits) and how to record with an Optitrack system, the Etisense plethysmography system and audio in synchrony. You can find our tutorial here: https://osf.io/dx2fy/?view_only=0a445ae953a04274853dc1d104c1fdec.

3.1 Overview of the study

Participants' main task was to watch stories and recall them in different conditions of limb movements. This task is referred to as "a narrative task". It was done in four main conditions in a within-subject design:

- *speech_ armFree*: participants are free to gesture with the hands
- *speech_ armBlock*: participants have to put the hands under the chair
- *speech_ legMot*: participants are biking with the feet on a mini bike
- *speech_ armMot*: participants are biking with the hands on a mini bike

To answer our questions, we decided to design an experiment in which speech co-occurs with spontaneous co-speech gestures, no gestures, as well as biking with the hands on a mini bike, and biking with the legs on the same mini bike. Cycling motions on a mini bike were chosen to enable the comparison between upper and lower limb activity. Movements are of the same amplitude and shape for both limbs. They were also chosen to avoid the strong coupling that may occur between flexion-extension motion of the arm such as that used by Pouw et al. (Pouw et al., 2019; Pouw et al., 2020b). As the focus is on the motor and biomechanical nature of limb movements, confounding interactions due to physiological reactions regarding physical effort, and cognitive interferences at a higher level of movement planning were avoided as much as possible. Cycling movements are relatively automatic. They do not require a large attention span, spatial navigation or visual control. To minimize physical effort, participants sat on a chair, cycling on a mini-bike with no resistance. We chose a within-participant design to: (1) decrease the required number of participants; (2) limit the bias that could be due to the sample: in a between-subject design, the participants' characteristics in terms of cognitive capacities or movement skills could have had different levels from one condition to another, which would have impacted the results; (3) to be able to evaluate the consistency of inter-individual variability between the conditions, and correct it if necessary; (4) minimize the inter-participant variability between the conditions. During the different tasks, speech, breathing and limb movement were recorded synchronously. The study took place in 3 sessions on different days. The material and procedure are detailed here after, as well as the experimental design.

3.2 Participants

Twenty-five healthy participants (19 females and 6 males) aged 20–29 years (mean = 23.3 years; standard deviation (SD) = 2.5 years) were recruited using email announcements sent via a study participant database. All participants spoke German as their native language and reported no respiratory, motor, neurological, speech, or hearing pathologies. They signed a consent form and received 10€ per hour for their participation.

3.3 Material for the narrative task and pre-tests

A real-time story telling paradigm was chosen (McNeill, 1992). Many studies used this paradigm to investigate spontaneous gestures of participants during a narrative task, presenting short cartoons (Cassell et al., 1999; Pouw et al., 2020a) or comics strips (Stevanoni and Salmon, 2005; Cravotta et al., 2019) without any verbal content. It has been shown that gestures are more elicited by visual content than read or verbal content only (Hostetter and Hopkins, 2002). This effect seems to be due to a stronger spatial representation in the visual content conditions (Hostetter and Skirving, 2011). In our study, the insertion of unknown pseudo-words in the stories was needed to have an objective evaluation of content recall. A pseudo-word is a word which does not exist in the language concerned but respects the phonotactic rules of this language.

It has been shown that during story-telling, gestures are elicited by explaining spatial information units, and that limiting the access to lexical units when participants see the story (providing them with only pictures and no word or text) does not increase their gesture production (Hostetter and Skirving, 2011). The support chosen was visual as well a verbal, to respectively elicit gestures in the *speech_ armFree* condition and insert unknown pseudo-words in the text.

Because of the within subject-design, participants had to watch a different story for each condition. Creating new stories seemed the best option (compared to reusing cartoons from the literature) to control the content in detail so that the stories were balanced and could be randomized over the conditions, and to reduce bias related to previous knowledge about existing stories. Four stories were created.

3.3.1 Text of the stories and audio-visual support

As the study was also set-up to evaluate the short-term and long term memory performances of the participants, stories needed to be neither too easy (to avoid a ceiling effect) nor too difficult (to avoid a floor effect) to memorize. The structure and content was constrained by the fact that the SALAMMBO project also aimed to study the learning of unknown words in a narrative context. These new words were conceived as names of categories for items inserted in the stories: character-vehicle-house.

The stories were first created in French, using the same structure for each of them. The number and type of action units, of characters and the grammatical construction of the narration were controlled (see Table 3.1). The stories were translated in English so they could be understood by our German colleagues, who translated them to German.

Each story was about an alien arriving on earth and attempting to build a house. The stories' structure was based on Mandler and Johnson's work (Mandler and Johnson, 1977) (see Figure 3.1). The number of details in the descriptions and actions was precisely controlled, and all stories describe actions with similar narrative content. The scripts were then put on video with a voice telling the story while pictures scrolled by. The stories were presented as short animations. One story was uploaded to the OSF repository as an example (https://osf.io/w6nq7/).

The shape of the main characters and objects of the story were also selected to elicit gestures (see Table 3.1). They are presented in shades of gray to avoid a salient effect due to the color of an object or character. The main elements of the stories (character, vehicle, house) with the support of the pseudo-words and their pictures were created by Estelle Gillet-Perret, a speech pathologist, and a Master student as part of two research projects on the role of gestures in children's learning of new words (GeSLI and ComEns projects).

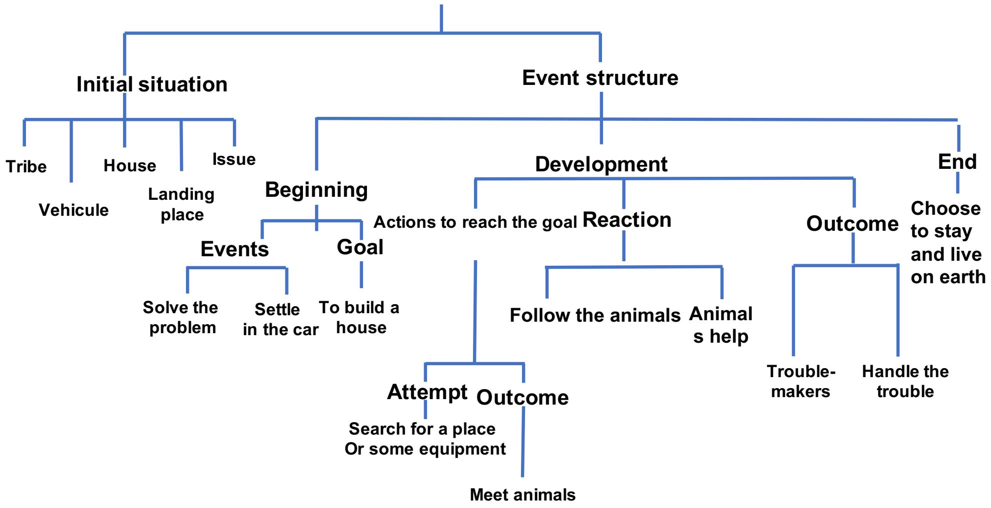


Figure 3.1: Structure of the stories, inspired from Mandler and Johnson (Mandler and Johnson, 1977).

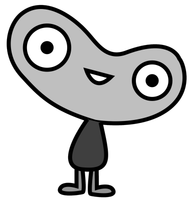
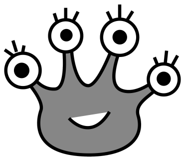
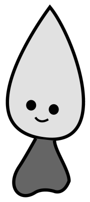
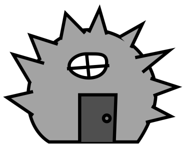
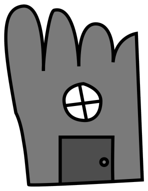
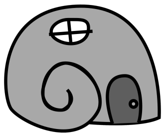
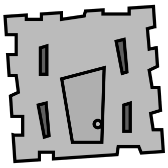
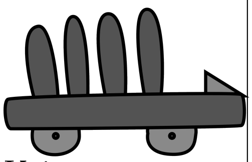
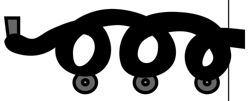
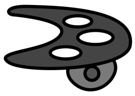
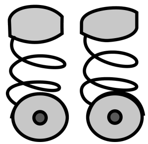
Element	story 1	story 2	story 3	story 4
Main character	 Mielbete	 Tereinat	 Soltete	 Lasgelich
House	 Sonistik	 Melare	 Wecktelin	 Mattendich
Vehicle	 Keimase	 Ligete	 Madikte	 Zulerge
Duration (minutes)	2.25	2.25	2.30	2.33
Number of words	292	264	291	300
Number of pictures	25	24	26	31

Table 3.1: Characteristics of the stories: pseudo-words and associated images for the alien, the house and the vehicule; duration of the story; number of words and pictures.

3.3.2 Generation and selection of pseudowords

In order to have an objective measure of memory performance, pseudo-words were inserted in the stories. They were created based on the methods described in Appendices B and C. Since the retrieval of the pseudo-words have not been studied in this thesis, their creation is not detailed here.

3.4 Experimental set-up and data acquisition

The challenging part was to set-up a multimodal recording with a proper coordination between the recording systems (audio, breathing, motion) to be able to synchronize the different signals in the post-processing. The set-up is represented in Figure 3.2 and a picture of a participant in the main experimental conditions is provided in Figure 3.4.

3.4.1 Set-up

The experiment took place in a dedicated 10m^2 room. To capture limb motion, speech and breathing at the same time, three recording systems were synchronized together. Each of them is described in the following sections, as well as their synchronization mechanism. The following equipment was also used:

- two desktop computers represented on Figure 3.2. Both computers were running under windows 10®. Computer 1 controlled the motion capture system. Computer 2 drove and combined the breathing recordings, the audio and synchronization signals.
- an acquisition card receiving the audio, breathing and synchronization signals and sending them to computer 2;
- a synchronization box, receiving a binary signal (0 when recording was off and 1 when recording was on) from the motion capture system and sending it to the acquisition card;
- a laptop Macintosh® computer displaying the story videos. The laptop was placed on a table in front of the participant each time the participant had too watch a story, or fill in a test;
- a rolling table. This table was used to carry either the laptop to present the material to the participants, or the bike when participants had to bike with the arms;
- a mini-bike SportPlus® for the cyclic movements of the arms and legs in the corresponding conditions;
- a camera to get a complete picture of the participants' motion, to detect any problems that may occur during the recording (sensor detachment, movement, etc.) and to observe motions that are not tracked with the Optitrack system for further analyses. The camera was on a stand, at a height of 120 cm, at one meter in the left diagonal of the participant.

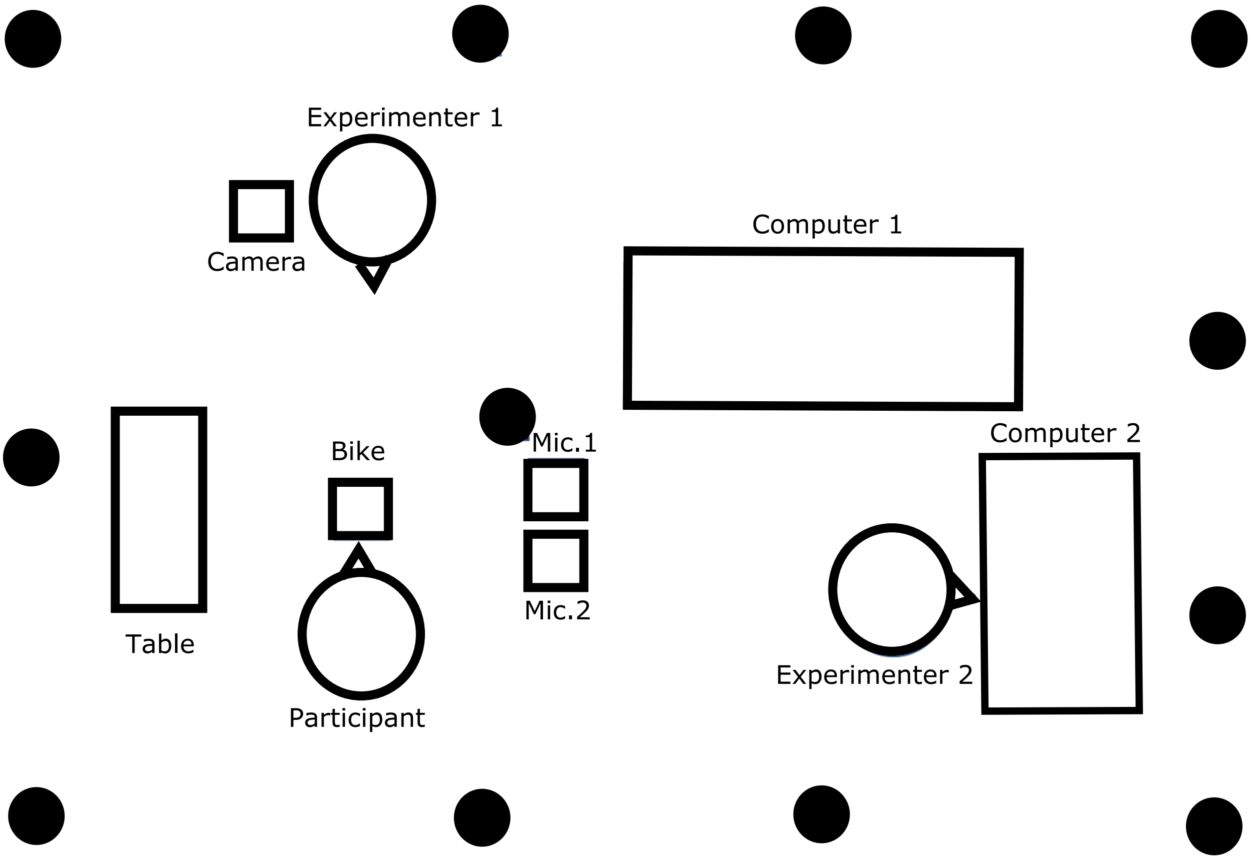


Figure 3.2: Set-up of the experimental room when participants are biking with the legs. The black circles are the motion capture cameras. Mic = microphone.

3.4.2 Recording of limb movements

3.4.2.1 The Optitrak system

The limb movements were recorded using an Optitrack® motion capture system from Natural Point®.

Twelve infrared cameras model Prime 13® were settled all around the room (see Figure 3.2). 11 cameras were located at the junction between the walls and the ceiling, surrounding the room, and one camera was fixed on a stand, at 70 cm in the right diagonal of the participant. The system was handled on computer 1, through the software Motive® 1.9.

This motion capture device allows the recording of the three dimensional trajectories over time of markers placed on the participants. The markers consist in spheres covered by reflective material and attached to the body parts by a velcro strap, or stuck on a velcro jacket (see figure 3.3). The cameras emit infrared radiation, which is reflected by the markers and sent back to the cameras.

Through the angle captured by each camera, the system makes a 3D reconstruction of the scene, and records the coordinates of the trajectories of each marker at a certain rate, in this 3D space. The trajectories of the markers are saved in a csv file, with a column for each dimension (three columns for one marker, then three columns for another one etc.).

With the Optitrack® motion capture system, when a marker is not visible over a certain time interval, this marker is considered as a new one when it reappears. This issue requires a post-processing that consists in finding all the trajectories created for the same marker and merge them. Reconstructing the trajectories of all markers becomes excessively time-consuming. To avoid loss in marker tracking, rigid bodies can be used. Rigid bodies are made up of three markers always at equal distance from each other (see figure 3.3). Before being tracked, the rigid body is labelled with its three markers, so that when any marker is occluded, the rigid body is still tracked with its other visible markers. More than one marker must be visible to maintain the tracking of the rigid body. The final coordinates provided by the software are the ones from the center of gravity of the rigid body, computed with its markers. In addition to the positional coordinates, the system also provides the orientation coordinates of the rigid bodies (six degrees of freedom).

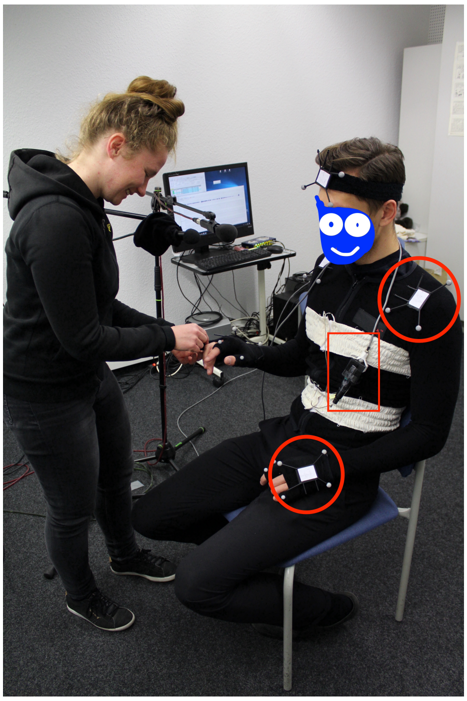


Figure 3.3: Example of a participant being equipped with the motion capture jacket and gloves and the belts of the inductotrace plethysmography system. Rigid bodies are inside red circles. They are stuck on a velcro jacket. The oscillator of the inductance plethysmography system is inside the red rectangle.

3.4.2.2 Capture of arm and leg movements

The Optitrak was used to track the body motion of the participant and the cyclic movements of the bike.

The participant was fitted with a headband and a black jacket on which the rigid bodies and markers were scratched as follows (see Figure 3.3):

- One rigid body on the head to track head motion;
- One rigid body on each shoulder and one on the back. They allow in particular to analyse head movements in a body reference frame;
- One rigid body on each hand plus individual markers on each thumb, index finger and little finger;
- One marker on each pedal of the bicycle.

The sampling rate of the motion tracking was set at 200 samples (frames) per second. The positional coordinates over time of the motion capture markers and rigid bodies' centers were extracted in a csv format. Video recordings were used to get a complete picture of the participants' motion, to detect any problems that may occur during the recording (sensor detachment, movement, etc.) and to observe motions that are not tracked with the Optitrack for further analyses. The camera was on a stand, at a height of 120 cm, at one meter in the left diagonal of the participant (see Figure 3.2). Examples of video recorded images can be seen on Figure 3.4.

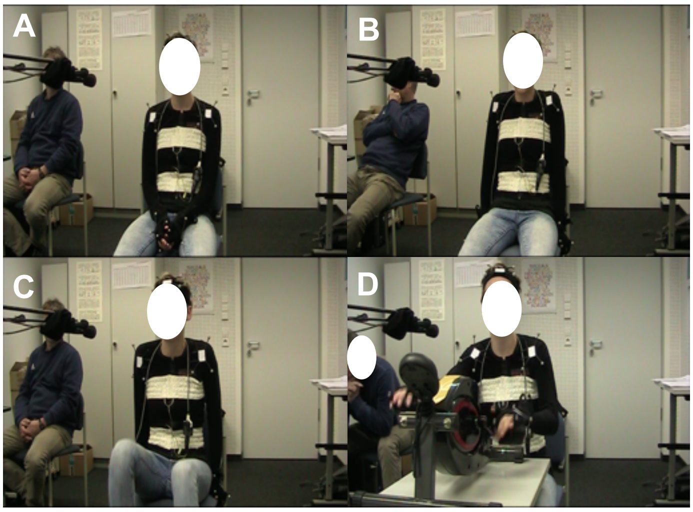


Figure 3.4: Screenshots of a participant performing the four conditions. A: *speech_ armFree*; B: *speech_ armBlock*; C: *speech_ legMot*; D: *speech_ armMot*.

3.4.3 Recording of breathing

3.4.3.1 The inductotrace inductance plethysmography

Breathing was recorded using a non-invasive system for measuring respiratory movements: the Inductotrace inductance plethysmography system (Ambulatory Monitoring Inc, New York, USA). Plethysmography is a procedure to measure changes in volumes in different parts of the body. The flow of an electric current going through a circuit creates a magnetic field around the wire conducting the current (the conductor). This magnetic field, also called self-inductance, changes depending on the area encircled by the conductor. The circuit is connected to an oscillator module through an electrode and an interface cable to obtain digital waveforms. The monitoring of breathing at rest or even during exercise via a respiratory inductive plethysmography system has been validated by several studies (Caretti et al., 1994; Clarenbach et al., 2005). Respiratory inductive plethysmography has been medically validated as a method to diagnosis different diseases from respiratory disorders.

The Inductotrace system has been widely used in research to estimate the current breathing volume over time as a non-invasive method so the participants can speak at the same time (Fuchs et al., 2008; Huber, 2008; Rochet-Capellan and Fuchs, 2014; McKenna and Huber, 2019; McFarland et al., 2020). Via elastic belts placed on the abdomen and the thoracic cage of the participant, the inductotrace system records the variations of the circumference of thoracic and abdominal volumes: insulated wires are sewn in each belt. Because the thorax and the abdomen are independent breathing volumes (Konno and Mead, 1967), one belt for each unit is necessary in order to record the respiratory volume. During inhalation and exhalation, the cross-sectional areas of respectively the rib cage and abdomen expand and compress, altering the self-inductance of the coils and the frequency of their oscillation. The electronics convert this change in frequency to a digital respiration waveform. The amplitude of the waveform is proportional to the inhaled or exhaled breath volume.

The final output is two signals of one dimension, one for each volume measured (thorax and abdomen, see Figure 3.5), from which the following variables can be extracted:

- the respiratory rate;
- the air volume;
- the minute ventilation;
- the peak inspiratory flow;
- the symmetry of the cycles (inspiration time divided by total duration of a cycle).

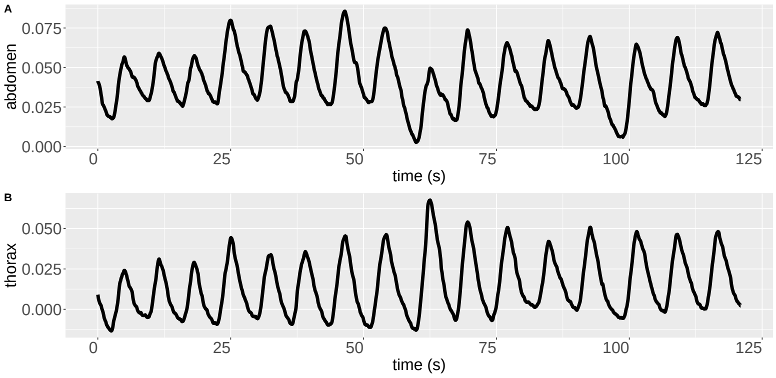


Figure 3.5: Output of the Inductotrace during a breathing recording session.

To enable the saving of the breathing and speech signals in the same file, the sampling rate of the Inductotrace was set at 44100 samples (frames) per second, like the speech signal. The two signals were compressed in an edt (Adobe eBook Download Settings format) file, together with the synchronization signal and the audio signal, through the software EdWin®.

3.4.3.2 Monitoring of breathing

Above the jacket of the motion capture, the abdominal and thoracic belts of the Inductotrace plethysmography system were placed to record breathing kinematics (see Figure 3.3). The belts were connected to an oscillator module hanged around the neck of the participant (see Figure 3.3), plugged in computer 2 through a data acquisition device DIC6B from DataRec®.

3.4.4 Speech recording

Speech was recorded using two stand-up microphones (MKH 50 P48, Sennheiser). The first one was connected to an ULTRAGAIN DI-BOX MODEL MIC 2000 sound card and was sampled at 16000 Hz. The mouth–microphone distance was 50 centimeters. Its signal was transmitted to computer 2. The second microphone was connected to the motion capture system and could be used for synchronization between the two computers signals. It was sampled at 44100 Hz. The signals recorded by the first microphone and the Inductotrace (abdomen and thorax) were sent to computer 2 through the acquisition device (see Figure 3.6), and saved in an edt (Adobe®eBook Download Settings format) file. The signal from the second microphone was stored in a wav file in computer 1.

3.4.5 Synchronizing the three set-ups

The material described here is represented in Figure 3.6. The challenge was to synchronize three different acquisition hard-wares. The Esync®, a box dedicated to the synchronization of the motion capture system with other devices, was connected to the switch of the cameras. This box was receiving a constant binary signal from the motion capture, and was sending it to the acquisition device on computer 2. The signal was equal to one when the motion capture was recording, and 0 else, and was stored in the same edt file as audio and breathing signals. Movement acquisition started after the onset of breathing acquisition, so that there is a visible jump from 0 to 1 on the synchronization signal, indicating the start of the movement recording. The first step of the data processing was to open the edt files and to discard the first values of the speech and breathing signals for which the value of the synchronization signal was equal to 0. This was done in Matlab®. The main issue encountered with this method is that the clocks of the different devices may not be exactly the same and could progressively deviate from each other. We investigated the potential lag between the motion capture device and the acquisition card by applying a cross-correlation analysis to the audio signals coming from the two microphones. In average, the lag between the two was about 10 ms. Another potential issue is the fact that the samples of the different devices do not occur at the same time. This can be the case if the recordings are not starting exactly at the same time, or if the different sample rates of each device are not the multiple of one another. In our experiment, the motion was recorded at 200 Hz, the first audio sample rate was set at 16000 Hz, and the second one at 44100 Hz. As 100 is a divider of each sample rate, we down sampled every signals at 100 Hz. Because of the lag between the clocks, the samples may not have occurred at exactly the same time. The interval duration between each sample can be up to 10ms.

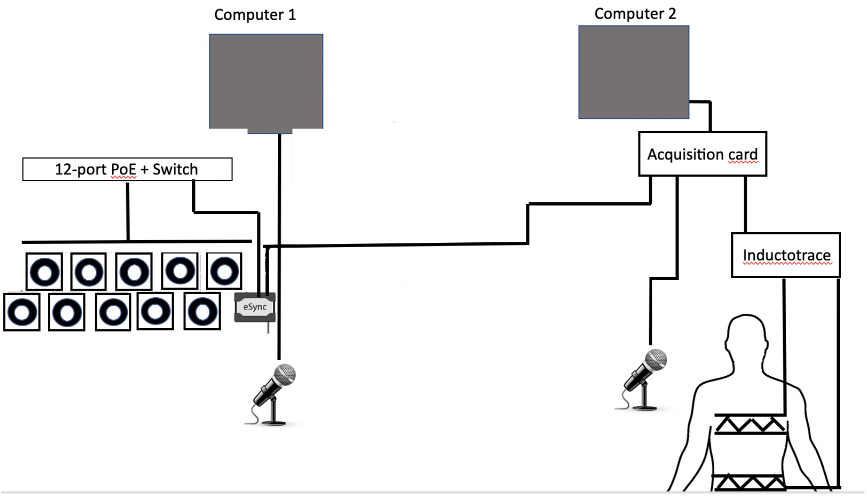


Figure 3.6: Illustration of the different units of the data acquisition and their connections.

3.5 Experimental procedure

The study was longitudinal over 3 sessions: the first day, the next day, and 10 days later. The three sessions respectively lasted 1 hour and 30 minutes, 1 hour, and 30 minutes. Each session has its own structure. They are summarized on Figure 3.7

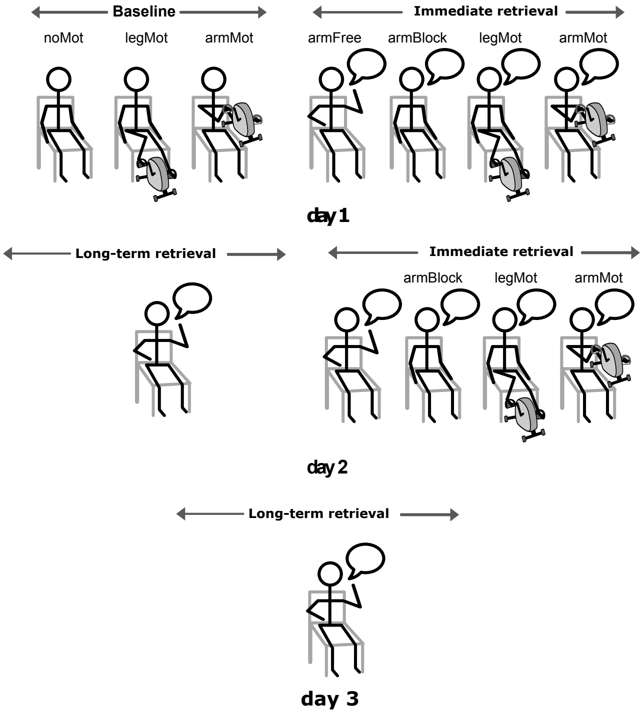


Figure 3.7: Participants' tasks according to the day. On day 1, From left to right: Quiet doing nothing (*quiet_ noMot*), quiet with legs biking (*quiet_ legMot*), quiet with arms biking (*quiet_ armMot*), speech with arms free (*speech_ armFree*), speech with arms blocked (*speech_ armBlock*), speech with legs biking (*speech_ legMot*), and speech with arms biking (*speech_ armMot*).

3.5.1 First day: installation, pre-tests and main task

The first day, after reading the information sheet and signing the consent form, participants were asked to fill in questionnaires about their daily physical activities: the IPAQ (International Physical Activity Questionnaire, 2003) to know more about their physical activity level. Other questions were about their sport, dance, music and theatre practice and on his or her educational level, as well as their frequency of smoking (see Appendix D). A working memory test of about 15 min was also performed (von der Malsburg, 2015) but at the end of the second day to reduce the duration of the first session. This individual-specific information was collected to investigate any potential correlation between life habits/working memory and kinematic, cognitive or breathing behavior. Participants were then fitted with the recording equipment: the jacket and the markers for the motion capture system, and the two respiratory belts of the breathing tracking system. While setting the equipment on the participant, the experimenter explained the audio, video, breathing and motion capture recordings that would be made, and gave the instructions to be read. The instructions presented the task and the conditions in which they were going to take place (see Appendix A). Still in the objective to elicit gestures, it was specified that the person listening to the narration was not speaking German well, and that it was important to make sure that this person would understand everything.

After reading the instructions, participants were placed on a chair in the center of the volume covered by the motion capture system. Breathing and movement was first recorded for two minutes in each of the following quiet conditions: at rest (*quiet_noMot*), biking with the arms (*quiet_armMot*), and biking with the legs (*quiet_legMot*) at a self-selected comfortable pace. This reference was used to characterize the movements without speech. Participants were given a short introduction of 2 min, to present the outline of the four stories, as well as the pseudo-words with their corresponding images.

The participant's main task consisted in watching the short animated films attentively, and retelling them just after each one, in one of the four limb conditions. They also had to fill in a test of explicit memorization of the pseudo-words. The free recall took place first. Participants had to retrieve everything he/she remembered, in particular the pseudo-words, descriptions and actions of the story. The explicit memorization test was divided in two parts: a naming test, during which participants saw the pictures of the alien, the vehicle and the house, and were asked to provide the pseudo-word corresponding to each picture. The order of presentation of the items was chosen randomly. The second test was an identification test: within a table, the pseudo-words were in rows, and the pictures were presented in columns. Participants were asked to tick the picture corresponding to each pseudo-word. The presentation of both the pseudo-words and the images was random, and differed from one participant to another (see Appendix F). For the sake of maintaining session duration as short as possible, the pseudo-words of the naming test were typed on the keyboard, and the pseudo-words of the identification were written on the screen.

3.5.2 Day 2: overall recall and repetition of the main task

Participants were asked to come back the next day (Day 2). The second day consisted in a free recall of the 4 stories in a sitting position, hands free, followed by the explicit memorization tests of the 12 pseudo-words. Participants then went through the main task again, following the same procedure as on the first day (see Figure 3.7).

3.5.3 Day 3: overall recall

Participants were asked to come back ten days later (Day 3). They had to recall the 4 stories freely, followed by the explicit memorization tests of the 12 pseudo-words (following the same guidelines as at the beginning of day 2) (see Figure 3.7).

3.5.4 Counterbalancing of orders

In order to prevent potential interactions between story preference and conditions, the order of the presentation of the stories, conditions, and the presentation of a given story in a given motion condition were balanced as much as possible. There were 24 possible condition orders and 24 possible condition–story associations. For complete randomization, 576 participants would have been necessary, but such an endeavor was impossible in the time given. Twenty-four participants were planned such as to have each story presented six times in each condition and each position. One participant did not come back on day 2 and was replaced by another one who also did not come back on day 2. We decided to keep the data of these two participants, which slightly unbalanced the number of times a given story appeared in a given condition (six or seven times) and the number of times a condition was presented at a given position (five to eight times).

The data collected in this study were analyzed to describe the interactions between speech, breathing and limb movements. Each actor was characterized by measured variables already defined in the literature. These variables are presented below.

3.6 Characterization of breathing, motion and speech

3.6.1 Characterization of breathing

Breathing was recorded by inductance plethysmography. The output of a recording was one signal characterizing the variations of the thoracic cage and one signal characterizing the variation of the abdomen (see Figure 3.5). These signals were embedded in a edt file, with the speech and the synchronization signals. First, through a Matlab script, each breathing signal was extracted from the edt file, and downsampled at 100Hz to reduce the time of file processing during the analyses, and optimize storage space. The respiratory signals of the thorax and abdomen were low-pass filtered at 10 Hz. To convert the changes in the thoracic and abdominal sections recorded by the plethysmography system to overall relative change in lung volume, a transformation is required. We used the method suggested by Banzett

(1995), who evaluated the contribution of thoracic and abdominal circumference changes to lung volume to be: $2 * thoraxdisplacement + 1 * abdomendisplacement$. According to Mckenna and Huber (2019), Banzett's method is more reliable than more complex methods based on isovolume maneuvers. Inhalation peaks and onsets were automatically detected with the Matlab function *findpeaks*, and then manually corrected after visual inspection and the help of the audio signal. The number of breathing cycles per condition and day is reported in table 5.1.

Breathing profiles have been defined and recorded differently in various domains. Different features can be considered to characterize and analyze these profiles. We essentially focus on temporal parameters and shape. This choice is motivated by the methods used to record breathing and because these parameters allow to distinguish speech breathing from silent breathing. On the basis of previous work on speech breathing (Benchetrit et al., 1989; Fuchs and Rochet-Capellan, 2021), each breathing cycle was characterized by three variables, which are illustrated in Figure 3.8:

(1) The total duration of the cycle (*cycleDur*) measured from inhalation onset to next inhalation onset, expressed in seconds (Fig. 3.8.A). This variable has been taken as a proxy for cognitive and physical demands (Mitchell et al., 1996).

(2) The symmetry of the cycle (*cycleSym*), computed as the ratio of inhalation duration (*inhDur*) divided by total cycle duration (*cycleDur*) (Fig. 3.8.A). The symmetry is a typical feature of breathing in speech production in comparison with quiet breathing (Conrad and Schoenle, 1979).

(3) A multivariate variable introduced in previous work on the "ventilatory personality" and representing the shape of the cycle (*cycleShape*).

Each breathing cycle can be described as a sum of an infinite number of cosine functions. The amplitude and phase angle of each cosine are obtained via the Fourier transform, that is, the frequency decomposition of each breathing cycle: each cosine represents one of the frequencies in the frequency decomposition. These frequencies are also called harmonics (see Figure 3.8.B). Each harmonic is thus associated with an amplitude and a phase, corresponding, respectively, in the signal processing terminology, to the modulus and argument of the Fourier transform for that frequency. According to Bachy et al. (1986), the breathing cycle at rest can be reliably reconstructed from the modulus and the argument of only the first four harmonics of the cycle (see 3.8.B and C) as they account for 95% of the original signal's power. We also used the modulus and the argument of the first four harmonics to characterize each breathing cycle. Even if less interpretable than univariate measures, this cycleShape variable is particularly interesting to evaluate the consistency of breathing profiles since speech breathing can be determined by multiple factors, such as the length of the upcoming sentence, prosodic phrasing, and prominent syllables. In order to calculate the cycleShape variable, each cycle was normalized in duration (1) and amplitude (2):

(1) The cycles were normalized to 64 points, corresponding to the largest and closest power of 2 to the number of points included in the shortest cycle (87 points) (Bachy et al., 1986).

(2) The amplitude of each cycle was normalized between 0 and 1 using the following formula, where y is the cycle signal:

$$y = \frac{y - min(y)}{max(y) - min(y)} \tag{3.1}$$

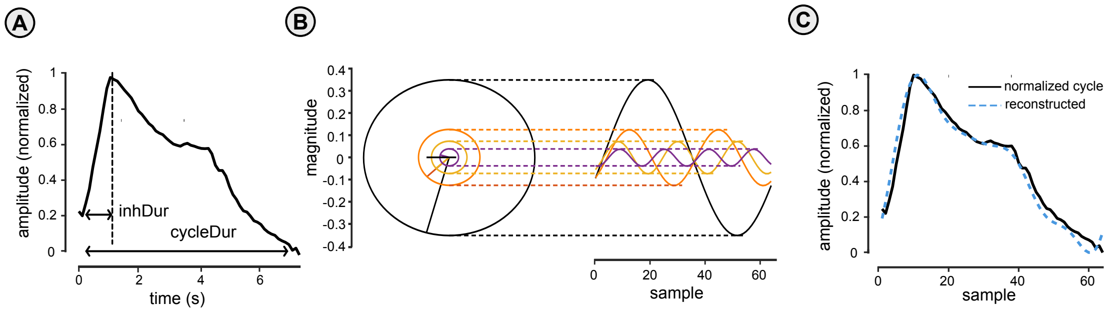


Figure 3.8: Characterization of the speech breathing cycle. (A) The measurement of the cycle duration (*cycleDur*). Symmetry (*cycleSym*) was obtained by dividing the duration of inhalation (*InhDur*) by *cycleDur*. (B) A schematic representation of the process to describe the cycle in A: the different cosine (waves) that are summed for the reconstruction are displayed. Their modulus (*amplitude*) and argument (*phase angle*) are both represented by the circles on the left: their amplitude corresponds to the radius of the circle, and their phase angle corresponds to the orientation of the radius. This vectorial representation does not have axes per se. (C) An example of a normalized cycle (solid line) and the corresponding reconstructed airflow profile from the first four harmonics (dashed line). The axes are unitless because of the normalization.This figure has been published in Serré et al. (2021)

Each normalized cycle was then decomposed in the frequency domain using the fast Fourier transform. The parameters of the first four harmonics (modulus and argument) were extracted to characterize the cycle (see Figure 3.8 part B). They can be used to reconstruct the cycle with a sum of four cosine functions, the frequency, amplitude, and phase angle which correspond to the frequency, modulus, and argument parameters described above (see Figure 3.8 part C) (note that because of temporal normalization, the frequencies were the same for all the cycles). On average, the first four harmonics accounted for

95 % (1st quartile = 0.92; 3rd quartile = 0.99; SD = 0.059) of the signal power for the cycles at rest and 90 % (1st quartile = 0.88; 3rd quartile = 0.97; SD = 0.11) for the speech cycles.

The amplitude was not exploited because we considered that since the belts of the plethysmography system can move with upper limb movements, this value was not reliable.

These variables are analyzed in Chapter 5 Sections 5.1 and 5.3

3.6.2 Characterization of limb movements

The limb movements involved in the experiment are free gestures (*speech_ armFree*), leg cycling (*speech_ legMot*) and arm cycling (*speech_ armMot*).

3.6.2.1 Hand gestures

Quantifying hands' motion based on 3D trajectories (*HandMvt*)

To compute the movement quantity achieved by both hands within a trial and model it as a function of the day, the 3D distance (in cm) of both hands were summed for each time point and the distances were averaged over the trial in the condition *speech_ armFree*. The movement data were extracted from 3D coordinates of the two rigid bodies attached to the participants' hands, which first required to identify correctly these coordinates in Optitrack files.

First, for each file, the rigid bodies of the hands were identified. Depending on the recording, the x-axis can be the depth, the vertical or the horizontal dimension. The identification of each axis was needed. This labeling was done by watching carefully each 3D scene through the function *plot3* of Matlab (see Figure 3.9). Each label was written in a csv file, assigning a filename to each row, and its corresponding axes and rigid bodies. This csv file was called in all the analyses in which the rigid bodies were involved.

Since some files have been mistakenly recorded at 120 Hz instead of 200 Hz, the vectors of coordinate of the rigid body on each hand were downsampled to 100 Hz to put all vectors to the same sample rate. The quantity of motion was measured as follows: to reduce the three dimensions to one, the euclidean distance of the three dimensions was computed for each row corresponding to a time point. Since the signals were downsampled at 100 Hz, 100 rows represent one second. The outcome was an unidimensional vector for each rigid body. To get the distance between two successive samples, each point was substracted from the successive point. To compute the average quantity of hand motion (*HandMvt*), the distance vector was averaged over each trial.

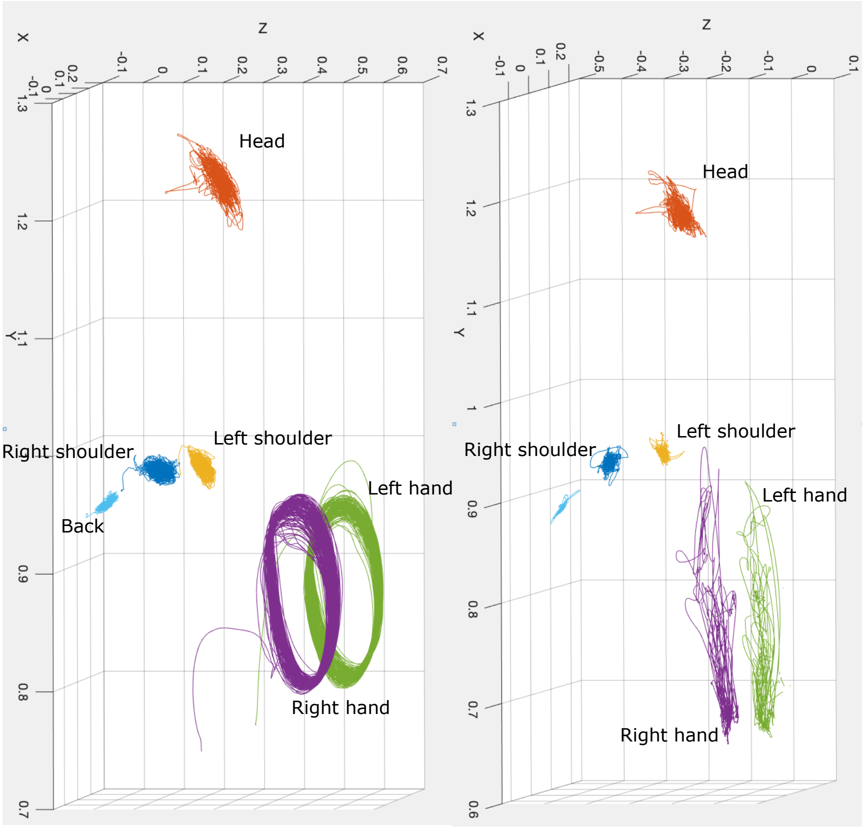
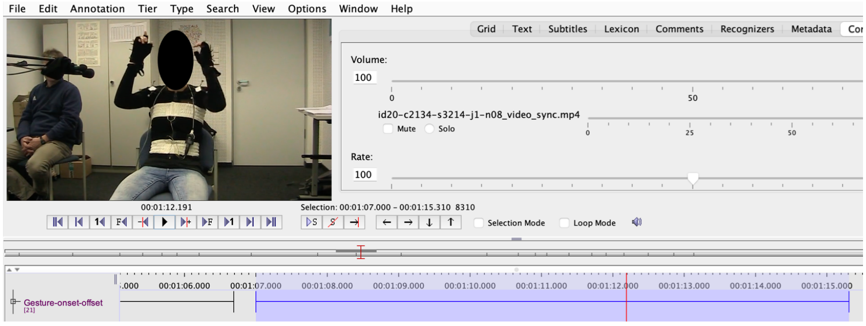


Figure 3.9: 3D scene (matlab plot) with the rigid bodies. Left: *speech_armMot* condition. Right: *speech_armFree* condition. red: head; clear blue: back; dark blue: right shoulder; yellow: left shoulder; purple: right hand; green: left hand.

Quantifying hands' motion based on ELAN annotations (*RateGest*)

As displayed on Figure 3.10, the videos of each session in which participants were free to gesture were annotated with ELAN (Lausberg and Sloetjes, 2016) by a research assistant who was instructed to detect the onset and offset of each hand gesture. Hand movements were considered as gestures if they were going along with speech. For instance, scratching movements (on the nose or the ear) were not considered as a gesture. Hand movements are not always discrete and distinct gestures, but are often continuous mere motor movements accompanying speech prosody (Chawla and Krauss, 1994). Because of this continuity in hand movements, counting the number of gestures per trial is not representative of the motion quantity of the hands. To get a more reliable measure of hand movement quantity and able to compare it to the motion capture measure (*HandMvt*), the rate of the total duration in which the participant move the hands over the total duration of the trial was computed. This variable was labeled *RateGest*.



[redacted]

Volume: 100

id20-c2134-s3214-j1-n08_video_sync.mp4

Rate: 100

00:01:12.191

Selection: 00:01:07.000 – 00:01:15.310 8310

Figure 3.10: Annotation of hand gestures through the application ELAN (Lausberg and Sloetjes, 2016) (https://archive.mpi.nl/tla/elan). The tier 'gesture onset-offset' is used to annotate the start and stop of a gesture.

3.6.2.2 Quantifying biking movements

Identification of the movement of the pedals

Each pedal was equipped with a motion capture marker. The time series of the positional coordinates of the markers were extracted in the same csv format file as the rigid bodies. These markers were not within rigid bodies. As mentioned in the previous section, these markers could be masked from time to time, in which case a new marker ID was added when they became visible again by the cameras. Therefore, these markers' ID, which correspond to the same physical marker at different time points, needed to be merged together into one marker ID. A python script was created for this purpose by our research assistant.

To detect the circular trajectories and merge them when they were from the same marker, the 3D trajectories of each marker were parsed and kept if they were fitting a circle. To check this last point, the coordinates of the center of a 3D trajectory was calculated by averaging the coordinates on each dimension:

$$x_c = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.2}$$

$$y_c = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.3}$$

$$z_c = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{3.4}$$

where n is the number of elements in the vector of coordinates. A vector was created encompassing the euclidean distance of each point of the trajectory vector to its center as follows:

For i from 1 to n,

$$D_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2} \tag{3.5}$$

Residuals are defined as the sum of the distances between the average of D and each point of D:

$$R = \frac{1}{n}\sum_{i=1}^{n}(D_i - D_m)^2 \tag{3.6}$$

Where D_m is the mean of D:

$$D_m = \frac{1}{n}\sum_{i=1}^{n} D_i \tag{3.7}$$

Two criteria are required to judge the circularity of the trajectory:

(1) The average distance should be close to the ray of the bike (17 cm). The maximum error accepted was 0.005 and was empirically assessed over several files; (2) Theoretically, if the trajectory is a circle, the distance from the center to each point of the trajectory should be constant. The residuals should equal 0. The upper bound of the residual value was 0.5. This value was also empirically assessed over several files.

Once the circular trajectories were detected, the most circular trajectory (minimizing the value of the residuals, see equation 3.6) was selected and completed by other trajectories: if a trajectory was circular and there was no overlap between the time intervals of their coordinates, this trajectory was merged to the most circular one. Intervals without any value were interpolated. This way, a complete vector of the pedal marker was reconstructed. Since the bike is a rigid body and does not have any degrees of freedom between the two pedals (the movement of one pedal determines the movement of the other), the reconstruction of one marker was sufficient.

The coordinates of the vertical position of this vector were saved in a csv file at 200 Hz. The other dimensions were not used because the bike underwent some small shifts through the trials, especially on the horizontal plane. Since some movement files were mistakenly recorded at 120 Hz instead of 200 Hz, the extracted coordinate vector was resampled at 100 Hz to have all the vectors at the same sample rate, and filtered with a Butterworth filter at a cutoff frequency of 10Hz. The signal peaks (one peak for each cycle) were detected using the peak finder function of Matlab (*findpeaks*, *MATLAB version 9.10.0.1613233 (R2021a)*, 2021a) with adequate parameters validated for each participant based on signal visualization. Results of the detection were visually inspected. Each session started by two seconds with no movement during which the breathing recording was launched, and each participant needed one to two biking cycles to reach their comfortable rhythm and start speaking. This part was removed by discarding the coordinates before the first complete biking cycle. The signal was transformed with the arc sinus function to express displacement as an angle of rotation. The acceleration was computed as the second derivative of the position signal.

Characterization of the biking cycles (*BikDur* and *SdBikDur*)

The biking cycles were characterized by their duration and by the standard deviation of their duration. The peak finder function of Matlab (*findpeaks*, *MATLAB version 9.10.0.1613233 (R2021a)*, 2021a) was applied on the vertical position of the pedals. The onset of a cycle at a time *t* was the peak at t, and its offset was the peak at *t+1*. The duration of a cycle (*BikDur*) was computed by substracting *t* to *t+1*. The biking cycle duration was averaged over each trial. The standard deviation of the biking cycle durations (*SdBikDur*) was computed over the biking cycle durations of each trial.

3.6.3 Acoustic characterization of speech and characterisation of narration

Three dimensions of speech were analysed: speech fluency, speech acoustics and content recall. Speech fluency was characterized first at the scale of the trial, then at the scale of the breathing cycle, and was described by the number of syllables, the speech rate and the proportion of time spent speaking (over the trial, or over the breathing cycle). Content recall was characterized by the rate of information units correctly recalled. Speech acoustics were described by the amplitude envelope and the fundamental frequency (F0) of the speech signals.

3.6.3.1 Extraction of parameters related to fluency (*NumSyll*, *SpRate* and *SpeechTime*) over the trial and over the breathing cycle

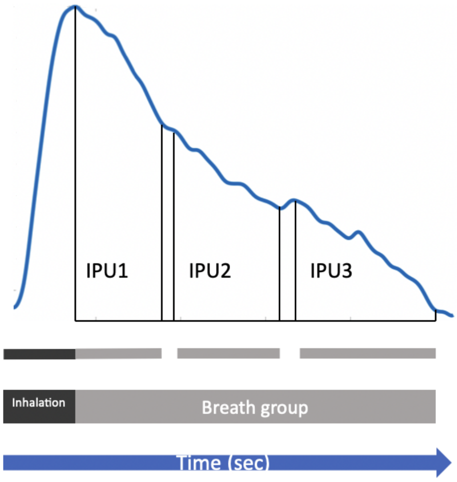


Figure 3.11: Representation of the notions of interpausal units and breath group within a speech breathing cycle. The duration of the breath group is the sum of the durations of the interpausal units.

The parameters related to fluency were extracted at the scale of the breathing cycle. As seen on Figure 3.11, an interpausal unit (IPU) is an interval containing speech between two silent pauses. A breath group is the interval containing all the IPUs within a breathing cycle. The onset of the breath group is the onset of the first IPU, and the offest of the breath group is the offset of the last IPU. The audio recording of the experiment were transcribed by a translator hired specifically for this task. The translator transcribed speech content within each breathing cycle. Through the software PRAAT (Boersma, 2009), a German speaking research assistant annotated the interpausal units in each breath group. Number of syllables (*NumSyll*) per IPUs were detected automatically through a script in python (Van Rossum and Drake, 2009) taking these annotations as input. This detection was corrected in PRAAT by a research assistant speaking German. *NumSyll* was summed per breathing cycle ($NumSyll_b$) or per trial ($NumSyll_t$) depending on the ongoing analysis. When *NumSyll* was analysed per breathing cycle, *NumSyll* was averaged over all the breathing cycles of the trials.

The speech rate over breathing cycles ($SpRate_b$) was calculated by dividing the number of syllables within a breathing cycle over the duration of the exhalation of the corresponding breathing cycle. $SpRate_b$ was averaged over all the breathing cycles of the corresponding trial. The speech rate over the trial ($SpRate_t$) was calculated by dividing the number of syllables within a trial by the duration of the trial.

At the scale of the trial, the proportion of time spent speaking ($SpeechTime_t$) was calculated as follows: the cumulated duration of IPUs within a trial was divided by the cumulated duration of the exhalations within the same trial. At the scale of the breathing cycle, the proportion of time spent speaking ($SpeechTime_b$) was the duration of the breath group over the duration of the corresponding exhalation. $SpeechTime_b$ was averaged over all the breathing cycles of the trial.

3.6.3.2 Computation of F0 (*F0*) and amplitude envelope (*Env*)

The fundamental frequency (*F0*) was extracted from each speech signal using PRAAT (Boersma, 2009; autocorrelation method, range = 50–250 Hz for males and 100–300 Hz for females; cf. Figure 4.8, row 3). The values above the third quartile plus three times the inter-quartile space and below the first quartile minus one time the inter-quartile space were discarded. These thresholds were chosen after looking deeply in the data and noticing that most of the values between the third quartile plus one time and three times the inter-quartile range were not outliers. On average 3% of all data were discarded. After downsampling the speech signal to 11025 Hz, the amplitude envelope (*Env*) was extracted and processed using the method of Pouw et al. (2020a): the *envelope* function of Matlab (2021b), using a Hilbert transformation, was applied to the speech signal. The upper bound of the output was filtered using a 5-Hz Hanning window, and the signal was smoothed using a Gaussian-weighted moving average over each window of 1000 points (see 4.8, row 2). The code is available at https://osf.io/npkdh/. To temporally align the values of the acoustical parameters F0 and intensity with the movement signals, all were sampled to 100 values per second (100 Hz).

3.6.3.3 Scoring the content of the recalls (*ContRec*)

To assess the content of the free recalls (*ContRec*), a grid was created to isolate each information unit contained in the audio of the story. Two points were attributed to each information unit. An annotator was hired to listen to the free recalls of the participants and identify the information units of the audio that were recalled. An information unit could be an adjective or a complete sentence (Hyman Jr, 1994). Two points were given if the information unit was recalled correctly. One point was given if it was imprecise or incomplete. For instance, if the participant said 'the director asked her if she wanted to play in a movie' whereas the information was 'the director offered her a role in a series', one point was given. No point was given if the information unit was not recalled at all. One point was removed if the information was falsely recalled. For instance, if the participant said 'she had a small head' instead of 'she had a big head'. The score corresponds to the ratio of the sum of the points of the participants over the sum of the points of the information units (so over two times the number of information units). The stories respectively accounted for 71, 73, 73 and 76 points. To evaluate the reliability of this grid, a second annotator worked on a subset of the recalls (88 out of 365), and an inter-rater reliability score was established by taking the Pearson's correlation between the rating scores. As seen in Figure 3.12, the scores of the two annotators were highly correlated (cor=0.99,t $= 61.985$, df $= 86$, p-value < 0.001).

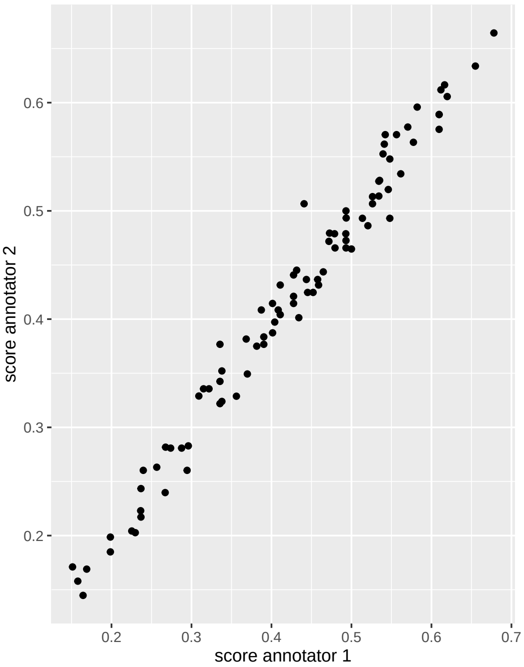


Figure 3.12: Scores of annotator 1 versus scores of annotator 2 on the common subset of recalls.

In this section, the variables to be analyzed were described. In next section, the statistical methodology of the models applied on these variables will be presented.

3.7 Statistical analyses

The statistical tool mainly used is the family of linear mixed models. The analyses were run in the R environment (Team, 2022) using the following packages:

(1) To run the models, the packages *lme4* (Bates et al., 2015), *nlme* (Pinheiro et al., 2022), *mgcv* (Wood, 2017), *glmmTMB* (Brooks et al., 2017) and *brms* (Bürkner, 2021) were used. *lme4* allows considering complex structure of random effects when needed, while

nlme enables the analysis and correction of the variance-covariance matrix. *glmmTMB* is for generalized linear models such as logistic regressions. *brms* is to apply generalized additive mixed models. (2) To check whether there is some autocorrelation left in the model, the *DHARMa* package (Hartig and Hartig, 2017) was used. (3) To estimate post-hoc comparisons, *multcomp* (Hothorn et al., 2008) and *emmeans* (Mandler and Johnson, 1977) packages were employed.

The procedure of the statistical analyses using linear mixed models always complied with the following steps:

- an initial model containing all the random and fixed factors as well as their interactions was created;
- the evolution of the inter-individual variability over the different modalities of each fixed factor was checked and corrected if needed by adding random slopes to the participant random intercept;
- Non-homogeneous variance of the residuals across conditions was corrected for when present;
- A term to consider the potential auto-correlation between the consecutive points within the time series was included if needed;
- The best model was selected by a backward selection process: at each step, one interaction or one factor (interactions first) was removed and the models were compared with either a likelihood ratio test through the anova function, or using the AIC criteria. The AIC seeks for the best trade-off between the variance in the data explained by the model and parsimony of number of factors within the model;
- the model was diagnosed by looking at the distribution of the residuals and checking the absence of structure in the residuals (no more interindividual differences, no more autocorrelation etc.).

The dependent variables of the models are presented on jitter plots designed with the package *ggplot2* (Wickham, 2016), allowing to see the distributions of the data depending on the fixed factors.

The principal factors of the models applied are the day (*day 1*; *day2*), the limb condition (*speech_ armFree*; *speech_ armBlock*;*speech_legMot*; *speech_ armMot*), the vocal condition (*quiet*; *speech*) and the type of recall (*shortTerm*; *longTerm*). The type of recall distinguishes the short-term narration task (story recalled just after being watched) on day 1 and 2 and long-term recalls performed at the beginning of days 2 and 3 without constraints on hand motion, and grouping all the stories.

For each method detailed in the next chapters, the data set involved 25 participants. The database should contain 100 rows (25 participants x 2 types of recall x 2 days). However, some files are missing for three reasons: (1) the participants did not come back (participants 17 and 25); (2) because of failures in saving certain files coming from the motion capture; (3) because of failures in recording the sessions with the camera (memory full).

In chapter 4, for describing the interactions between speech and limb movements, as well as for measuring the effect of speech and limb movements on breathing cycles at the beginning of chapter 5, the statistical method consists in linear mixed models, adapted to the dependent variable when necessary. For each model, Post-hoc comparisons were made using Tukey's HSD correction. The critical significance threshold was fixed at 0.05.

The models will be specified along with the corresponding result section. Chapter 4 reports the analysis of the interactions between speech and limb movements. Chapter 5 reports the analysis of breathing. These analyses will be done on the variables presented in this section.

CHAPTER 4

Interactions over time between speech and limb movements

The aim of this chapter is to describe and compare the interactions between speech and limb movements across the four different limb conditions and over time. The first part details the effect of speaking conditions on limb movements and focuses on: (1) the *speech_ armFree* condition, and the evolution of the time spent gesturing across days and the different retrievals; (2) the effect of speech on biking parameters compared to biking while being quiet, and these effects over the two first days. The second part compares the effect of the different limb movement conditions on speech fluency, speech content and speech acoustics, over the two first days.

Our contribution is: (1) to assess the spontaneity of gesturing during story retrieval over several days; (2) to compare the interaction between speech and free gestures to the interaction between speech and a constrained non communicative movement; (3) to compare the interaction between speech and upper limb movements to the interaction between speech and lower limb movements; (4) to assess the points (1) to (3) with a quantitative approach using motion capture data that has been poorly used to analyse body motion during narration and in the analysis of the effect of task repetition on three different days.

4.1 How much do speakers move their hands during the narrative task? Does this quantity of motion change with task repetition?

The aim of this section is to assess the evolution of time spent gesturing spontaneously across days and task repetition. This section describes spontaneous hand gestures in the *speech_ armFree* condition over the three days. We also detail and compare the quantitative and qualitative approach to analyze gestures.

The focus is first on arm-hand gestures that speakers may produce during the narrative task when able to move their hands (*speech_ armFree* condition). More specifically, the question of the impact of task repetition on gesture quantity between days 1 and 2 for the short-term and story-specific recalls, and between days 2 and 3 for the long-term and grouped recalls, is addressed.

To answer these questions, the variables analyzed are the quantity of hand motion (*HandMvt*) and the proportion of time spent gesturing (*RateGest*). These variables are defined in Chapter 3 Section 3.6.2.1 and are analyzed through the day as the only factor of interest, on the short-term recalls on one hand and the long term recalls on the other hand.

4.1.1 Statistical analyses

The effect of the day on hand gesturing was assessed through linear mixed models applied on *HandMvt* and *RateGest*. The day was fitted as a fixed effect, and a random intercept by participant was fitted in the model. A first model was applied on the short term recall of the stories (days 1 and 2), in which the position of the recall within the four recalls was added as a fixed effect, and the story recalled was added as a random factor. It resulted in a two-factor model: *day (1,2)* and *position (1,2,3,4)*.

The models on the short term recalls are:

- *HandMvt* ∼ *position*day + random(1|participant) + random(1|story)*
- *RateGest* ∼ *position*day + random(1|participant) + random(1|story)*

A second model was applied on the grouped long term recalls (days 2 and 3), resulting in a one factor model: *day (2,3)*. Since the variable *RateGest* is a ratio defined in the interval [0;1[, we applied a generalized linear mixed model with a logit link function.

The models on the long term recalls are:

- *HandMvt* ∼ *day + random(1|participant)*
- *RateGest* ∼ *day + random(1|participant)*

4.1.2 Results on the hand motion of the participants across days (task repetition)

Variables averaged by day and type of recall are presented in Table 4.1.

Table 4.1: Average of the variables concerning gestures, per day and type of recall. *cum. gest. dur* is the cumulative gesturing duration per trial. *Trial duration* and *cum. gest. dur.* are in seconds. Average 3D distance is in cm per second.

day	type of recall	trial duration	cum. gest. dur.	RateGest	HandMvt
1	shortTerm	136 ± 46.39	28.5895 ± 31.78	0.22 ± 0.28	9.08 ± 10.6
2	shortTerm	139 ± 31	24.1778 ± 36.52	0.18 ± 0.27	6.48 ± 9.5
2	longTerm	384 ± 158.21	92.4942 ± 95.97	0.25 ± 0.27	6.48 ± 4.23
3	longTerm	453 ± 178.53	78.2018 ± 103.87	0.16 ± 0.20	5.53 ± 4.41

Table 4.2: Post-hoc comparisons from the linear mixed models : *gestureQuantity* ∼ *condition*day* (short-term recall) and *gestureQuantity* ∼ *day* (long-term recall). *RateGest* refers to the proportion of time spent gesturing, and *HandMvt* refers to the average 3D distance.

variable	type of recall	contrast	Estimate	Std.Error	df	z-value	Pr(>|z|)
RateGest	short term	day1 - day2	0.0986	0.267	38	0.370	0.7136
	long term	day2 - day3	0.662	0.172	41	3.851	0.0004
HandMvt	short term	day1 - day2	2.46	0.653	22	3.762	0.0011
	long term	day2 - day3	0.878	0.644	21	1.365	0.1869

Table 4.2 displays the principal results of the factors having an effect on *HandMvt* and *RateGest*. Neither the position (in which occurred the recall compared to the other conditions) nor the story had any effect on these variables. As seen on Figure 4.1, in average, *HandMvt* decreases on day 2 compared to day 1 (b = 2.46, z = 3.76, p = 0.001); *RateGest* decreases on day 3 compared to day 2 (b = 0.67, z = 3.85, p = 0.0004); there is a tendency of *HandMvt* to decrease from day 2 to day 3 (b = 0.8, z = 1.36, p = 0.19).

4.1.3 Analysis of the consistency between motion capture and annotation analyses

The results from the quantitative approach on one hand and the qualitative approach on the other hand are not consistent. While *HandMvt* decreases from day1 to day2, *RateGest* decreases from day2 to day3. However, although their effect are not on the same type of recall, they are both showing a decrease of hand movement over time. To go further on the consistency of these two analyses, we calculated the Pearson correlation between the average distance of the hands and the ratio of time spent gesturing (see Figure 4.2). The correlation coefficient is 0.53 and significant (pearson=0.53, t = 5.8155, df = 87, p-value = 9.811e-08).

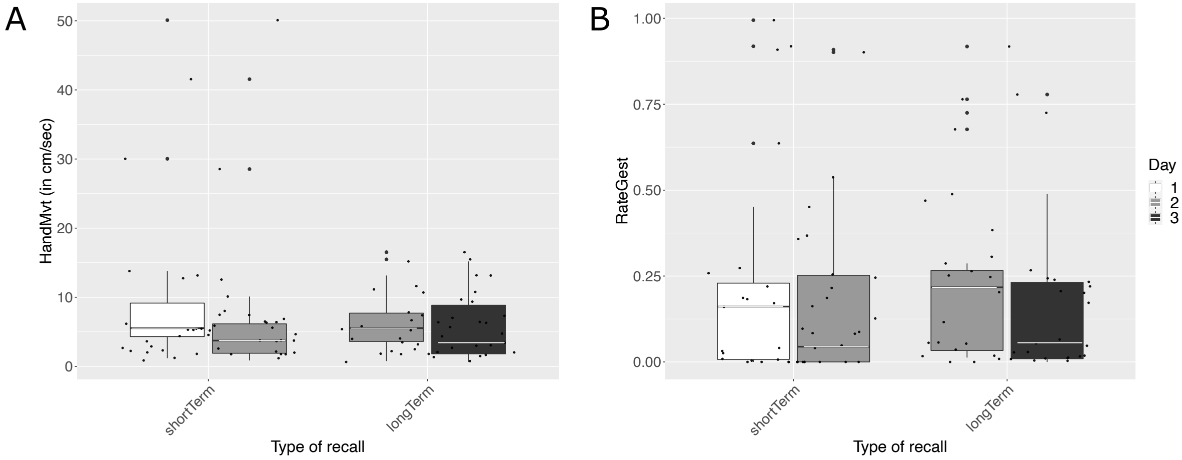


Figure 4.1: Gesture quantity depending on day and type of recall. A: average 3D distance of both hands (*HandMvt*); B: proportion of time spent gesturing (*RateGest*).

On short-term recalls, *HandMvt* decreases with the repetition of the task across days, but not *RateGest*, suggesting that participants do fewer movements not considered as gestures from one day 1 to day 2. On long-term recalls, *RateGest* decreases with the repetition of the task across days, but not *HandMvt*, suggesting that participants do more movements not considered as gestures from day 2 to day 3. The correlation between the two variables is not high and may reveal the discrepancy between *HandMvt* and *RateGest*, lying in their definition: *HandMvt* accounts for the movement quantity of the hands, while *RateGest* accounts for the proportion of co-speech gestures. The next section focuses on the effect of speech on biking movements.

4.2 Does the narrative task affect biking motions with the arms vs. the legs ?

Previous research shows that in dual-task paradigms, speech impacts the variability of the position, frequency or speed of concurrent body motion such as gait balance (Verghese et al., 2007), hand drawing (Dayalu et al., 2013) or finger tapping, but this question has never been assessed on biking movements. The latter have the particularity of being controlled in position and phase: the trajectory and the anti-phase dynamic between both hands/feet are imposed by the structure of the bike. Yet, can biking movements be disturbed in their speed? Are hand biking movements more disturbed because of the natural link between hand and mouth motor control? The aim of this section is to assess the impact of speech on biking parameters such as cycle duration and its variability, and to compare upper and lower limbs.

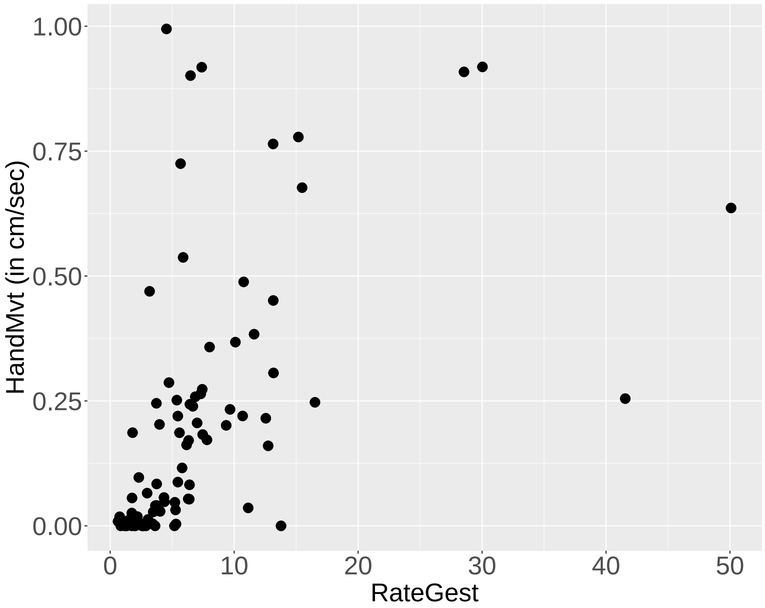


Figure 4.2: Correlation between *HandMvt* and *RateGest*.

This section describes biking cycles of the *legMot* and *armMot* conditions across quiet and speech phases and over the three days. The contribution of this section is to assess the effect of a cognitivo-motor task such as speech on biking movements, and to compare this effect between upper and lower limbs. More specifically, we address the following questions:

1. Does speech impact cycle duration and cycle duration variability compared to being silent?
2. Is this impact higher for arm than for leg biking movements?

To answer these questions, the variables analyzed are the biking cycle duration (*BikDur*) and the standard deviation of the biking cycle duration (*SdBikDur*). These variables are analyzed through three factors: the limb condition (*legMot*; *armMot*), the vocal condition (*speech*; *quiet*) and the day (*day1*; *day2*) in the short-term narration task.

4.2.1 Statistical analyses

To test whether speech impacts cycle duration and its variability compared to being silent, a first model was applied to the data from day 1. Its aim was to examine the effect of speech and limbs on biking cycle duration. It included two fixed effects: vocalCondition (*speech*;*quiet*) and limbCondition (*legMot*;*armMot*). The participant was included as a random effect, with a random slope on vocal and limb conditions. For *BikDur* as the dependent variable, the residual variability was corrected on the vocal condition factor. For *SdBikDur* as the dependent variable, the residual variability was corrected for the interaction between the two factors.

The models are:

- *BikDur* $\sim$ *vocalCondition*limbCondition* + *random(1+ vocalCondition* + *limbCondition/participant)*
- *SdBikDur* $\sim$ *vocalCondition*limbCondition* + *random(1+ vocalCondition* + *limbCondition/participant)*

To test whether the effect of speech on *BikDur* and *SdBikDur* changes over days, a second model was applied to the data from speech conditions only. Its aim was to examine the effect of limb motion (*speech_ legMot*;*speech_ armMot*) and day (*day1*;*day2*) on biking cycle duration. Participant was included as a random effect, with a random slope on limb condition and day. The residual variability was corrected on the day.

The models are :

- *BikDur* $\sim$ *limbCondition*day* + *random(1+ day* + *limbCondition / participant)*
- *SdBikDur* $\sim$ *limbCondition*day* + *random(1+ day* + *limbCondition / participant)*

4.2.2 Results: do speech and days impact biking cycle durations and their variability?

The distribution of the number of cycles per condition and per day is described in Table 4.3. The mean values and standard deviation of the cycle durations per trial are summarized in Table 4.4.

The results of linear mixed models are available in Table 4.5 for the effect of speech and limb condition on day 1, and in Table 4.6 for the effect of limb conditions and day. The Figures 4.3 and 4.4 illustrate the results.

The first analysis concerns only the data from day 1. Its purpose was to assess the effect of limb motion and vocal condition on the duration of biking cycles (*BikDur*). The selected statistical models included the interaction between *vocalCondition* and *limbCondition*. In speaking conditions, biking with the arms increased the average duration of the cycles and its standard deviation compared to biking with the legs (*BikDur*: b = 0.28, z = 6.58, p <0.0001; *SdBikDur*: b = 0.04, z = 2.69, p = 0.04). In limb conditions *legMot* and *armMot*,

Table 4.3: Summary of the number of biking cycles per day and condition. The mean is presented with the standard deviation.

day	conditions	min	mean ± sd	max
1	quiet_legMot	70	117.59 ± 26.06	179
1	quiet_armMot	79	108.45 ± 22.09	161
1	speech_legMot	29	108.91 ± 49.90	222
1	speech_armMot	24	88.14 ± 37.00	158
2	speech_legMot	36	122.29 ± 54.24	248
2	speech_armMot	31	93.23 ± 39.95	171

speaking decreased the average duration of the cycles compared to quiet (*quiet_ armMot* - *speech_ armMot*: b = -0.32, z = -7.66, $p < 0.0001$; *quiet_ legMot* - *speech_ legMot* : b = -0.12, z = -2.77, p = 0.036). When biking with the arms, speaking increased the standard deviation of cycle durations (b = - 0.05, z = -5.54, $p < 0.0001$).

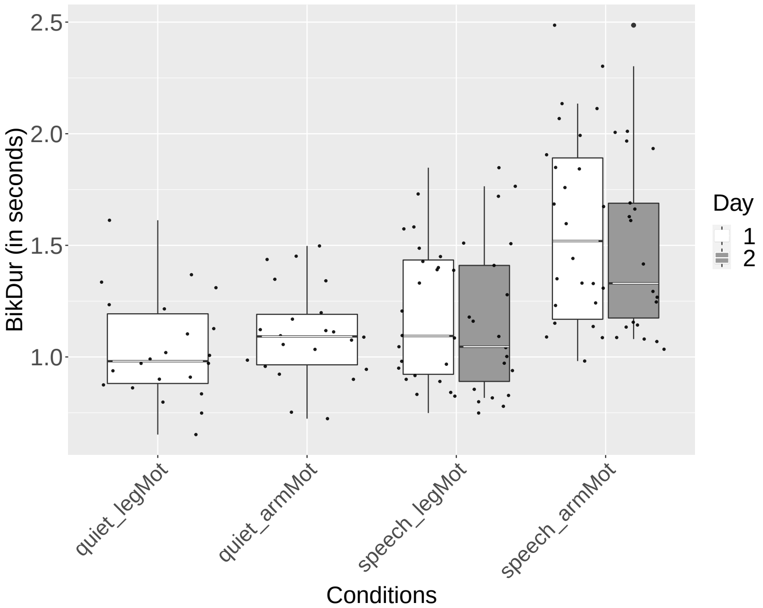


Figure 4.3: Biking cycle duration per condition and day.

The second analysis concerns only the data from speech conditions (on days 1 and 2). Its purpose was to assess the effect of day and limb motion on the duration of biking cycles (*BikDur*) and its standard deviation (*SdBikDur*). The selected statistical model included the main effect of limb condition on the cycle duration, and main effects of limb condition and day on the standard deviation of cycle durations. Arm biking cycles were longer and

Table 4.4: Summary of the distribution of the durations (in seconds) of the biking cycles (*BikDur*), and the standard deviation of the durations (*SdBikDur*), per condition and day.

BikDur

day	condition	min	mean ± sd	max
1	quiet_legMot	0.65	1.04 ± 0.23	1.61
1	quiet_armMot	0.72	1.11 ± 0.21	1.5
1	speech_armMot	0.98	1.55 ± 0.39	2.13
1	speech_legMot	0.75	1.18 ± 0.32	1.85
2	speech_armMot	1.08	1.52 ± 0.42	2.49
2	speech_legMot	0.82	1.17 ± 0.31	1.76

SdBikDur

day	condition	min	mean ± sd	max
1	quiet_legMot	0.02	0.05 ± 0.024	0.12
1	quiet_armMot	0.03	0.06 ± 0.017	0.08
1	speech_armMot	0.04	0.12 ± 0.066	0.3
1	speech_legMot	0.03	0.08 ± 0.036	0.15
2	speech_armMot	0.05	0.11 ± 0.046	0.2
2	speech_legMot	0.02	0.06 ± 0.027	0.11

Table 4.5: Post-hoc comparisons from the models estimating the effect of limb motion and vocal condition on biking cycle duration (*BikDur*) and its standard deviation (*SdBikDur*) on day 1.

Variable	contrast	estimate	SE	df	t.ratio	p.value
BikDur	quiet_arms - quiet_legs	0.0712	0.0421	63	1.693	0.3357
	quiet_arms - speech_arms	-0.3223	0.0421	63	-7.661	<.0001
	quiet_legs - speech_legs	-0.1167	0.0421	63	-2.774	0.0357
	speech_arms - speech_legs	0.2768	0.0421	63	6.580	<.0001
SdBikDur	quiet_arms - quiet_legs	0.00318	0.00455	63	0.700	0.8967
	quiet_arms - speech_arms	-0.05941	0.01072	63	-5.544	<.0001
	quiet_legs - speech_legs	-0.02364	0.01072	63	-2.206	0.1328
	speech_arms - speech_legs	0.03896	0.01446	63	2.695	0.0436

more variable than leg biking cycles (*BikDur*: b = 0.26, z = 7.1, p = 0.0001; *SdBikDur*: b = b = 0.043, z = 5.27, $p < 0.0001$). Biking cycles were more variable on day 1 than on day 2 (b = 0.01, z = 2.92, p = 0.005).

Table 4.6: Post-hoc comparisons from models estimating the effect of limb conditions and days on biking cycle duration (*BikDur*) and its standard deviation (*SdBikDur*).

variable	contrast	estimate	SE	df	t.ratio	p.value
BikDur	arms - legs	0.268	0.0377	64	7.103	<.0001
SdBikDur	arms - legs	0.0429	0.00814	63	5.270	<.0001
	day1 - day2	0.0142	0.00486	63	2.918	0.0049

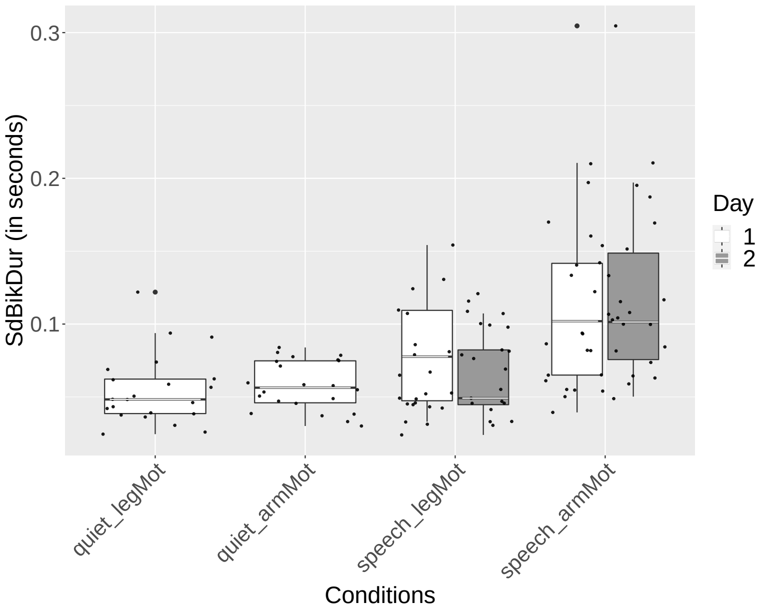


Figure 4.4: Standard deviation of the duration of biking cycles per condition.

In this section, we saw that: (1) the quantity of hand motion decreased over days and task repetition. (2) arm biking cycles were longer than leg biking cycles in speech conditions. Speech decreases the duration of biking cycles for both arm and leg movements, and increases the variability of the arm biking cycle duration compared to being silent. Do speech parameters related to fluency, acoustics or content change depending on the co-occurrent limb movements?

4.3 How do limb movements influence speech across several days?

As seen in Chapter 2 section 2.1, the effect of limb movements on speech fluency and content recall are still under debate: while in some studies, gesturing with the hands enhances vocabulary richness and decreases disfluencies (Hostetter et al., 2007a), in other studies, gesturing does not have any effect (Hoetjes et al., 2014). In dual-task paradigms, speech does not seem to be perturbed by other movement tasks, and content recall seems to be enhanced by a co-occurrent locomotion task. Speech acoustics are impacted by physical impulses of upper limb movements (Pouw et al., 2020a). The aim of this section is to assess the effect of different limb movements on speech fluency, speech content and speech acoustics.

4.3.1 Do movement conditions and days have an effect on speech parameters related to fluency?

This first section describes the effect of the day and the limb conditions on speech fluency. The variables chosen to characterize speech fluency are the number of syllables ($NumSyll_t$), speech rate ($SpRate_t$), and the proportion of time spent speaking ($SpeechTime_t$). These variables are defined in Chapter 3 Section 3.6.3.1 and analyzed through two factors: limb condition (*speech_ armFree*;*speech_ armBlock*;*speech_ legMot*;*speech_ armMot*) and day (*day1*;*day2*).

4.3.1.1 Statistical analyses

For each variable characterizing speech fluency and content recall, the analyses were divided in two parts: one model was applied on the short-term recalls of the first two days, and includes limb conditions. One model was applied on long-term recalls of days 2 and 3.

The models on short-term recalls included limb condition and day as fixed effect factors, and participant as a random effect factor. The story was also considered as a random effect factor, and the position in which the condition took place was among the fixed effect factors.

The models are:

- $NumSyll_t \sim limbCondition * day + random(1/participant) + random(1/story)$
- $SpRate_t \sim limbCondition * day + random(1/participant) + random(1/story)$
- $SpeechTime_t \sim limbCondition * day + random(1/participant) + random(1/story)$

The models applied on long-term data included the day as a fixed factor, and participant as a random factor. As for content recall, the model applied on long-term data also included the condition and the position as fixed effects.

The models are:

- $NumSyll_t \sim day + random(1/participant)$
- $SpRate_t \sim day + random(1/participant)$
- $SpeechTime_t \sim day + random(1/participant)$

The linear mixed model were applied to the number of syllables per trial ($NumSyll_t$) and speech rate ($SpRate_t$). Since the proportion of speech time ($SpeechTime_t$) and the content recall score (*contRec*) are ratios, a generalized linear mixed model with a logit link function was applied to them. $NumSyll_t$ was log-scaled to avoid contrasts of dissimilar magnitudes.

4.3.1.2 Results: how do limb movements and day impact parameters related to speech fluency over the trials?

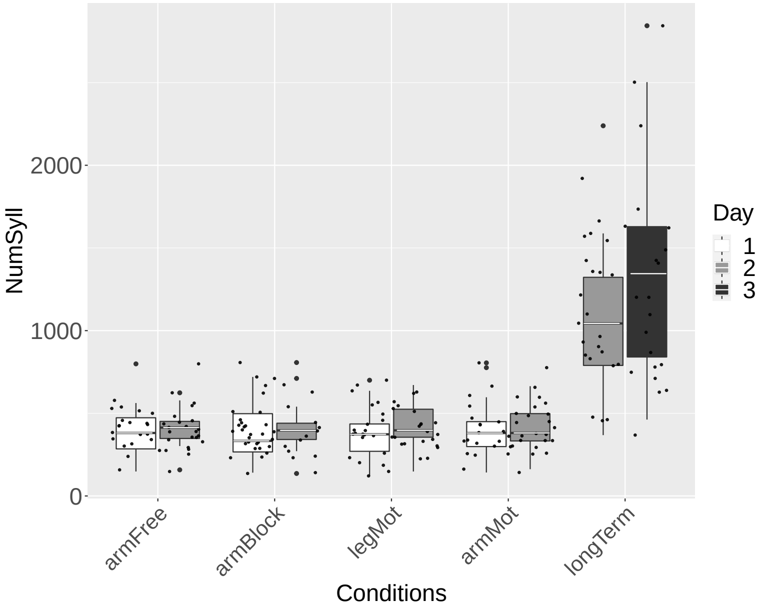


Figure 4.5: Distribution of number of syllables ($NumSyll_t$) over a trial, per limb condition and day.

Table 4.7: Post-hoc comparisons from the models estimating the effect of limb conditions and day on speech parameters.

variable	type_recall	contrast	estimate	SE	df	t.ratio	p.value
SpeechTime	short-term	day1 - day2	-0.0657	0.0356	187	-1.846	0.0665
	long-term	day2 - day3	-0.169	0.0572	42	-2.958	0.0051
NumSyll	longTerm	day2 - day3	-256	93.8	21	-2.726	0.0127
SpRate	short-term	day1 - day2	-0.163	0.0426	21	-3.831	0.0010

Tables 4.8, 4.9 and 4.10 summarize distributions of each variable per limb condition and day. The results of the remaining factors of the linear mixed models are available in Table 4.7. As displayed in Figure 4.5, concerning the short-term recalls, neither the day nor the condition have an effect on the number of syllables per trial. $NumSyll_t$ increases significantly on day 3 compared to day 2 (*day2 - day3*: b = -256, z = -2.726, p = 0.01). As displayed in Figure 4.6, for the short-term recalls, the limb condition does not have any effect on $SpRate_t$, but $SpRate_t$ increases on day 2 (b = -0.16, z = -3.83, p = 0.001). For the long-term recalls, the day does not have any effect on the $SpRate_t$. As displayed in Figure 4.7, on the short-term recalls, $SpeechTime_t$ increases from day 1 to day 2 (b = -0.06, z = -1.85, p = 0.066). Concerning the long-term recalls, $SpRate_t$ increases from day 2 to day 3 (b = -0.17, z = -2.96, p = 0.005).

Table 4.8: Summary of the distribution of number of syllables ($NumSyll_t$) over a trial per condition and day.

day	condition	min	mean ± sd	max
1	speech_armMot	142	411.3 ± 165.2	805
1	speech_armFree	148	395.6 ± 141.5	799
1	speech_armBlock	141	394.7 ± 170.4	720
1	speech_legMot	121	371.1 ± 146.8	700
2	speech_longTerm	368	1062.4 ± 429.2	2239
2	speech_armMot	162	418.6 ± 136.8	664
2	speech_armFree	158	411 ± 100.5	624
2	speech_armBlock	136	412.5 ± 141.1	807
2	speech_legMot	148	430.3 ± 132.1	671
3	speech_longTerm	462	1318.2 ± 610.5	2844

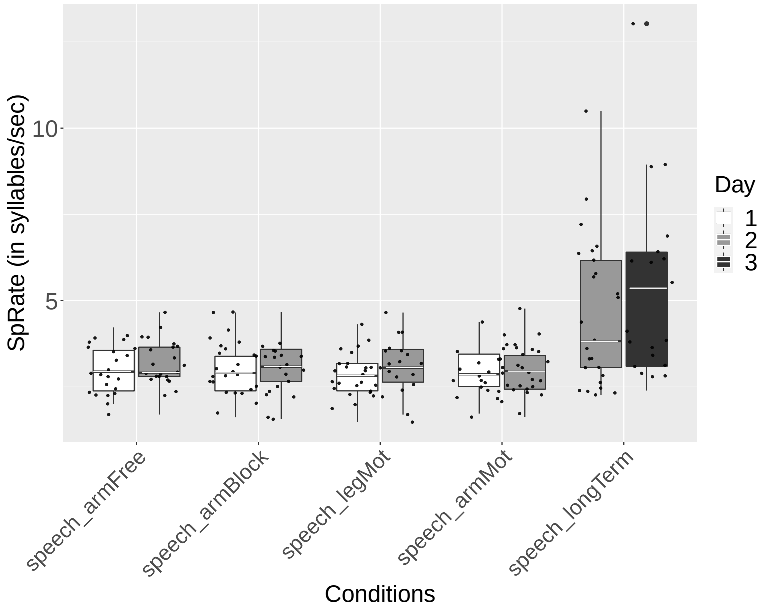


Figure 4.6: Distribution of speech rate ($SpRate_t$) over a trial, per limb condition and day.

Table 4.9: Summary of the distribution of the speech rate ($SpRate_t$) over a trial (in syllable/second), per condition and day.

day	condition	min	mean ± sd	max
1	speech_armMot	1.73	2.94 ± 0.67	4.38
1	speech_armFree	2.01	3.03 ± 0.67	4.23
1	speech_armBlock	1.62	2.91 ± 0.72	4.66
1	speech_legMot	1.48	2.83 ± 0.68	4.32
2	speech_longTerm	2.27	4.53 ± 2.14	10.49
2	speech_armMot	1.63	2.99 ± 0.70	4.77
2	speech_armFree	1.7	3.15 ± 0.66	4.66
2	speech_armBlock	1.56	3.13 ± 0.73	4.67
2	speech_legMot	1.7	3.1 ± 0.69	4.66
3	speech_longTerm	2.39	5.31 ± 2.63	13.02

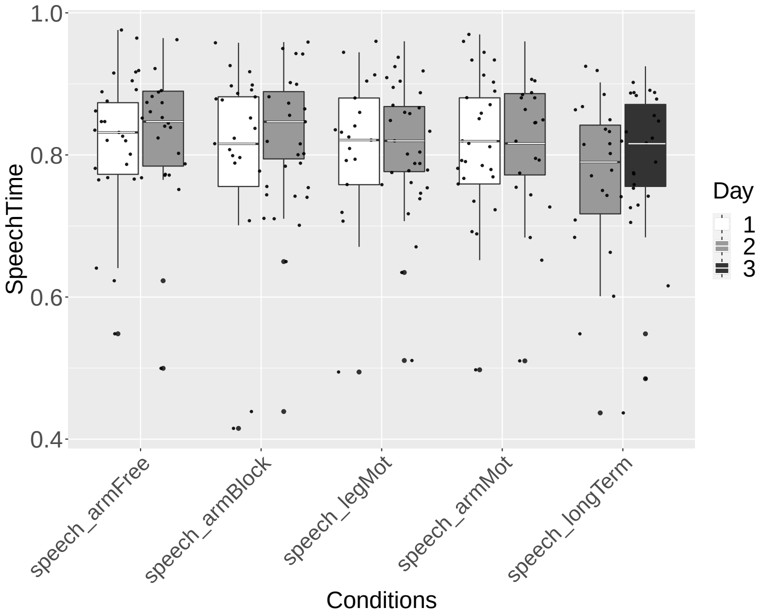


Figure 4.7: Distribution of time spent speaking ($SpeechTime_t$) over a trial, per limb condition and day.

Table 4.10: Summary of the distribution of time spent speaking ($SpeechTime_t$) over a trial, per condition and day. Since $SpeechTime_t$ is a ratio, it does not have unit.

day	condition	min	mean ± sd	max
1	armMot	0.50	0.81 ± 0.10	0.97
1	armFree	0.55	0.82 ± 0.09	0.98
1	armBlock	0.42	0.81 ± 0.11	0.96
1	legMot	0.49	0.81 ± 0.10	0.94
2	longTerm	0.44	0.77 ± 0.11	0.90
2	armMot	0.51	0.81 ± 0.10	0.96
2	armFree	0.50	0.83 ± 0.11	0.96
2	armBlock	0.44	0.82 ± 0.12	0.96
2	legMot	0.51	0.81 ± 0.10	0.96
3	longTerm	0.49	0.79 ± 0.11	0.92

After describing speech fluency with quantitative terms, the next section presents analyses on the effect of biking movements on speech acoustics.

4.3.2 Do movement conditions and days have an effect on speech amplitude envelope and fundamental frequency?

The work described in this part has been submitted to *Journal of Neurophysiology* and received with minor revisions. In this part, the question addressed is whether speech acoustics (f0 and amplitude envelope) are impacted by the different limb movements. More specifically, in order to further address the biomechanical link between limb motion and speech, the present part investigates the effect of limb movements on acoustic parameters of spontaneous speech during biking. Intensity and fundamental frequency of narrative speech are analyzed during biking motions with the arms or with the legs in comparison to a control condition with no motion. Arm or leg biking motion may induce the contraction of muscles involved in breathing or antagonist muscles impacting the thoracic cage. We aimed to test whether these movements, with low energy demands and no synchrony constraints, generate an F0 and intensity peak in speech at the highest physical impulse of the motion (the acceleration peak within each movement cycle). The data were analyzed using an approach similar to Pouw et al. (2020a), investigating the non-linearity in the speech amplitude envelope and fundamental frequency (F0) around the acceleration peak in limb movements.

4.3.2.1 Statistical analyses

Speech amplitude envelope and F0 are defined in Chapter 3 Section 3.6.3.2 and analyzed at two different temporal scales: the scale of the whole trials, and the scale of the time window in which acceleration peaks of biking movements occur. To do so, it is necessary to consider the evolution of speech acoustics within such time windows, and to see if there is a consistent event (like a bump) in this evolution at the occurrence of cycling peak accelerations, or if the acoustics stay linear. Investigating non-linearity in the time course of acoustic parameters is relevant for testing whether there is a local effect of the physical impulse on acoustics, resulting in a ‘bump’ in the acoustic time course within 400 ms around the acceleration peaks.

After pre-processing F0 and the amplitude envelope (see Chapter 3 Section 3.6.3.2), the movement and synchronous acoustic signals were paired for *legMot* and *armMot* conditions. As there was no biking signal in the *armBlock* condition, a surrogate condition was created. For each *speech_ armBlock* trial, a biking signal was randomly chosen from the *legMot* or *armMot* conditions of another participant. The longest signal was then cut to align with the shortest one.

Each biking cycle encompasses two acceleration peaks, corresponding to the effort on the right pedal and the effort on the left pedal. We considered the acceleration peaks of both pedals. To extract them, we used the signal of the right pedal only: since the bike is a rigid body, the acceleration applied on the left pedal was fully reflected on the right pedal. Since the goal of the study is to investigate the impact of motion on F0 and intensity, only cycles with speech, but not within a pause, were extracted.

To analyse both scales at the same time, a generalized additive mixed model (GAMM) was applied to the F0 and amplitude envelope time series using a time window of 200 ms before and after the acceleration peak of each biking cycle (400 ms in total). The GAMM uses basis functions, called smooth, to estimate the shape of the curve over time. The time series of F0 and amplitude envelope were averaged by participant, condition, and day to follow the method of Pouw and collaborators (2020a) and avoid high intra-speaker variability due to spontaneous speech. This variability was also compromising the goodness of fit of the model and the compliance with the model assumptions. Non-linearity in the armMot and legMot conditions was compared to that of the surrogate armBlock condition. To test whether the biking conditions are non-linear and different from the surrogate condition, a smooth was set per condition for both F0 and amplitude envelope. For the F0 analysis, sex, condition and day were also added as fixed factors. After a backward selection, all the factors remained in the model.

The differences in the overall level of the amplitude envelope between the conditions was an artefact from different mouth-microphone distances. Since the participant was moving her/his trunk forward to reach the bike when biking with the arms, the mouth-microphone distance changed, and so did the intensity. The amplitude envelope was then z-scaled per trial. Because of this transformation, the trials in the different conditions and days were no longer comparable. Thus, we did not add sex, condition and days for predicting the amplitude envelope, and the analysis differed from the F0 model.

For both F0 and envelope, the participant variable was included as a non-linear random effect with a random smooth for participant and condition. Since this analysis focuses on F0 and amplitude envelope as time-series, the correlation from one point of the time-series to the next has to be considered: the value at a given time t is dependent on the value at the time t-1, and possibly before. A term to consider the potential autocorrelation between the consecutive points within the time series was therefore included. We ran the analysis following the tutorial of Martijn Wieling (2018). The non-linearity of the curves and the difference between the curves across the conditions were tested in two different models (see Wieling, 2018).

The generalized additive models measure the effect (general and local) of limb conditions on speech acoustics:

- $F0 \sim sex + day + limbCondition + smooth(limbCondition) + random(smooth(1{+}condition|participant))) + autocorr$

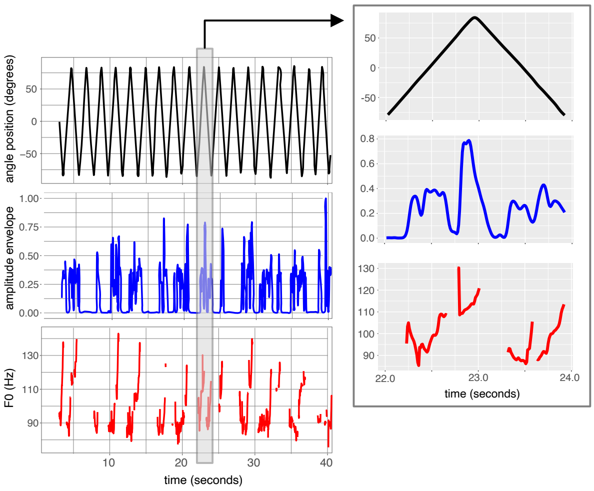


Figure 4.8: Example of analyzed signals for one speaker in the legMot condition. Left: sample of biking angular trajectory (top) and corresponding parameters in the acoustic signal: amplitude envelope (middle) and F0 (bottom). Right, zoom in on one biking cycle for each parameter. This figure has been submitted in a journal paper to *Journal of Neurophysiology*.

- $\mathit{ampEnv} \sim \mathit{limbCondition} + \mathit{smooth}(\mathit{limbCondition}) + \mathit{random}(\mathit{smooth}(1{+}\mathit{condition}/\mathit{participant}))) + \mathit{autocorr}$

4.3.2.2 Results: do cycling movements impact F0 and intensity?

The effect of the condition on the time series of the acoustic parameters around the movement acceleration peak is illustrated in Figure 4.9.

Table 4.12 summarizes the overall effects of the factors sex, day, and condition on F0. The overall level of F0 is higher for females than for males ($b = 97.39$, $z = 19.11$, $p < 0.0001$). As we can see in Figure 4.9, it is also lower on day2 compared to day 1 ($b{=}3.8, z{=}14.67, p{<}383\ 0.0001$). The overall level of F0 does not increase significantly when biking with the legs or with the arms as compared to when the hands are blocked.

The output of the generalized additive mixed models testing for non-linearity is given in Table 4.11 and the one for differences between the conditions in Table 4.13. For F0, the only significant non-linearity is when biking with the legs ($edf = 1$, $F = 5.5$, $p{=}0.02$; see Table 4.11). The curves in the legMot and armBlock conditions are not significantly different. The same is true for the armMot and armBlock conditions (see Table 4.13). For the amplitude envelope, the curves are significantly non-linear over time in the armBlock and legMot conditions ($edf = 4{,}47$, $F = 3{,}17$, $p = 0.007$; $edf = 6.05$, $F = 7.77$, $p < 0.0001$), with a tendency to be non-linear in the armMot condition (edf = 4.55, F = 2.15, p = 0.05). The curve in the legMot condition is significantly different from the curve in the armBlock condition (edf = 5.22, F = 2.35, p = 0.03).

After describing speech in quantitative terms with the acoustics and the fluency, the next section assesses the effect of the condition and the day on the content of the recall.

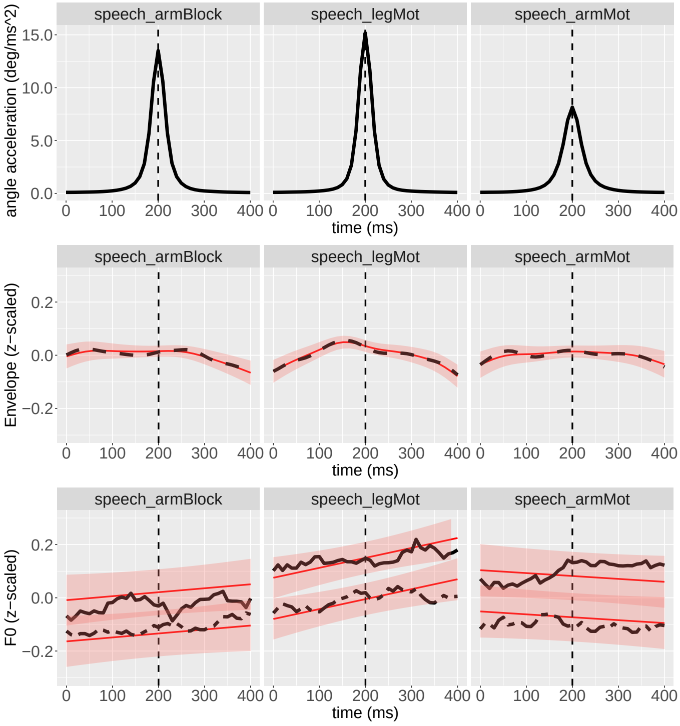


Figure 4.9: Average acceleration of the biking movement (top), amplitude envelope (center) and F0 (bottom) of the speech signal 200 ms before and after the biking acceleration peak. Red: smooth estimates of the evolution of the acoustic parameters over time, per condition. The shadow part represents the 95% confidence interval. For F0 (third line), the solid line on F0 represents the first day. The irregular dashed line represents the second day. The vertical dashed line represents the time of occurrence of the biking acceleration peak. This figure comes from the paper under review for *Journal of Neurophysiology*.

Table 4.11: Approximate significance of the non-linearity of the smooth terms. 'edf' can be seen as an estimate of the number of parameters needed to compute the smooth. 'Red.df' is the number of degrees of freedom for testing the hypothesis.

		edf	Red.df	F	P
Env (z-scaled)	Smooth_armBlock	4.47	5.01	3.17	0.007
	Smooth_armMot	4.55	5.07	2.15	0.05
	Smooth_legMot	6.05	6.63	7.77	<2e-16
F0	Smooth_armBlock	1.00	1.00	1.28	0.26
	Smooth_armMot	1.00	1.00	1.45	0.23
	Smooth_legMot	1.00	1.00	5.5	0.02

Table 4.12: Parametric coefficients of the intercepts of the GAMM for F0.

	contrast	estimate	Std.Error	t-value	p-value
F0	female-male	97.39	5.09	19.11	<2e-16
	day1-day2	3.8	0.26	14.67	<2e-16
	armBlock - legMot	1.71	5.54	0.31	0.76
	armBlock-armMot	3.43	5.5	0.62	0.53

Table 4.13: Approximate significance of the difference between the smooth terms in the different conditions. 'edf' can be seen as an estimate of the number of parameters needed to compute the smooth. 'Red.df' is the number of degrees of freedom for testing the hypothesis.

		edf	Red.df	F	P
Env (z-scaled)	S(legMot - armBlock)	5.52	5.76	2.35	0.03
	S(armMot - armBlock)	1.00	1.00	1.15	0.28
F0	S(legMot-armBlock)	1.00	1.00	0.68	0.41
	S(armMot - armBlock)	1.00	1.00	2.27	0.13

4.3.3 Do movement conditions and days have an effect on speech parameters related to content recall?

In this part, the question addressed is whether participants recall more information units in one condition or one day compared to the others. More specifically, do participants recall more information when they can gesture compared to no gesturing or constrained biking? Do they recall more information on day 2 than on day 1? As for the long-term recall, do they recall more information on day 3 than on day 2 ?

4.3.3.1 Statistical analyses

The analyses were divided in two parts: one model was applied on the short-term recalls of the first two days, and includes limb conditions. One model was applied on long-term recalls of days 2 and 3.

The model on short-term recalls included limb condition and day as fixed effect factors, and participant as a random effect factor. The story was also considered as a random effect factor, and the position in which the condition took place was among the fixed effect factors.

The model on the short term recalls is:

- *ContRec ∼ limbCondition * day + position + random(1|participant) + random(1|story)*

The model applied on long-term data included the day as a fixed factor, and participant as a random factor. As for content recall, the model applied on long-term data also included condition and position as fixed effects.

The model on the long term recalls is:

- *ContRec ∼ limbCondition * day + position + random(1|participant) + random(1|story)*

4.3.3.2 Results: Do limb movements and days impact content recall?

Table 4.14 summarizes the distribution of each variable per limb condition and day. The results of linear mixed models are available in Table 4.15. Figure 4.10 summarizes these analyses.

Table 4.14: Summary of the distribution of scores on content recall (*ContRec*), per condition and day

day	typeRecall	condition	min	mean ± sd	max
1	shortTerm	armBlock	0.24	0.42 ± 0.13	0.66
1	shortTerm	armFree	0.26	0.43 ± 0.1	0.64
1	shortTerm	legMot	0.17	0.43 ± 0.12	0.73
1	shortTerm	armMot	0.18	0.43 ± 0.11	0.64
2	shortTerm	armBlock	0.33	0.53 ± 0.11	0.73
2	shortTerm	armFree	0.22	0.54 ± 0.13	0.8
2	shortTerm	legMot	0.24	0.54 ± 0.14	0.8
2	shortTerm	armMot	0.28	0.53 ± 0.13	0.8
2	longTerm	armBlock	0.06	0.26 ± 0.11	0.42
2	longTerm	armMot	0.04	0.25 ± 0.12	0.46
2	longTerm	legMot	0.05	0.25 ± 0.14	0.53
2	longTerm	armFree	0.04	0.23 ± 0.12	0.5
3	longTerm	armBlock	0.17	0.33 ± 0.12	0.57
3	longTerm	armMot	0.11	0.34 ± 0.15	0.62
3	longTerm	armFree	0.07	0.34 ± 0.14	0.56
3	longTerm	legMot	0.13	0.34 ± 0.14	0.65

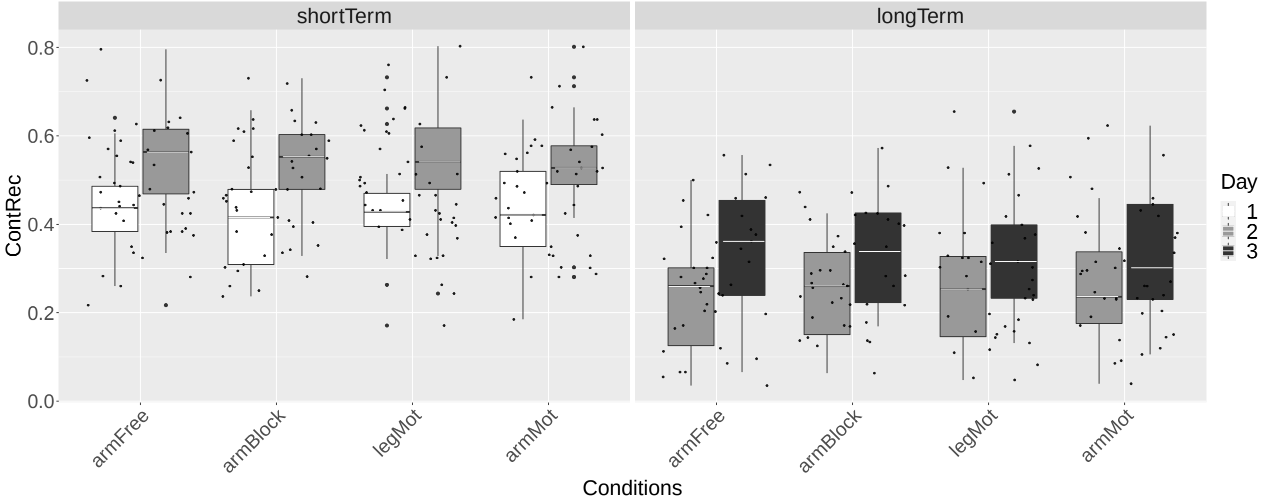


Figure 4.10: Distribution of the scores of the recalls per limb condition and days.

There is no effect of the condition on the retrieval of the narrative content, but there is an effect of the day: scores improve significantly from one day to the day after (short-term:

Table 4.15: Post-hoc comparisons from models estimating the effect of limb conditions and days on content recall (*ContRec*).

type of recall	contrasts	estimate	Std.Error	z-value	p-value
short-term	day 2 - day 1	0.42902	0.04059	10.57	0.01
long-term	day 3 - day 2	0.50605	0.06701	7.552	0.01

b = 0.43, z = 10.57, p = 0.01; long-term: b = 0.5, z = 7.5, p = 0.01).

4.4 Overall Discussion

The aim of this chapter was to examine the interactions between speech and limb movements in different limb conditions and over time. The results are discussed in the same order as they were presented.

4.4.1 Why do participants gesture less over time?

Participants tend to move their hands less with the day, while they recall more information with time. One possibility is that participants gesture less with task repetition. People make fewer lexical gestures when speech is rehearsed (Chawla and Krauss, 1994), and less gestures when they have to tell the story again to the same listener (Jacobs and Garnham, 2007).

Another possibility is that the quantity of hand movements may decrease along with the consolidation of memory: while on the first day, the cognitive load within working memory may be high because the stories just have been watched for the first time, it may decrease on the second and the third day, because of a possible transfer of the information from the working memory to a long-term memory, helped by sleep (Cherdieu et al., 2018). This decrease of cognitive load in working memory would result in a decrease in hand movements. Indeed, as seen in chapter one, gestures tend to emerge with high working memory cognitive load (Chu and Kita, 2011; Cook et al., 2012). This suggestion is supported by the increase over time of content recall performances. The evolution of speech parameters also suggest a better management of the cognitive load over time. Participants speak more from day 2 to day 3, faster from day 1 to day 2, and do fewer silent pauses from day 1 to day 2 and from day 2 to day 3 (deduced from the increase of *SpeechTime*, which corresponds to a decrease of proportion of silent pauses over a trial). The number of syllables encompasses both word syllables and disfluencies. It would be interesting to distinguish filled pauses from word syllables, to see if word syllables increase and disfluencies decrease over time.

The use of both qualitative annotation of gestures and quantitative analyses of the movement signal are complementary. To estimate the quantity of hand movements, the average 3D distance of the hands (by seconds) has been taken as the principal variable. This quantity decreases over time, but without the segmentation of gestures, it would not be possible to know if this decrease is due to the fact that participants move more slowly or move less. The distance covered by the hands only informs about hand displacements, but cannot tell anything about the number of co-verbal gestures. The use of motion capture is more objective and reliable in terms of body movement analyses, but the equipment is heavy and may not be comfortable for the participants. This discomfort could inhibit the natural gesture dynamics that they experience in everyday life. The best way of tracking gestures reliably and in an ecological context may be to use algorithms of video tracking (Pouw et al., 2020e).

4.4.2 What can biking cycle duration tell about speech-limbs interactions?

Speaking makes arm biking cycles longer and more variable. Previous studies also found that talking while drawing circles (Dayalu et al., 2013) of finger tapping (Verghese et al., 2007) lowered movement frequency, and increase its variability. Verghese and colleagues (2007) also found that walking while talking lowered stride frequency. We find that talking lowers leg biking frequency, but does not increased its variability compared to quiet leg biking. Besides, speech arm cycles are longer and more variable than speech leg cycles, but this difference is not present when quiet. These results suggest stronger interference between speech and upper limbs than speech and lower limbs. Most of the studies investigated walking because of the crucial stakes that the dual-task walking-talking represents for elderly people (Li et al., 2001), but the effect of talking on cycling dynamic has not been widely described. As well as examining the effect of speech on cycling, the motor activity we chose enables the comparison between upper and lower limbs and points towards a stronger interference between speech and arms. This is not surprising, since it has been showed that motor control systems of spoken language and arm gestures overlap (Gentilucci and Volta, 2008). Our result suggests that the speech-arm link is also at work with non-communicative motor activity, supporting the idea of a general motor link between the two that goes beyond the specific task of speech, and can be elicited during other tasks. This link could be used for motor rehabilitation of the arms or the articulators (Farley and Koshland, 2005). On the other hand, biking with the arms is not a familiar motor activity. The novelty of the task may influence its regularity. The variability of both leg and arm cycle durations decreased with day, suggesting that participants were more familiar with the task on the second day. This decrease in variability also goes with the increase of performance in content recall. Variability in a concurrent motor task may be a marker of the cognitive load involved in the speech task, especially when moving the upper limbs.

4.4.3 Why don't limb movements have an effect on speech fluency?

On the variables analysed, the arm-free condition do not have a different effect on speech compared to the biking conditions, or no-movement condition. This result is in line with studies using dual-task paradigms and not finding any effect of motor task on speech (Verghese et al., 2007; Hoetjes et al., 2014; Whitfield et al., 2019). Whitfield (2019) measured speech rate, mean pause duration, F0 variability and formant space during drawing a circle, and did not find any effect of the motor task on speech fluency. Hoetjes and colleagues (2014) did not find any effect of not being able to gesture on the number of words, on speech rate or on the rate of filled pauses. However, Hostetter and colleagues (2007a) found that not being able to gesture increased the proportion of units starting with 'and' and decreased percentage of events described with rich verbs. Kemper and collaborators (2003) found an effect of walking and complex finger tapping on grammatical complexity and propositional content. It seems that limb movements have an effect on grammatical and propositional properties of speech, but not on speech fluency. We did not conducte any semantic or syntactic analyses, but it could be interesting to see if the limb conditions have an effect on these dimensions of speech. It could be also that in our experiment, participants may not spontaneously gesture enough to enhance an effect of hand gestures on speech. This can be due to the discomfort of the motion capture equipment.

4.4.4 Does the effect of biking on speech acoustics depend on the acceleration magnitude?

Concerning acoustic analyses of speech, contrary to the literature, our results did neither show a correlation between the acceleration peaks during biking motion and fundamental frequency peaks, nor a correlation with the speech amplitude envelope. We also investigated the potential non-linearities of the acoustic parameters around the acceleration peaks. For F0, the data did not show any effect in any of the biking conditions. For the amplitude envelope, however, acceleration peaks yielded a significant non-linear effect when biking with the legs but not with the arms. The absence of an effect is interpreted in line with a threshold of movement acceleration that is needed to induce the cascading mechanisms of gesture-speech physics. In Pouw et al. (2020a), acceleration peaks were higher than 4 cm/s^2 with one maximal value at 16 cm/s^2. In our study, acceleration values for arm motion ranged from 0.4 to 10 cm/s^2 with a median value of 2 cm/s^2 while for leg motion it ranged from 0.5 to 7.4 cm/s^2, with a higher median value of 3.5 cm/s^2 relative to the arms. When participants biked with their legs, they produced higher acceleration peaks than when they biked with their arms. This is congruent with the fact that legs have larger muscles, they are heavier and have an advantage for endurance tasks such as biking (Calbet et al., 2005). Although we did not find any linear relationship between acoustic peaks and physical impulse, we did find a non-linear effect of the leg acceleration peaks on the time course of intensity. It means that the legs' physical impulse promptly increases intensity, but the magnitude of this increase does not depend on the magnitude of the acceleration. Again, this result can be due to the range of values of acceleration. This range may not be wide enough, with higher values, to see a linear effect.

4.4.5 Content recall rating: a questionable methodology

The way content recall was scored can be debated. This technique of scoring was adopted in order to give a first idea of the free recalls' content. Indeed, only the audio of the stories were taken into account, and not the potential information units' recall that the images could have triggered. Besides, the grid used does not characterize the structure of the recalls: for instance, it does not distinguish participants who talked a lot but do not recall a lot of information, and participants recalling more information but in a more concise and short talk. In order to characterize the content of the recalls in more details, a more complete method of annotating was set-up, but not yet used. In this method, the score relies directly on the recall of the participants, by evaluating the different characteristics of the information units. Whether an information unit affords a valuable information, or if it is a repetition/exaggeration/invention, if the information unit is visual or auditive, are considered in the evaluation. But the annotations using this grid take more time and the agreement rate may not be as high as for the first grid. This work has to be continued. The recall of the new words is another important marker of the evolution of the content over time. Analyses of the recall of the new words have been conducted on the 20 first participants, by our research assistant during her internship. The pronunciation of the new words were phonetically annotated and a score representing the distance between the original and the pronounced word was calculated. In short-term free recall, there is no effect of limb condition on the recall of the new words. In long-term recall, scores are better in the biking conditions compared to the arm free condition, although the difference is not significant. Note that the progression between the second and the last day for the arm biking condition is more important than for the leg biking condition. It is possible that the motor task first interfered with memorization, and then, after learning, the movement facilitated memorization. It would be interesting to pursue this learning further in time to see if this effect persists.

4.4.6 Conclusion

In its content or in its fluency, speech is mainly influenced by the day, but there is no effect of the type of limb movements on the speech parameters analysed. Analysing other speech parameters linked to the content, such as richness of vocabulary could be more informative. Besides, we can notice some pitfalls in the parameters chosen. The patterns chosen were computed on the whole trials, but speech dynamics at the scale of the breathing cycle may be different. For instance, although the speech rate at the scale of the trial does not change from one condition to another, it may be that there are more syllables per breathing cycle, but also more hesitations, which lengthen the duration of the trial. Since breathing is inherent to speech and speech shapes respiration for its own purpose, analyses at the scale of breathing cycles are necessary to deeper analyse speech dynamics and organization. The next chapter will introduce the analysis of breathing to better understand the interactions between speech and limb movements, and how they share this common resource.

CHAPTER 5

The interactions between speech and limb movements through the lens of breathing

---

Speech is structured in breathing cycles due to ventilation needs. These cycles are yet also shaped by communication constraints with a short inhalation phase and a long exhalation phase. This chapter analyzes respiratory signals and how they structure speech and interact with movement. Describing breathing, and in particular speech breathing, during movement is an important issue and some questions have not been explored, in particular the respiratory stability for the same speaker over time and according to movement condition (see Chapter 2 section 2.2.2). The aim of this chapter is to analyse the limb movement-speech interaction through the physiological process of breathing. The first section describes the impact of speech and limb conditions on breathing cycles over time. The second section details the analyses of speech parameters at the scale of the breathing cycles. The third section investigates whether speech breathing is speaker-specific, and if this specificity is consistent across limb conditions and over time. The fourth section examines the motor respiratory coupling on biking movements, in quiet and speech conditions. The results of the first and the last section of this chapter have been published in *the Annals of the New-York Academy of Sciences* in August 2021 (Serré et al., 2021). Some passages of this paper have been included without changes.

Our contribution is: (1) to analyse speech through the window of breathing to specify speech re-organization across limb conditions and over time; (2) to assess, in line with previous work, speaker-specific respiratory stability in speech. Speech breathing stability has not been widely studied before; (3) to investigate the motor respiratory coupling in presence of natural speech. The methodologies used to address these points are adapted from physiological studies related to ventilation. These analyses should provide a more precise picture of the entanglement between speech, breathing and limb movements.

5.1 Impact of limb motion on breathing cycles over quiet and speech conditions

As mentioned in Chapter 2 Section 2.2.4, on the one hand, speaking triggers restructuring of breathing cycles through neural connections specific to voluntary breathing. On the other hand, physical activity modulates breathing patterns for metabolic needs through neural connections specific to automatic ventilation. When both speech and limb movements occur at the same time, the main strategy is to maintain speech quality by reducing the airflow (compared to being quiet for the same level of physical activity). Yet, without deteriorating speech quality, speech breathing patterns such as the symmetry of the cycle or the airflow shape may be impacted by limb movements.

5.1.1 Analysis of the effect of motion, speech, and day on breathing.

We first characterized the effect of speech, motion, and day on the cycle duration (*cycleDur*) and symmetry (*cycleSym*) of the breathing cycles to allow for comparison with previous work and to provide an overview of the parameters related to the breathing cycle. The variables are defined in Chapter 3 section 3.6.1.

Linear mixed models were used to assess the effects of limb motion, speech, and day on both the log of breathing cycle duration and the logit of symmetry (data were transformed for normality constraints). Participant was included in the model as a random effect. A term to consider the potential autocorrelation between the consecutive breathing cycles within a participant's session (*autocorr*), and the position of the cycles (*cyclePosition*) were included in the model. Nonhomogenous variance across conditions (*var*) was corrected for when present. A backward model selection was used to select significant factors. These models were applied on the parameters of all cycles for each participant, day, and condition.

The first set of models was applied to the data from day 1. Its aim was to examine the effect of limb motion on breathing cycle duration and symmetry in quiet and speech conditions. Note that we did not compare quiet and speech conditions as it is already evident from previous work that speech breathing cycles are longer and clearly less symmetric than quiet breathing cycles. Three factors were included in the models: *vocalCondition (speech; and quiet)*, *limbCondition (legMot; armMot; noMot)* and *cyclePosition.* Since participants did not move their hands in either the *quiet_ noMot* or the *speech_ armBlock* conditions, *speech_ armBlock* was considered to be the equivalent speech condition of *quiet_ noMot.*

After the statistical steps described in Chapter 3 Section 3.7, the final models are:

$cycleDur \sim cyclePosition + limbCondition * vocalCondition + random(1 + limbCondition + vocalCondition|participant) + var(vocalCondition) + autocorr$

$cycleSym \sim cyclePosition + limbCondition * vocalCondition + random(1 + limbCondition + vocalCondition|participant) + var(vocalCondition) + autocorr$

The second set of models was applied to the data from speech conditions only. The aim was to examine the effect of limb motion *(armFree; armBlock; legMot; armMot)* and day (*day 1; day 2*) on speech breathing cycle duration and symmetry. Three factors were included: limb motion *(armFree; armBlock; legMot; armMot)*, day (*day 1; day 2*) and cycle position ($cyclePosition$).

After the statistical steps described in Chapter 3 Section 3.7, the final models are:

$cycleDur \sim limbCondition + cyclePosition + day + random(1 + day|participant) + autocorr$

$cycleSym \sim cyclePosition + limbCondition * day + random(1 + day|participant) + autocorr$

The third set of models was applied to the data from long term recalls, from day 2 and 3. Its aim was to examine the effect of time (*day 2; day 3*) on speech breathing cycle duration and symmetry. The models included two factors: day (*day 2; day 3*) and cycle position ($cyclePosition$).

After the statistical steps described in Chapter 3 Section 3.7, the final models are:

$cycleDur \sim cyclePosition + random(1 + day|participant)$

$cycleSym \sim cyclePosition + jourExp + random(1 + day|participant)$

5.1.2 Results: how does limb motion impact breathing cycle duration and symmetry?

A summary of the distribution of the number of breathing cycles as well as their duration and symmetry are presented in tables 5.1, 5.2 and 5.3. Only significant differences are described in the results section below; all the other comparisons did not result in significant differences.

Table 5.1: Summary of the distribution of the number of breathing cycles, per condition and day.

day	condition	min	mean ± sd	max
1	quiet_noMot	13	30.72 ± 9.07	49
1	quiet_legMot	17	36.2 ± 9.09	52
1	quiet_armMot	20	34.6 ± 9.83	57
1	speech_armMot	6	24.72 ± 8.77	47
1	speech_armFree	7	24.4 ± 8.09	49
1	speech_armBlock	10	24.16 ± 8.57	44
1	speech_legMot	11	25.48 ± 7.62	39
2	speech_armMot	10	23.65 ± 8.38	46
2	speech_armFree	9	23.48 ± 8.38	41
2	speech_armBlock	7	23.48 ± 9.26	49
2	speech_legMot	9	26 ± 9.7	47
2	speech_longTerm	18	69.74 ± 36.39	171
3	speech_longTerm	30	77.35 ± 35.66	160

Table 5.2: Summary of the distribution of the cycle duration (in seconds), per condition and day.

day	condition	min	mean ± sd	max
1	quiet_noMot	2.54	4.24 ± 1.43	8.6
1	quiet_legMot	2.3	3.58 ± 1.11	7.12
1	quiet_armMot	2.1	3.7 ± 0.98	5.73
1	speech_armMot	3.33	5.42 ± 1.16	8.67
1	speech_armFree	4.08	5.23 ± 1	7.95
1	speech_armBlock	3.97	5.11 ± 0.81	7.61
1	speech_legMot	3.49	4.86 ± 0.99	7.29
2	speech_armMot	4.18	5.72 ± 1.04	7.98
2	speech_armFree	3.8	5.52 ± 1.18	8.89
2	speech_armBlock	4.22	5.51 ± 0.96	7.98
2	speech_legMot	3.86	5.24 ± 0.97	7.33
2	speech_longTerm	3.86	5.37 ± 0.92	6.99
3	speech_longTerm	3.77	5.34 ± 0.96	7.39

Table 5.3: Summary of the distribution of the cycle symmetry (in seconds), per condition and day.

day	condition	min	mean ± sd	max
1	quiet_noMot	0.32	0.39 ± 0.05	0.54
1	quiet_legMot	0.37	0.43 ± 0.03	0.51
1	quiet_armMot	0.34	0.42 ± 0.04	0.49
1	speech_armMot	0.1	0.15 ± 0.04	0.27
1	speech_armFree	0.11	0.15 ± 0.03	0.24
1	speech_armBlock	0.1	0.15 ± 0.03	0.24
1	speech_legMot	0.11	0.16 ± 0.04	0.28
2	speech_armMot	0.09	0.15 ± 0.05	0.28
2	speech_armFree	0.09	0.15 ± 0.05	0.28
2	speech_armBlock	0.1	0.15 ± 0.04	0.28
2	speech_legMot	0.08	0.15 ± 0.05	0.3
2	speech_longTerm	0.11	0.17 ± 0.05	0.36
3	speech_longTerm	0.09	0.16 ± 0.05	0.32

As displayed in Figures 5.1 and 5.2, the presence of limb motion decreased the average duration of the breathing cycle in quiet conditions (*quiet_noMot - quiet_armMot*: b=0.13, z=3.29,p=0.009; *quiet_noMot-quiet_legMot*: b=0.16, z=3.96, p<0.001); while it increased the symmetry of the cycle (*quiet_noMot-quiet_armMot*:b=-0.11, z=-3.98, p<0.001; *quiet_noMot-quiet_legMot*: b=-0.14, z=5.1, p<0.001). Both the decrease in cycle duration and the increase in symmetry are due to the shortening of exhalation combined with the lengthening of inhalation. By contrast, adding leg or arm motion when speaking (*speech_armMot and speech_legMot*) induced no statistically significant effect compared with speaking while not moving the arms (*speech_armBlock*). The position of the cycle had a significant effect the breathing cycle duration (b=-0.002, z=-5.66,p<0.001).

On average, the duration of the breathing cycle was smaller in *speech_legMot* than in *speech_armFree* (b=0.06, z=3.4,p=0.004) and in *speech_legMot* than in *speech_armMot* (b=0.09, z=4.82, p<0.001). Other comparisons were not significant. Regardless of the condition, on *day2*, the cycles tended to be longer than on *day 1*, but the difference did not reach significance (b=-0.07, z=-2.61, P=0.06). The symmetry of the cycle was larger in *speech_legMot* than in *speech_armFree* (b=-0.07, z=-3.52,p=0.003) and in *speech_legMot* than in *speech_armBlock* (b=-0.07, z=-3.54,p=0.0031). Other comparisons were not significant. The cycles were also less symmetric on *day2* than on *day1* (b=0.1, z=3.02,p=0.02). *cycleSym* decreased on day 3 compared to day 2 (b = 0.0153, z = 4.576 , p <.0001).

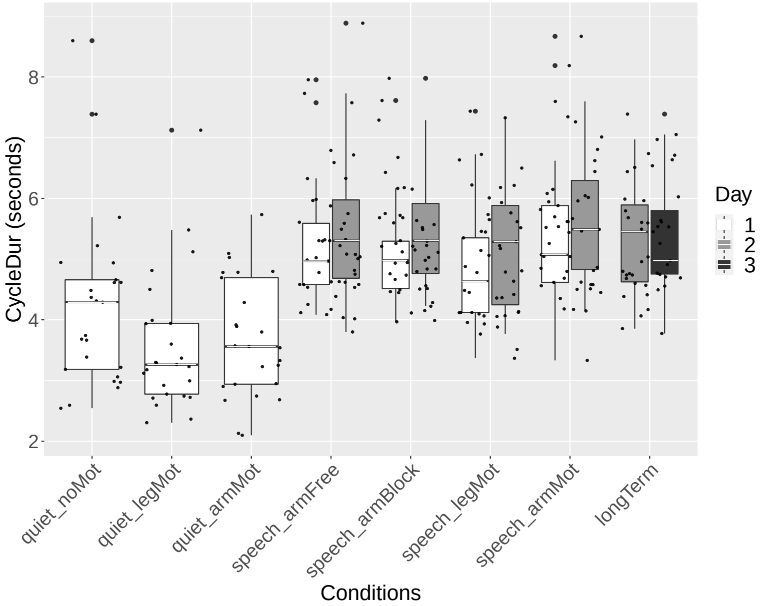


Figure 5.1: Cycle duration (*cycleDur*) in different conditions of limb movements and over days.

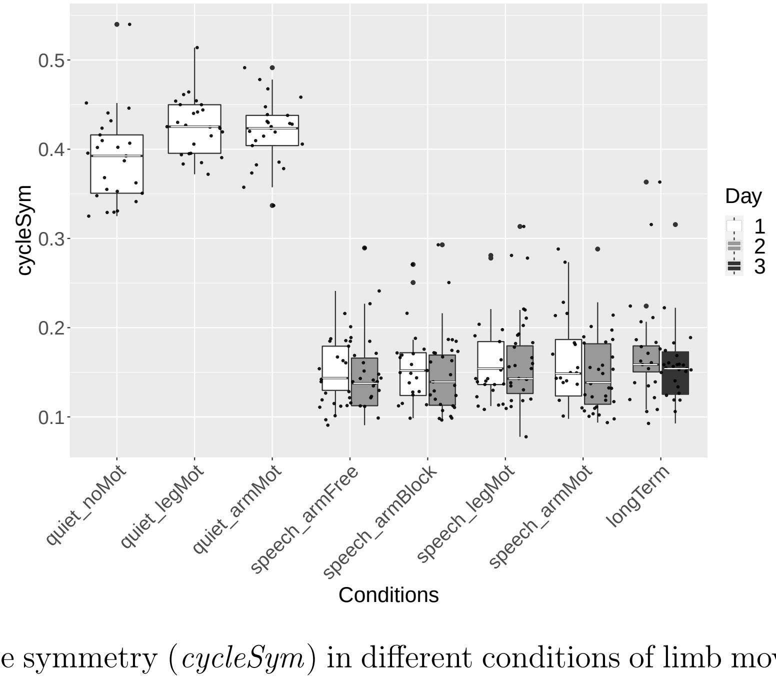


Figure 5.2: Cycle symmetry (*cycleSym*) in different conditions of limb movements and over days.

In summary, in quiet conditions, breathing cycles are shorter and more symmetric when biking than when doing nothing. In speech conditions, breathing cycle duration and symmetry increase with leg biking compared to arm biking and arms blocked. Symmetry decreases over days. Are the changes due to biking conditions impacting speech? Does the symmetry decrease on day 2 because of longer exhalation due to more syllables?

5.2 How do limb conditions and days impact speech fluency at the breath group scale?

At the scale of the trial, limb movements did not have any impact on the number of syllables, the speech rate or the rate of time spent speaking (see Chapter 4 Section 4.3.1). As seen in the previous section, breathing cycles are impacted by motor activity. Since speech is constrained by breathing capacities and shapes breathing cycles, do limb movements impact speech fluency within a breathing cycle?

As in Chapter 4, the variables chosen to characterize speech fluency are the number of syllables ($NumSyll_b$), speech rate ($SpRate_b$), and the proportion of time spent speaking ($SpeechTime_b$). The number of breathing cycles per trial ($NbCycle_b$) was also analyzed. Two factors were included in the models: limb condition (*speech_ armFree*; *speech_ armBlock*; *speech_ legMot*; *speech_ armMot*) and day (*day1*; *day2*).

5.2.1 Statistical analyses

For each variable characterizing speech fluency and content recall, the analyses were divided in two parts. The models on the short term recalls measure the effect of limb conditions and days on speech. After the statistical steps described in Chapter 3 Section 3.7, the final models are:

- $NumSyll_b \sim limbCondition * day + random(1/participant) + random(1/story)$
- $SpRate_b \sim day + random(1 \mid storyVec) + random(1 + day \mid part)$
- $SpeechTime_b \sim day + random(1 \mid participant) + random(1 \mid story)$
- $NbCycle_b \sim random(1 \mid participant) + random(1 \mid story)$

The models on the long term recalls measure the effect of the day on speech. After the statistical steps described in Chapter 3 Section 3.7, the final models are:

- $NumSyll_b \sim day + random(1/participant)$

- $SpRate_b \sim day + random(1/participant)$
- $SpeechTime_b \sim day + random(1/participant)$
- $NbCycle_b \sim random(1 \mid participant) + random(1 \mid story)$

5.2.2 Results: how do limb movements impact parameters related to speech fluency at the breath group scale?

$NumSyll_b$, $SpRate_b$ and $SpeechTime_b$ are respectively summarized in Tables 5.4, 5.6 and 5.7.

Table 5.4: Summary of the distribution of the number of syllables within a breath group ($NumSyll_b$), per condition and day. The data were first averaged over each trial.

day	condition	min	mean ± sd	max
1	speech_armMot	9.05	16.33 ± 4.48	24.83
1	speech_armFree	10.12	16.21 ± 4.66	28.35
1	speech_armBlock	7.92	15.18 ± 4.03	22.8
1	speech_legMot	7.73	14.37 ± 4.57	24.08
2	speech_longTerm	5.05	16.16 ± 4.44	25.96
2	speech_armMot	7.7	17.76 ± 5.57	31.11
2	speech_armFree	8	17.32 ± 4.92	28.63
2	speech_armBlock	8.44	17.62 ± 5.07	28.83
2	speech_legMot	6.66	16.79 ± 5.35	29.95
3	speech_longTerm	7.84	16.81 ± 4.18	25.23

Table 5.5: Post-hoc comparisons from the linear mixed models applied to $NumSyll_b$, $SpRate_b$ and $SpeechTime_b$, on short-term and long-term recalls.

variable	type_recall	contrast	estimate	SE	df	t.ratio	p.value
$SpeechTime_b$	shortTerm	day1 - day2	-0.0787	0.039	187	-2.016	0.0452
	longTerm	day2 - day3	-0.189	0.0717	42	-2.642	0.0115
$NumSyll_b$	shortTerm	day1 - day2	-0.1	0.0317	22	-3.169	0.0044
$SpRate_b$	shortTerm	nday2 - day3	-0.512	0.118	19	-4.351	0.0003

Table 5.6: Summary of the distribution of the speech rate within a breath group ($SpRate_b$), per condition and day. The data were first averaged over each trial.

day	condition	min	mean ± sd	max
1	speech_armMot	2.23	3.48 ± 0.72	5.11
1	speech_armFree	2.47	3.56 ± 0.69	4.86
1	speech_armBlock	2.02	3.39 ± 0.73	5.31
1	speech_legMot	2.41	3.39 ± 0.7	4.83
2	speech_longTerm	1.87	3.41 ± 0.76	5.13
2	speech_armMot	2.24	3.53 ± 0.73	5.15
2	speech_armFree	2.41	3.63 ± 0.68	5.03
2	speech_armBlock	2.13	3.7 ± 0.77	5.42
2	speech_legMot	2.37	3.65 ± 0.75	5.34
3	speech_longTerm	2.42	3.59 ± 0.61	5.16

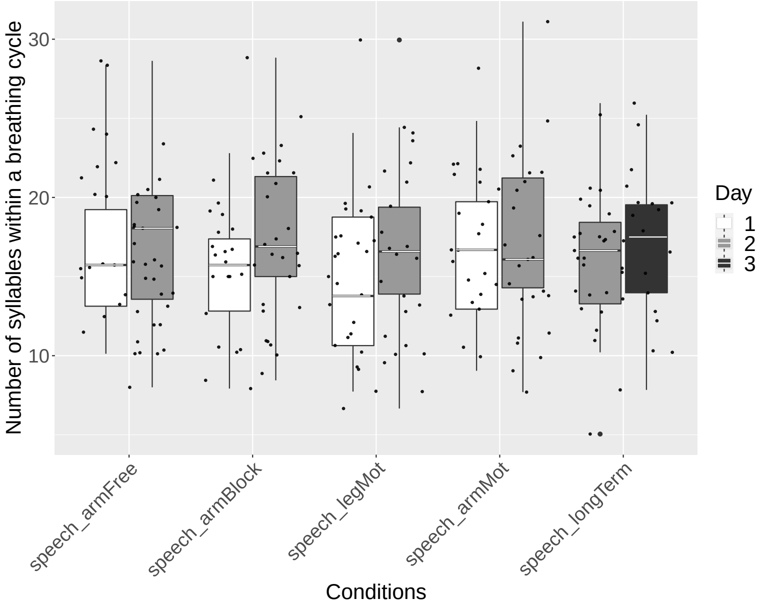


Figure 5.3: Distribution of $NumSyll_b$ within a breath group per limb condition and day. Each black point represents the average number of syllables for one participant.

Figures 5.3, 5.4 and 5.5 illustrate the data respectively for $NumSyll_b$, $SpRate_b$ and $SpeechTime_b$. The results of the linear mixed models are available in table 5.5. In the short-term recalls, $NumSyll_b$ and $SpeechTime_b$ increased from day 1 to day 2 ($NumSyll_b$: b = -0.1, z= -3.2, p = 0.004; $SpeechTime_b$: b = -0.08, z = -2.02, p = 0.04). In the long-term recalls, $SpRate_b$ and $SpeechTime_b$ increased from day 2 to day 3 ($SpRate_b$: b = -0.5, z =

-4.3, p = 0.0003; *$SpeechTime_b$*: b = -0.2, z = -2.64, p = 0.01). Neither the day nor the limb condition had an effect on the number of breathing cycles per trial.

Table 5.7: Summary of the distribution of the proportion of time spent speaking within a breathing cycle (*$SpeechTime_b$*), per condition and day. The data were first averaged over each trial.

day	condition	min	mean ± sd	max
1	speech_armMot	0	0.82 ± 0.21	0.98
1	speech_armFree	0	0.73 ± 0.31	0.98
1	speech_armBlock	0	0.79 ± 0.24	0.98
1	speech_legMot	0	0.72 ± 0.3	0.98
2	speech_longTerm	0	0.71 ± 0.27	0.98
2	speech_armMot	0	0.8 ± 0.22	0.98
2	speech_armFree	0	0.66 ± 0.34	0.98
2	speech_armBlock	0	0.76 ± 0.31	0.98
2	speech_legMot	0.31	0.8 ± 0.2	0.98
3	speech_longTerm	0	0.8 ± 0.25	0.98

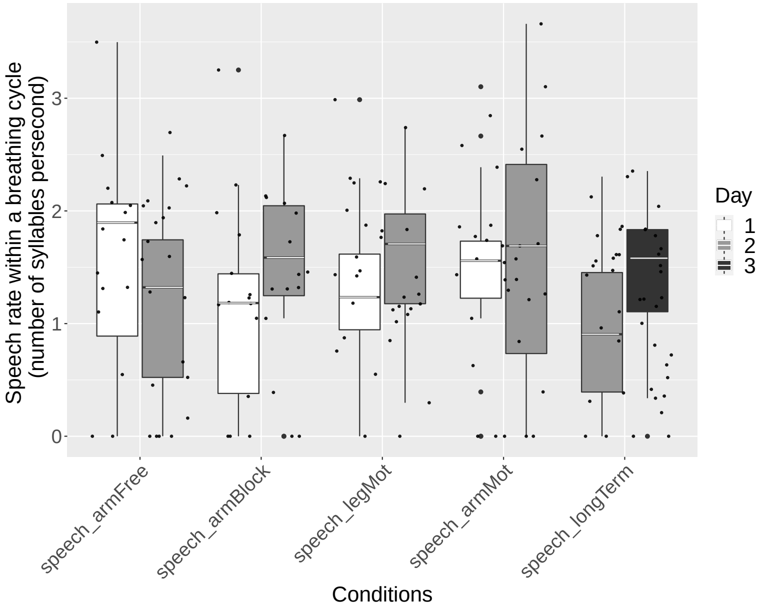


Figure 5.4: Distribution of the speech rate (*$SpRate_b$*) within a breath group per limb condition and day. Each black point represents the average speech rate for one participant.

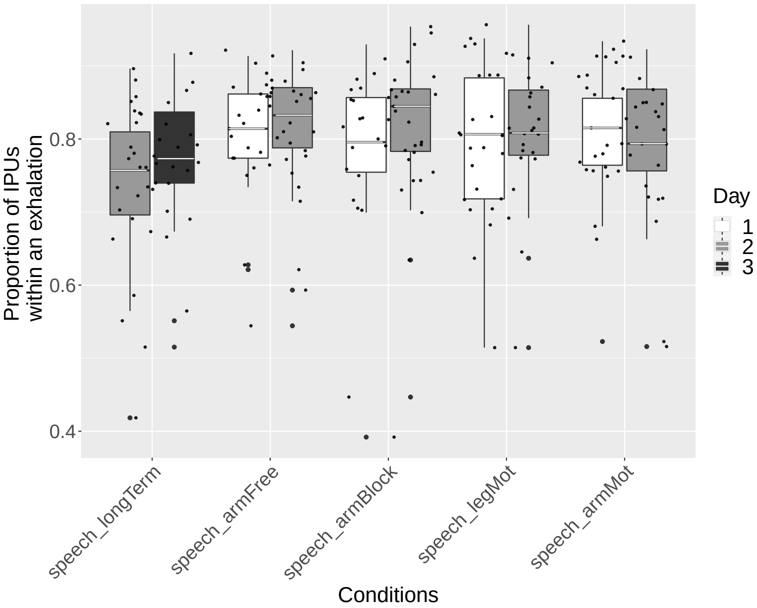


Figure 5.5: Distribution of the proportion of time spent speaking (*SpeechTime$_b$*) within a breath group per limb condition and day. Each black point represents the average proportion of speech time for one participant.

In summary, participants produced more syllables and spent more time speaking (compared to silent pauses) within a breathing cycle on day 2 than on day 1, and and spent more time speaking on day 3 than on day 2. They also spoke faster within a breathing cycle on day 3 than on day 2 in the long-term recalls.

Participants speak more from one day to another, and they do so by adding more syllables within each breathing cycle, and not by increasing the number of breathing cycles. Exhalations are longer from day 1 to day 2, and from day 2 to day 3. Breathing cycles may adapt to speech fluency improving over time. For instance, exhalations are longer and cycles are less symmetric when speakers produce more syllables. If speech fluency is speaker-specific, this specificity may be reflected in speech breathing.

5.3 Is speech breathing speaker-specific?

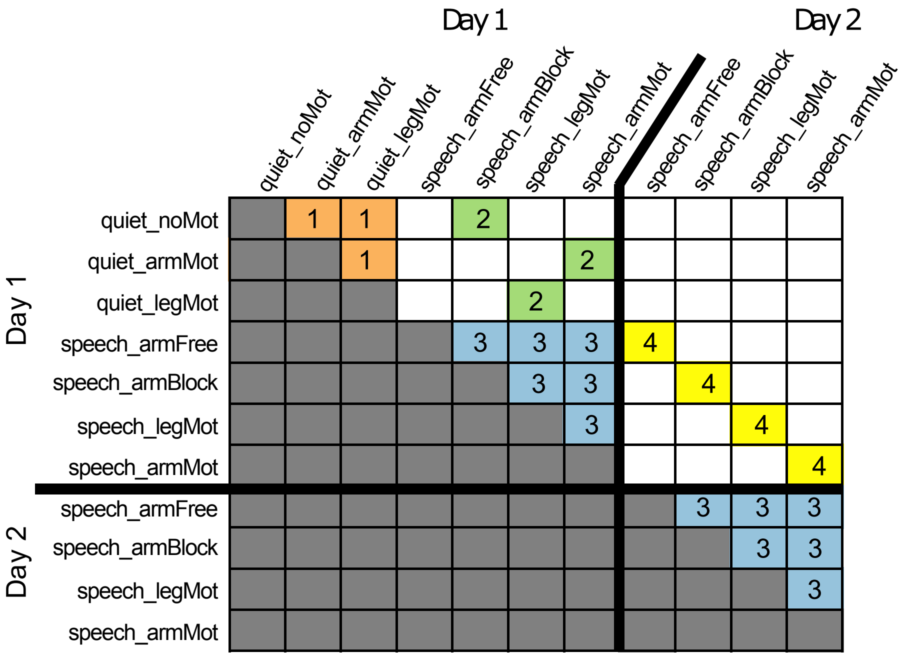
We assess the individuality of speech breathing in spontaneous speech by focusing on three general questions: (1) Is quiet breathing individuality related to potential speech breathing individuality? (2) Can speech breathing individuality be observed across multiple days, as previously observed for other non-spoken tasks? (3) Is speech breathing individuality sensitive to the global pressure induced by leg motion on the respiratory system and/or more specific changes induced by arm motion? We address these questions by applying methods previously used in physiological studies to assess the “ventilatory personality,” or “breathing individuality”. The work described in this section has been published in *the Annals of the New-York Academy of Sciences* in 2021 (Serré et al., 2021).

5.3.1 Analysis of within-participant consistency across conditions and days.

The consistency of the breathing cycles is assessed on cycle duration *cycleDur*, cycle symmetry *cycleSym* and cycle shape *cycleShape*. Their definition are detailed in Section 3.6.1 of Chapter 3.

In agreement with the literature, the individuality of the breathing patterns was assessed using the method developed by Benchetrit et al. (1989). The aim was to determine whether within-speaker variability is lower than between-speaker variability across conditions and over time for a given parameter describing the breathing cycle. This would be the case if the difference between two conditions for the same participant were to be significantly smaller than the difference between two participants in the same or another condition. This difference was quantified by computing the distances between a given set of values of each respiratory cycle parameter for one participant and one condition and each of the 49 other sets of values of the same parameter (2 conditions × 25 participants = 50 sets of values). Benchetrit et al. (1989) used the Mahalanobis distance. However, in our dataset, the distributions of the values of the breathing parameters per speaker were not Gaussian, and non normality was observed for all parameters. For this reason, we used the Cramér–von Mises distance (Baringhaus and Henze, 2017). Instead of using the mean and variance of the distribution of parameter values, the Cramér–von Mises distance is directly based on the distribution functions (Baringhaus and Henze, 2017).

For *cycleDur*, *cycleSym*, and *cycleShape*, S_{ij} corresponds to the distribution of the considered parameter for participant i in condition j. For each S_{ij}, the 49 distances to all the other distributions were computed. These distances were then ranked from the smallest to the largest. This procedure resulted in 50 sets of the 49 sorted distances. The ranks of the within-participant distances were then summed ($T_{within}(T_w)$ statistic) as well as the ranks from 25 pairs $(S_{ij}, S_{i'j'})$ taken randomly ($T_{between}(T_{btw})$ statistic). To estimate the distribution of T_{btw}, the Monte Carlo technique (Metropolis and Ulam, 1949) was used: a sample of 25 pairs $(S_{ij}, S_{i'j'})$ was taken randomly 1000 times to get 1000 values of T_{btw}. The null hypothesis (T_w is not different from T_{btw}, from condition j to condition j’) is rejected when the p-value resulting from the comparison between T_w and the T_{btw} distribution is smaller than 0.05. Bonferroni correction was applied by multiplying the p-value by the number of comparisons required for the tested hypothesis (corresponding to the numbers



		Day 1							Day 2			
		quiet_noMot	quiet_armMot	quiet_legMot	speech_armFree	speech_armBlock	speech_legMot	speech_armMot	speech_armFree	speech_armBlock	speech_legMot	speech_armMot
Day 1	quiet_noMot	▨	1	1		2						
	quiet_armMot	▨	▨	1				2				
	quiet_legMot	▨	▨	▨			2					
	speech_armFree	▨	▨	▨	▨	3	3	3	4			
	speech_armBlock	▨	▨	▨	▨	▨	3	3		4		
	speech_legMot	▨	▨	▨	▨	▨	▨	3			4	
	speech_armMot	▨	▨	▨	▨	▨	▨	▨				4
Day 2	speech_armFree	▨	▨	▨	▨	▨	▨	▨	▨	3	3	3
	speech_armBlock	▨	▨	▨	▨	▨	▨	▨	▨	▨	3	3
	speech_legMot	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	3
	speech_armMot	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨

Figure 5.6: Summary of the 22 comparisons achieved to assess the breathing individuality; each hypothesis is represented by a color and a number. Orange (1): The effect of moderate exercise while remaining quiet on within-speaker consistency of the breathing profiles; green (2): within-speaker consistency of breathing profiles between nonspeech and speech breathing; blue (3): within-speaker consistency of breathing profiles in spontaneous speech; and yellow (4): within-speaker consistency of breathing profiles in spontaneous speech over time; white: comparisons we did not investigate. This figure has been published in Serré et al. (2021)

of same-color boxes in Figure 5.6). The method developed by Benchetrit et al. Benchetrit et al., 1989 was used to assess the stability of speakers' breathing profiles in different conditions relative to the following aims:

First, to examine whether we could reproduce Eisele's et al. (1992) results on the inconsistency of respiratory patterns between rest and exercise, we compared three quiet conditions on day 1 (three comparisons; in orange (1) in 5.6). Second, to investigate the individual consistency of the breathing cycles between nonspeech and speech breathing, we compared quiet and speaking conditions on day 1 (three comparisons; in green (2) in 5.6). Third, to investigate the stability of breathing profiles in spontaneous speech, we compared the four speaking conditions between each other on days 1 and 2 (12 comparisons; in blue (3) in 5.6). Fourth, to investigate the stability of breathing profiles in spontaneous speech over time, we compared each speech condition between days 1 and 2 (four comparisons; in yellow (4) in 5.6).

5.3.2 Results on the individual consistency of the breathing cycles

The results on within-participant consistency for the different parameters across conditions and days are presented below. Consistency refers to the fact that the breathing cycle parameter values for a given speaker in a given condition are closer to those of the same speaker in another condition (individual cross-condition consistency) or the same condition on another day (individual cross-temporal consistency) than to those of other speakers in the same or different conditions.

5.3.2.1 Individual cross-condition consistency in quiet and quiet versus speech conditions on day 1

Table 5.8: Probability of T_w to pertain to the distribution of T_{btw} for each comparison between two quiet conditions, and between one quiet and one speech condition, within day 1. This table has been published in Serré et al. (2021).

		cycleDur			cycleSym			cycleShape		
conditions compared		T_α	T_w	P_t	T_α	T_w	P_t	T_α	T_w	P_t
Day 1 quiet	quiet_legMot-quiet_armMot	921	745	<0.003	937	998	0.021	939	1263	1
	quiet_noMot-quiet_legMot	893	1285	1	947	1127	0.26	935	1234	1
	quiet_noMot-quiet_armMot	995	1177	0.48	925	1168	0.39	880	1457	1
Day 1 quiet versus speech	quiet_noMot-speech_armBlock	932	1406	1	982	1838	1	906	1356	1
	quiet_legMot-speech_legMot	918	1357	1	917	1867	1	857	1221	1
	quiet_armMot-speech_armMot	932	1171	0.54	908	1862	1	931	1402	1

None of the comparisons between quiet conditions for *cycleDur*, *cycleSym*, and *cycleShape* (orange (1) comparisons in 5.6) were significantly consistent, except for breathing cycle duration between the *quiet_legMot* and *quiet_armMot* conditions ($p_t<0.003$, see Table 5.8). None of the comparisons between quiet and speech conditions (green (2) comparisons in 5.6) were significantly consistent ($p_t \geq 0.54$, see Table 5.8). Individual cross-condition consistency was thus not observed on day 1 between quiet conditions and between quiet and speech conditions.

5.3.2.2 Individual cross-condition consistency in speech conditions on days 1 and 2

Table 5.9: Probability of T_w to pertain to the distribution of T_{btw} for each comparison between two speech conditions, respectively, within days 1 and 2. This table has been published in Serré et al. (2021).

		cycleDur			cycleSym			cycleShape		
conditions compared		T_α	T_w	P_t	T_α	T_w	P_t	T_α	T_w	P_t
Day 1 speech	speech_armFree-speech_armBlock	918	780	<0.012	971	515	0.012	950	660	<0.012
	speech_armFree-speech_legMot	955	815	<0.012	890	638	<0.012	924	923	0.012
	speech_armFree-speech_armMot	881	837	<0.012	920	580	<0.012	917	873	<0.012
	speech_armBlock-speech_legMot	911	796	<0.012	979	613	<0.012	920	1085	<0.456
	speech_armBlock-speech_armMot	970	888	<0.012	947	647	<0.012	923	900	<0.012
	speech_armBlock-speech_legMot	934	779	<0.012	920	636	<0.012	962	845	<0.012
Day 2 speech	speech_armFree-speech_armBlock	769	623	<0.012	812	390	<0.012	774	653	<0.012
	speech_armFree-speech_legMot	745	735	0.012	778	479	<0.012	769	626	<0.012
	speech_armFree-speech_armMot	744	725	0.012	799	570	<0.012	726	745	<0.036
	speech_armBlock-speech_legMot	722	745	0.036	751	420	<0.012	785	722	<0.012
	speech_armBlock-speech_armMot	736	756	0.048	764	445	<0.012	749	735	<0.012
	speech_armBlock-speech_legMot	762	656	<0.012	761	322	<0.012	742	528	<0.012

On both days, within-speaker distances were ranked significantly lower than random association distances for all comparisons (blue (3) comparisons on Figure 5.6), for *cycleDur*, *cycleSym*, and *cycleShape* ($p_t \leq 0.048$, cf. Table 5.9). There was, however, a single exception: for *cycleShape*, the difference between *speech_ armBlock* and *speech_ legMot* was not significant (p_t=0.456).

5.3.2.3 Individual cross-day consistency of speech breathing profiles

Between days, for speech conditions (yellow (4) comparisons in Figure 5.6), participants were generally closer to themselves than to other participants (cf. Table 5.10) in terms of cycle duration (p_t<0.04), symmetry (p_t<0.004), and shape (p_t<0.04), except for the *armFree* condition, for which *cycleDur* is not consistent between days (p_t=0.08).

Table 5.10: Probability of T_w to pertain to the distribution of T_{btw} , for each comparison from day 1 to day 2, respectively, within each speech condition. This table has been published in Serré et al. (2021).

conditions compared from day 1 to day 2		cycleDur			cycleSym			cycleShape		
		T_α	T_w	P_t	T_α	T_w	P_t	T_α	T_w	P_t
Speech	speech_armFree	725	893	0.08	762	621	<0.004	749	818	0.04
	speech_armBlock	778	581	<0.004	771	554	<0.004	796	733	<0.004
	speech_legMot	756	838	0.04	786	460	<0.004	742	761	0.012
	speech_armMot	775	751	0.004	723	545	<0.004	797	742	0.004

The respiratory cycle and the breathing group change on different days and, to some extent, under certain experimental conditions. Speaker specificity of the speech breathing cycles is maintained across different limb movements and over time. We are now interested in the coordination between breathing and pedaling.

5.4 How does speech impact motor respiratory coupling?

As seen in the literature (see Chapter 2 section 2.3.2), breathing and motor activity show some pattern of synchronization, especially during physical activity, but also for movements such as ocular saccades or bi-manual coordination. Motor respiratory coupling is usually studied by varying the frequency of the motor or the breathing activity to measure changes in synchronization patterns. For this purpose, the tempo of the activity of reference is constrained. In the present experiment, neither biking nor breathing are constrained, and participants spontaneously chose their own pace. The synchronisation between biking and breathing was examined in this spontaneous context, depending on the limbs at work, and the effect of speaking on this coordination was assessed.

Because of their different frequencies (the movement of the limbs are faster than breathing), when they are in synchronization, breathing and biking signals adopt a multifrequency coordination. The mode of synchronization between two oscillators refers to the ratio m:n in which the respective frequencies w and z of the oscillators lock in the interaction such that nw=mz (Zelic et al., 2018). The stability of the synchronization is related to the duration in which the oscillators stays within the same mode, with a constant relative phase. Since neither the breathing nor the biking rhythms were constrained by a tempo, and since their potential coordination is not intentional, the breathing and biking time-series are non stationary and their frequency can vary over time (Hill et al., 1988). Their potential coupling is likely to be relative: their mode of synchronization may switch along the session. To visualize the data of these two irregular oscillators, and to investigate their motor respiratory coupling (MRC), two methods appropriate for the spontaneous and unconstrained nature of this coupling were applied. The first method is used to visualize the data. The second method gives an index of stability of the coupling.

5.4.1 Visualization of the data through the synchrogram

To enable a better visualization of the synchronization patterns between biking and breathing, we applied the method of Schäfer and colleagues (1998), who investigated the synchronization between heartbeat and ventilation. This method allows for visualizing periods of synchronization between irregular and weakly coupled oscillators. For each participant, the respiratory phase was stroboscopically observed: the values of the phase of the breathing signal were plotted for each time onset of the biking cycles. This representation is called a synchrogram (Schäfer et al., 1998). Below the synchrogram, the number of complete biking cycles within each breathing cycle was plotted. The third plot is a histogram of the phase values of the breathing cycle at each biking cycle onset. These four plots are displayed in Figures 5.8, 5.9, 5.10 and 5.11 for the same participant. If breathing and cycling signals are in synchronization, the following patterns should appear: on the synchrogram, the same phase values should occur for each group of n breathing cycles, depending on the mode of synchronization $\frac{n}{m}$. These phase values correspond to the time onset of the biking cycles. Graphically, the plots should present horizontal lines at the level of the phase values. The histogram of phase values should present some peaks corresponding to the level of the horizontal lines in the synchrogram. Plotting the number of biking cycles for each breathing cycle enables to see where the mode of synchronization switches.

To plot the synchrogram, the phase of the breathing signal and the mode of synchronisation have to be computed first. Because the time series were continuous and non-stationary, the instantaneous phase of the respiratory signal was first computed using the method of Varlet and collaborators (2011):

$$\phi(t) = arctan(\frac{x'(t)}{x(t)}) \tag{5.1}$$

where $x'(t)$ is the velocity of the oscillator and $x(t)$ is its position. The velocity was normalized by $\frac{2\pi}{hp}$, where hp is half the period in which t occurs:

$$\phi(t) = arctan(\frac{\frac{x'(t)*hp}{2\pi}}{x(t)}) \tag{5.2}$$

The main improvement of this method is the computation of the phase on half of each cycle, and the normalization of its velocity by half the period of a cycle. Computing the phase on each half of the cycles instead of the whole cycles enables to better consider the irregularity of the signal and be more accurate.

However, the respiratory signal presents some artefacts due to limb movements, especially when biking with the arms (see Figure 5.7). To remove these artefacts, the signals were filtered through a Kaiser window with an amplitude of 200 and a threshold of 2 Hz (see Figure 5.7). The phase still presented problematic irregularities even with this filtering (see Figure 5.7). To overcome this issue, it has been decided to compute the phase in a discrete way. The method applied was the one of Zelic and collaborators (2018).

For one participant, the synchrograms in quiet biking conditions are displayed in figures 5.8 and 5.10. To compare the patterns of these synchrograms to the synchrograms of signals that have no reason to be entrained by each other (because they did not occur at the same time), the synchrograms of the synchronization between biking signals and breathing in *quiet_ noMot* condition are displayed in figures 5.9 and 5.11.

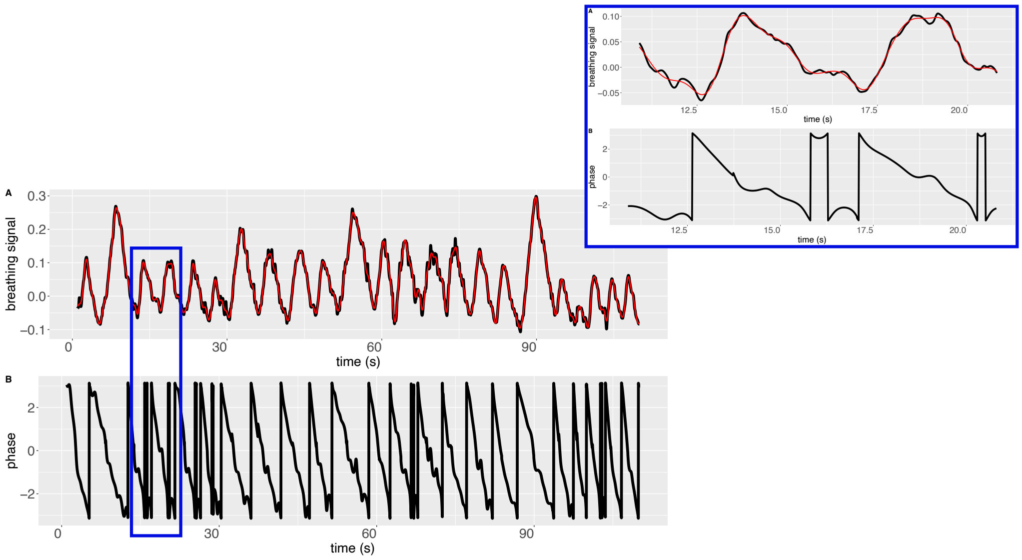


Figure 5.7: Breathing signal and continuous phase of one participant in the condition *quiet_armMot*. A : raw (black) and filtered (red) breathing signals. B : phase of the breathing signal computed using the method of Varlet et al. (2011). Rectangle: zoom in on the signalled part.

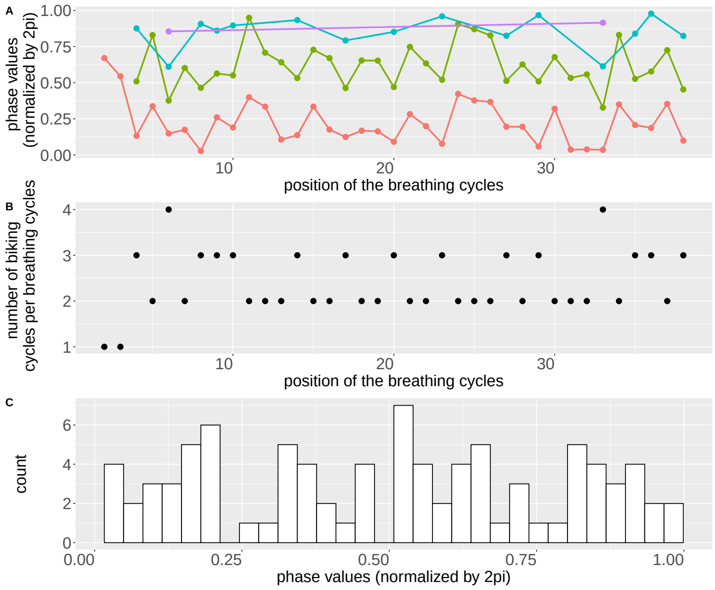


Figure 5.8: Synchronization between a breathing signal and the corresponding movement signal when quiet and biking with the legs, for one participant. A: synchrogram. Each color represents the order of a biking cycle within the breathing cycle. First biking cycles are in pink. B: number of complete biking cycles per breathing cycle. C: histogram of the phase.

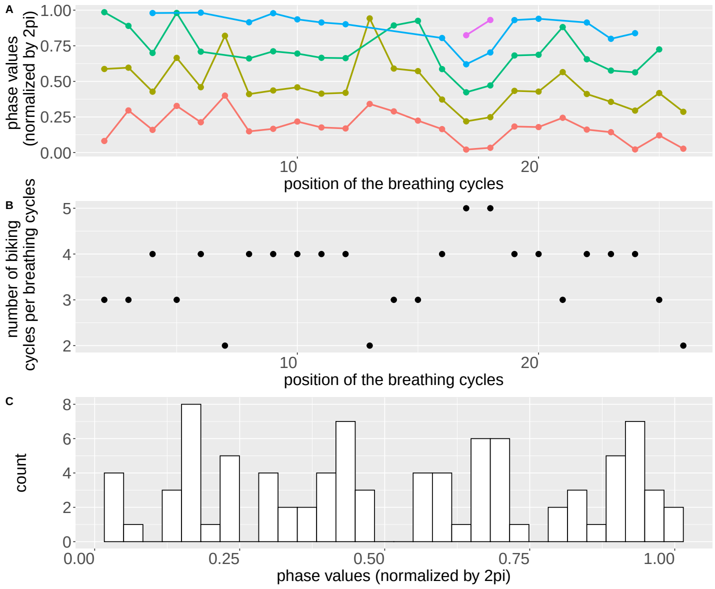


Figure 5.9: Synchronization between a breathing signal when quiet doing nothing, and the movement signal when quiet and biking with the legs, for one participant. A: synchrogram. Each color represents the order of a biking cycle within the breathing cycle. First biking cycles are in pink. B: number of complete biking cycles per breathing cycle. C: histogram of the phase.

Figure 5.8 shows weak synchronization patterns between breathing and legs biking: from one breathing cycle to the following, the phase of the breathing signal at biking cycle time onsets fluctuates (on part A, none of the coloured curves are constant) and the number of complete cycles per breathing cycle varies a lot (part B). Comparing figures 5.8 and 5.9, (Figure 5.9 illustrates the synchronization between two signals that are not coupled), enables to see that for this participant, breathing and legs biking when quiet do not present any remarkable consistency in synchronization patterns.

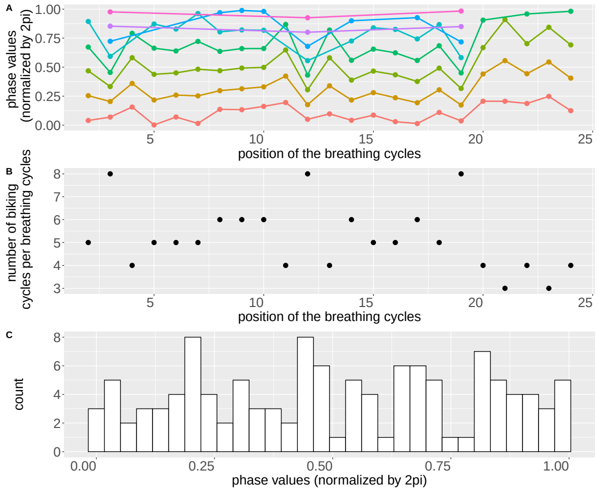


Figure 5.10: Synchronization between a breathing signal and the corresponding movement signal when quiet and biking with the arms, for one participant. A: synchrogram. Each color represents the order of a biking cycle within the breathing cycle. First biking cycles are in pink. B: number of complete biking cycles per breathing cycle. C: histogram of the phase.

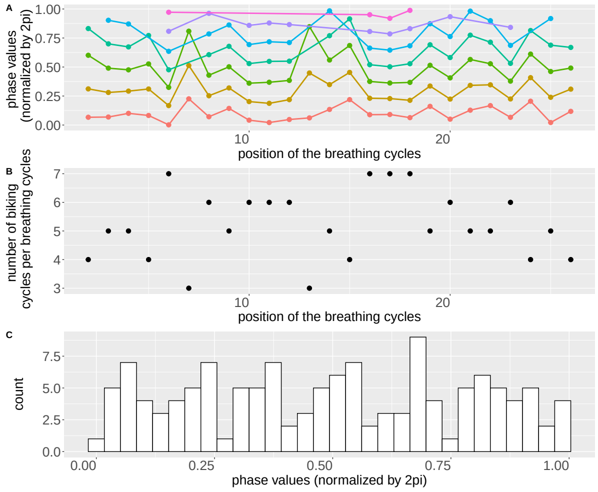


Figure 5.11: Synchronization between a breathing signal when quiet doing nothing, and the movement signal when quiet and biking with the arms, for one participant. A: synchrogram. Each color represents the order of a biking cycle within the breathing cycle. First biking cycles are in pink. B: number of complete biking cycles per breathing cycle. C: histogram of the phase.

Biking with the arms (see Figure 5.10) displays stronger synchronization patterns compared to biking with the legs, but not compared to the corresponding surrogate association (see 5.11). Both display periods in which the phase of movement cycle onsets is constant over several breathing cycles.

To quantify the synchronization between breathing and biking signals, we followed the procedure of Bouvet and colleagues (2019), computing the index of stability of synchronization (Zelic et al., 2018).

5.4.2 Analysis of the index of coordination stability

5.4.2.1 What is the index of stability?

The mode and stability of the coordination between breathing and limb movements were assessed by computing the index of stability developed by Zelic and colleagues (2018). This method enables the evaluation of the synchronization between two oscillators spontaneously and bidirectionally coordinated (without knowing which one entrains the other), which present multiple modes of coordination.

The index stability (IS) analysis is based on the circle map model. The circle map model predicts the attraction region of rational ratios in which the mode (frequency coupling) of coupled oscillators can fall. The stability of the attraction depends on the fluctuation of the relative phase and the span of the attraction region of the mode (Treffner and Turvey, 1993; Kelso, 1995; Pikovsky et al., 2003). Modes with small integers such as 1:1 or 1:2 have wider attraction regions. They are more stable than modes with higher integers. This has been empirically observed in human MRC (Bramble and Carrier, 1983; Treffner and Turvey, 1993; Daffertshofer et al., 2004). The Farey Tree describes the predictions of the circle map model, showing the paths from less stable to more stable ratios (see Figure 5.12). It enables to predict whether the oscillators will switch from one mode to another, depending on their coupling strength. The Farey tree has several levels. The first level includes modes $0/1$ and $1/1$. Level 2 includes the modes of level one, with $1/2$. Level three includes the modes of level 2, with $1/3$ and $2/3$, and so on.

First, the modes of the frequency coupling of the two signals at the time onsets of each breathing and biking cycle are computed. This vector of frequency ratios is then transformed in rational ratios of the Farey tree: for each level of a 10 level Farey tree (see figure 5.12 part B), the closest rational ratio of the frequency quotient is attributed to the corresponding time onset. We end up with ten vectors of rational ratios (still corresponding to the time onsets). For each time onset in each level, the relative phase between the two signals is computed, depending on the actual rational ratio. This results in ten vectors of relative phase (still corresponding to the time onsets). For each vector (corresponding to each level of the Farey Tree), the index of stability (IS) is computed, based on the

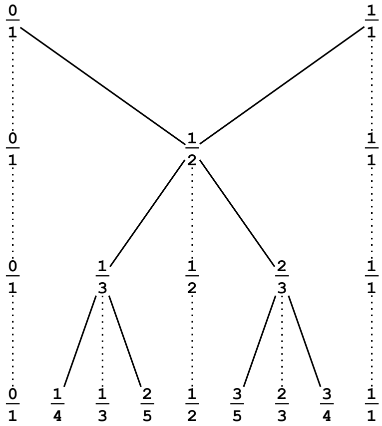


Figure 5.12: Illustration of the Farey tree with 4 levels (picture adapted from wikimedia). The paths appearing between the levels describe 'how coupled oscillators switch from less stable ratios at higher levels to more stable ratios of coordination at lower levels' (Bouvet et al., 2019, p.34).

mean resultant length of the relative phase time serie. The index of stability is defined between 0 and 1. *IS* is equal to 0 if there is no synchronization, and 1 if there is a perfect synchronization. The level which has the maximum *IS* is considered as the dominant level of synchronization, and the most frequent mode within this level is considered as the dominant mode of synchronization of the breathing and biking signals. See Appendix E for more details.

To assess whether there is an effective synchronization between breathing and biking, the index of stability is compared to the index of stability between the biking signals and the breathing signal from conditions with no biking. This association is the control condition. This comparison enables to test if the synchronization patterns are due to the fact that the participant keeps a stable breathing pace over time even when not biking, or due to chance, and not to the entrainment of breathing by biking movements. There are four control conditions in total. Two are quiet control conditions, in which the quiet biking signals (legs and arms) are associated with breathing during the quiet trial with no movement, and two speech control conditions, in which the speech biking signals are associated with breathing during the *speech_ armBlock* breathing.

5.4.2.2 Computing the vectors of time-onset

The computation of the index of stability requires the vectors of the time onsets of the biking and breathing cycles. The inhalation onsets constituted the breathing time series. The pedal strikes constituted the movement time series. The pedal strikes were computed by automatically annotating the acceleration peaks of the movement signals, after down-

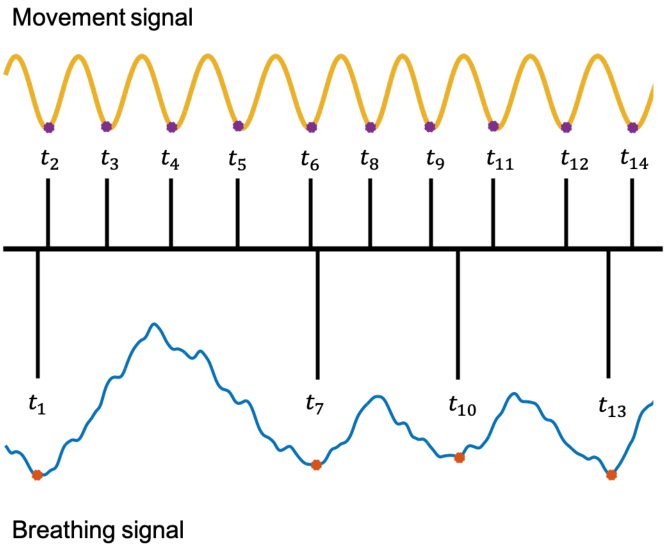


Figure 5.13: Schema of the vector of time onsets used to compute the index of stability. The times of inhalation onsets and the times of acceleration peaks are intertwined.

sampling the signals to 100 Hz to fit with the sample rate of breathing (see Chapter 3 Sections 3.6.2.2 and 3.6.3.2). As displayed in Figure 5.13, the time of inhalation onsets and acceleration peaks are intertwined in one vector.

5.4.2.3 Statistical analyses

Generalized Linear mixed models were used to assess the effects of control condition, limb motion, speech, and day on the index of stability (IS). A random intercept per participant was included. A backward model selection was used to select significant factors (see Chapter 3 section 3.7 for more details). Since IS is a ratio, we applied a generalized linear mixed model with a logit link function.

The first model was applied to the data from day 1. Its aim was to examine the effect of biking motion on breathing entrainment (control vs biking condition), and the effect of limb motion on the stability index in quiet and speech conditions. A three factor model was implemented: *controlCondition (control; biking)*, *vocalCondition (speech; quiet)*, and *limbCondition (leg; arm)*.

After the statistical steps described in Chapter 3 Section 3.7, the final model is:

$$IS \sim controlCondition * limbCondition * vocalCondition + (1|participant)$$

The second model was applied to the data from speech conditions only. Its aim was to examine the effect of biking motion on breathing entrainment *(control vs biking condition)*, and the effect of *limbCondition (armFree; armBlock; legMot; armMot)* and *day (day1; day2)* on the index of stability.

After the statistical steps described in Chapter 3 Section 3.7, the final model is:

$IS \sim (1 + limbCondition|participant)$

5.4.2.4 Results

A summary of the distribution of the index of stability is presented in Table 5.11.

Model 1 The selected statistical model included the interactions between *controlCondition*, *vocalCondition* and*limbCondition*. As displayed in Figure 5.14 and Table 5.12, in the *quiet* condition, biking *IS* is higher than control *IS* (*biking_ arm_ quiet* - *control_ arm_ quiet*: b=0.6, z = 10.2, p < 0.001; *biking_ leg_ quiet* - *control_ leg_ quiet*: b=0.3, z = 4.4, p < 0.001). In the *biking_ quiet* condition, biking with the arms increases *IS* compared to biking with the legs (*biking_ arm_ quiet* - *biking_ leg_ quiet*: b = 0.3, z = 5.5, p < 0.001). In the *biking* condition, being quiet increases *IS* compared to speaking (*biking_ arm_ quiet* - *biking_ arm_ speech*: b = 0.7, z = 11.3, p < 0.001; *biking_ leg_ quiet* - *biking_ leg_ speech*: b = 0.3, z = 4.9, p < 0.001).

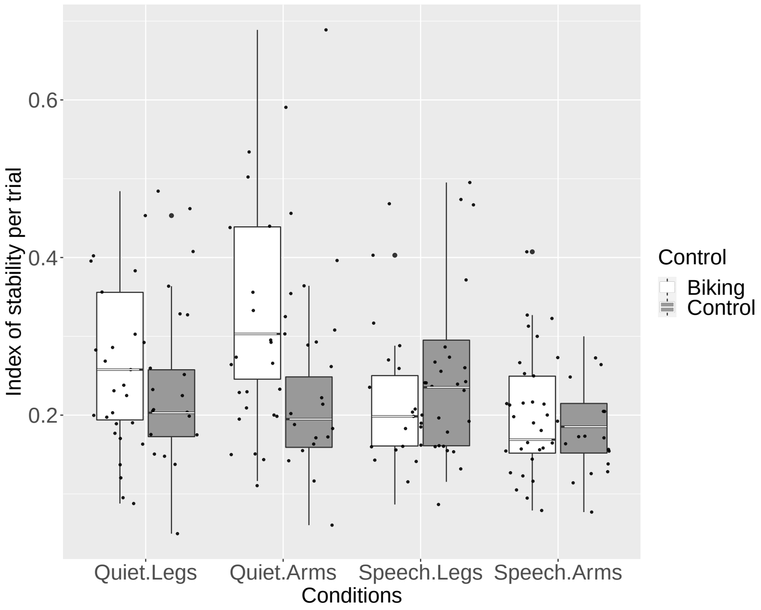


Figure 5.14: Index of stability (*IS*) on day 1, per control, vocal and limb conditions.

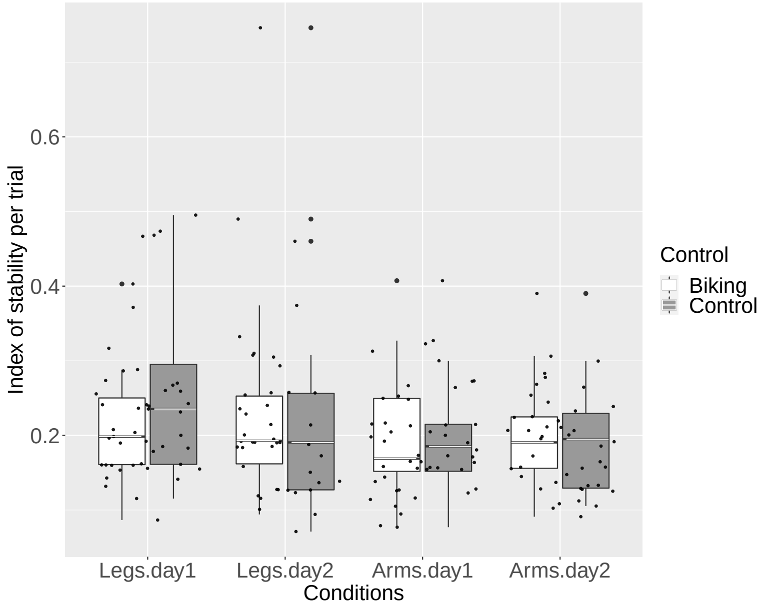


Figure 5.15: Index of stability (*IS*) in speech conditions, per control conditions and days.

Model 2 No factor remained in the selected statistical model. Figure 5.15 displays the results.

Table 5.11: Summary of the distribution of the index of stability, depending on the day and the condition of each signal implied in the coupling. *Condition mvt* refers to the condition of the limb movement signal, and *condition breath* refers to the condition of the breathing signal.

day	condition mvt	condition breath	min	mean ± sd	max
1	quiet_legMot	quiet_legMot	0.09	0.27 ± 0.11	0.48
1	quiet_legMot	quiet_NoMvt	0.05	0.22 ± 0.09	0.45
1	quiet_armMot	quiet_armMot	0.12	0.34 ± 0.15	0.69
1	quiet_armMot	quiet_NoMvt	0.06	0.21 ± 0.07	0.36
1	speech_armMot	speech_armMot	0.08	0.2 ± 0.08	0.41
1	speech_armMot	speech_armBlock	0.09	0.2 ± 0.06	0.31
1	speech_legMot	speech_legMot	0.09	0.21 ± 0.07	0.4
1	speech_legMot	speech_armBlock	0.1	0.25 ± 0.1	0.47
2	speech_armMot	speech_armMot	0.09	0.19 ± 0.06	0.31
2	speech_armMot	speech_armBlock	0.07	0.18 ± 0.08	0.39
2	speech_legMot	speech_legMot	0.09	0.21 ± 0.07	0.37
2	speech_legMot2	speech_armBlock	0.11	0.22 ± 0.08	0.38

Table 5.12: Post-hoc comparisons from the linear mixed model 1 applied on the index of stability. Only the significant contrasts are displayed.

	contrast	Estimate	SE	z-value	p-value
Model 1	biking arm quiet - control arm quiet	0.61316	0.06004	10.212	<0.001
	biking leg quiet - control leg quiet	0.27184	0.06113	4.447	<0.001
	biking arm speech - control arm speech	0.08485	0.06255	1.357	0.68
	biking leg speech - control leg speech	-0.25647	0.06142	-4.176	<0.001
	biking arm quiet - biking leg quiet	0.31963	0.05850	5.464	<0.001
	biking arm speech - biking leg speech	-0.05382	0.06240	-0.863	0.94
	biking arm quiet - biking arm speech	0.67644	0.06005	11.264	<0.001
	biking leg quiet - biking leg speech	0.30298	0.06128	4.944	<0.001

Table 5.13 and Figure 5.16 summarize and display the average number of biking cycles per breathing cycle, depending on the day and the condition. In average, speaking seems to increase the number of biking cycles per breathing cycle compared to being quiet. Most frequent modes of synchronization are 1:3 and 1:4 in quiet conditions, and 1:4, 1:5 and 1:6 in speech conditions.

Table 5.13: Summary of the distribution of the number of biking cycles per breathing cycle, depending on the day and the condition.

day	condition	min	mean ± sd	max
1	quiet_legMot	2.23	3.82 ± 2	9.83
1	quiet_armMot	1.78	3.7 ± 1.42	7.38
1	speech_armMot	2.47	4.73 ± 1.52	8.57
1	speech_legMot	2.88	5.02 ± 1.69	9.64
2	speech_armMot	2.09	4.72 ± 1.12	6.39
2	speech_legMot	3.09	5.37 ± 1.46	7.64

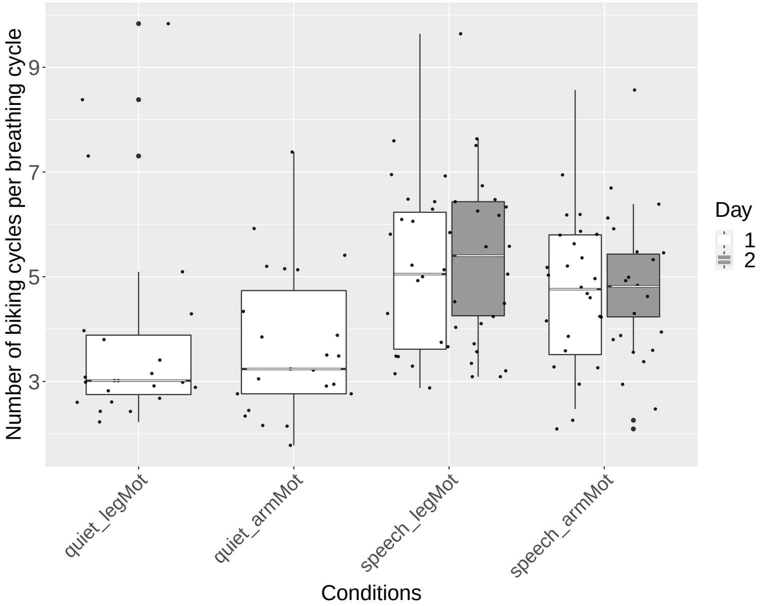


Figure 5.16: Number of biking cycles per breath cycles, averaged on trials.

These results suggest that motor respiratory coupling is only present in the *quiet* conditions (both for biking with the arms and biking with the legs), and that the stability increases with the arms compared to the legs.

5.5 Overall discussion

The goal of this chapter was to investigate breathing during the occurrence of speech and limb movements. The results on the speech parameters at the breath group scale, on the speech breathing individuality, and on the motor respiratory coupling, will be shortly discussed.

5.5.1 Speech evolution in breath groups may be related to narration performances

The movement conditions do not have any effect on the organization of speech within a breathing cycle (in terms of number of syllables and speech rate). Neither the day nor the limb condition impact the number of breathing cycles per trial. The cycle duration increases form day 1 to day 2, and the cycle symmetry decreases with time. Together with the increase of the number of syllables and the proportion of speech time with the day, these results suggest that participants speak more fluently within breathing cycles over time. Participants also speak faster between days 2 and 3 on the long-term recalls. The changes in speech parameters over time seem to be markers of learning, and impact the breathing cycle: the breathing cycle is less symmetric over time, because of longer exhalations due to more syllables.

5.5.2 Motor respiratory coordination is impacted by speech

Breathing seems to be entrained by biking cycles in quiet conditions. This is suggested by the increase of the stability index compared to control associations. Biking and breathing do not seem to present any synchronisation in speech conditions, since there is no significant differences between biking and control conditions when speaking. MRC in speech may be removed by the constraints that speech puts on breathing, drastically changing its frequency. Breathing can no longer be entrained by cycling. Limb cycling movements do not synchronize to speech breathing. This is confirmed by the fact that biking cycle durations are more variable during speech than quiet conditions (see Chapter 4 Section 4.2). Limb movements may not adapt to the new breathing frequency because of their role of leading oscillator. However, it has been shown that cycling can be entrained by breathing (Hoffmann and Bardy, 2015). The variability of breathing cycle durations, which are shaped by the upcoming utterance (Winkworth et al., 1994), may prevent the emergence of MRC. Neural commands from higher cortical areas to meet speech breathing needs may bypass central pattern generators responsible for MRC. Finally, cycling motion may be entrained by speech itself, interfering with co-speech gesture inhibition.

The index of stability is higher when quietly biking with the arms compared to quietly biking with the legs, suggesting a stronger coupling between arm movements and breathing than between leg movements and breathing. The nature of the task may not involve strong biomechanical interactions between the lungs and the legs. However, leg cycling impacts speech intensity whereas arm cycling does not (see Chapter 4 section 4.3.2). Besides, whereas biking conditions shorten the exhalations in quiet conditions, only biking with the legs has this effect in speech conditions. The synchronisation between breathing and cycling (both with arms and legs) has been previously assessed (Sporer et al., 2007), at different levels of intensity. The authors found a tendency of the entrainment to be higher during leg than arm exercise, suggesting that synchronisation is stronger in familiar tasks, such as biking with the legs, compared to unfamiliar ones, such as biking with the arms. Besides, MRC seems to increase with expertise. The results in the present study are not in line with these findings. MRC may emerge from other mechanisms than the ones related to physical exertion, and may be supported by higher cortical areas, such as sub-thalamic area, which sustains breathing when automatic mechanisms of the brain stem no longer work (Paterson et al., 1986; Shea, 1996). The difference of MRC stability between arm and leg biking may also lie in the different types of fibers recruited in the muscles (Shevtsova et al., 2019). Slow-twitch (type 1) fibers are more resistant to fatigue than fast-twitch (type 2) fibers. Stimulation of type 2 fibers likely triggers a higher and faster increase in ventilation than type 1 fibers. Upper limb muscles have a higher proportion of type 2 fibers than lower limb muscles. This physiological account may contribute to a higher MRC with upper limbs than lower limbs. More generally, the discrepancies in the results of studies on MRC may be explained by the differences in the tasks performed by the participants. Due to the different tasks, not the same muscles are recruited each time. The response of the respiratory system may change depending on the fiber composition of the muscles solicited.

Our results on MRC coupling suggest that MRC does emerge from spontaneous breathing and cycling, without any imposed rhythm, even in situations that does not require a lot of effort (Hill et al., 1988). These results are preliminary, and more analyses are needed to investigate further the patterns of this synchronization. The modes of coordination and their occurrence over a trial need to be examined for each participant, as well as the percentage of time during which breathing and cycling stay synchronized over the trial.

5.5.3 Speech breathing is speaker-specific

Individual consistency of breathing profiles was found over days even when speech is produced at the same time as light physical activity with the arms or legs. Our analyses also replicated the inconsistency in individual breathing profiles previously reported between breathing at rest and breathing during physical exercise (Eisele et al., 1992). In general, changes in breathing profiles related to leg motion did not affect the individual consistency of cycle duration, symmetry, and shape between speech conditions. This suggests that speech breathing individuality is maintained despite the co-occurrent motion of the arms or legs, at least when these are cyclic and do not strongly impact ventilation. The level of exercise to which participants were committed was, indeed, light. Regarding previous work (Doust and Patrick, 1981; Rotstein et al., 2004; Baker et al., 2008), it is well known that physical effort impacts speech breathing; what is unknown is whether this impact also affects the individuality of breathing and/or if speakers' strategies to compensate for physical effort are consistent over time. It would be interesting to examine if speech breathing patterns are maintained over days during intense physical activity. Assessing the consistency of speech breathing over different bodily contexts is a way to address the specificity of speech breathing skills. In the previous work, this bodily context was modulated by controlling the degree of effort. In the current work, we used leg versus arm movements. We found individual consistency of the breathing profiles from one day to another for all comparisons, and changes in cycle duration and symmetry between days 1 and 2 did not affect this consistency—except for cycle duration in *speech_armFree*. These results are in line with the individual consistency of tidal breathing profiles from one day to another (Shea et al., 1987). It would be interesting to examine whether, as for tidal breathing, we find this consistency across years.

In the next chapter, the results of chapters 4 and 5 will be discussed with respect to the state of the art detailed in chapters 1 and 2.

Chapter 6

Discussion and perspectives

The aim of this thesis was to contribute to a better understanding of speech-breath-movement interactions through a cross-disciplinary approach combining knowledge in phonetics, linguistics, neurophysiology and motor control. This thesis is an attempt to link these fields methodologically and theoretically, and to apply methods from motion analysis to understand a complex tripartite interaction.

Speech, breathing and limb movements have been extensively studied in pairs, but rarely together. However, arguments in favor of a complex interaction between these three actors are pouring in, and tend to show that breathing has a key role (cf. chapter 1). Breathing can drive cognition and attention processes (Allen et al., 2022), and can itself be driven by limb movements during physical activity (Bramble and Carrier, 1983) or even simply by movements such as eye or manual movements (Temprado et al., 2002; Rassler and Raabe, 2003). Breathing is highly influenced by behavioral factors related to the motor, emotional, physiological and cognitive processes at work (Shea, 1996). The breathing airflow is shaped by these processes. In this respect, the breathing profiles are markers of ongoing activities of the body and mind. Finally, breathing is a resource shared between speech and limb movements (Shea, 1996), and is intrinsic to both (Fuchs and Rochet-Capellan, 2021). It adapts to speech via utterance planning (Winkworth et al., 1994), and to limb movements via afferent connections and chemoreceptors that transmit metabolic needs (Shevtsova et al., 2019). Breathing can therefore reflect and influence speech-limb movement interactions.

Speech-limb movement interactions were studied in very distinct theoretical areas that do not necessarily communicate. On the one hand, the positive effects of coverbal gestures on speech has been extensively studied (Wagner et al., 2014), sometimes compared to simple rhythmic movements such as hand waving (Goldin-Meadow et al., 2009). On the other hand, studies with a dual-task paradigm show a deleterious effect (interference) of speech on limb movements (Verghese et al., 2007; Dayalu et al., 2013;), and a more debated effect of movement on speech, depending on the speech variables measured, and the type of the motor task (Kemper et al., 2003; Verghese et al., 2007; Dayalu et al., 2013). A third field, related to nonlinear dynamic system theories, studies the synchronization between simple rhythmic movements and the production of vocalizations, or even sentences

(Scott Kelso et al., 1983; Treffner and Peter, 2002; Parrell et al., 2014; Zelic et al., 2015). Synchronization patterns emerge from the production of vocalizations simultaneously with a rhythmic movement. However, it is difficult to determine whether this synchronization between the two emerges from low-level motor mechanisms, or from cognitive processes related to spoken communication.

This thesis links methods and theories of these different disciplinary fields in order to better understand the functioning of the body substrate in spoken communication. In this thesis, we have confronted communicative and non-communicative movements in a narrative situation in order to compare their interactions with speech over time. Each component (breathing, speech and movement) was recorded in order to be analyzed in regards to the different speaking and limb contexts. The methods employed to analyze the data come from the fields of psychology, physiology, phonetics and motor control.

Through this work, the following questions were addressed: (1) To what extent do communicative and non-communicative movements have a different impact on speech? (2) How does natural speech impact non-communicative movements? (3) How does speech impact motor respiratory coupling? (4) Is speech breathing speaker-specific, and is this speaker specificity maintained across different limb movements? (5) Do the interactions mentioned in the previous questions change over time? These questions have been addressed on the basis of a dataset recorded with broader aims, such as the assessment of the effect of the different limb movements on memorisation.

6.1 Interactions between speech, breathing and limb movements over time

In this section, the results of chapters 4 and 5 are crossed and discussed together: (1) the results about the effects on limb movements and time on speech and speech breathing; (2) the results about the effect of speech on limb movements, and on respiratory-motor coupling.

6.1.1 Effect of limb movements and time on speech and speech breathing

This section partly reports the introduction of our paper published in the proceedings of ISSP (Serré et al., 2020).

None of the movements involved had more effect on speech fluency or recall performance than the others, both at the trial and breathing cycle scale. In the arm free condition, participants did not gesture a lot, which could explain the lack of effect of gesturing on speech. It is also possible that participants compensate with other body parts, such as the head. As seen in Chapter 2, Hoetjes and collaborators (2014) did not find any effect of not being able to gesture on speech. They suggest that although people are sitting on their hands (in the no gesture conditions), they still can make tiny movements, or send motor commands that are not fully realized. These movements would stand for the gestures they can produce when allowed to. Besides, co-verbal gestures are not limited to the hands (Kendon, 2004). For instance, people also move their head and eyebrows when speaking. These movements were also shown to possibly synchronize with speech prosodic parameters (Hadar et al., 1983; Graf et al., 2002). Finlayson et al. (2003) suggested that participants could compensate for not being able to move their hands by means of moving (or moving more than typically) other parts of their body. This compensation might avoid effects of the hands being blocked on speech production. For example, Rimé et al. (1984) observed an increase of eyebrow movements when speakers could not gesture with their hands. This could also be the case for other body parts such as the head. During this thesis, the head movements were analyzed to examine whether the head compensates for the hand not being able to move in the hands constrained condition. The results were presented at the 12th International Seminar on Speech Production (Serré et al., 2020). The results rather suggest an entrainment of the head by the hands when speaker are allowed to gesture. The head moves less in the hands constrained condition than in the hands free condition. Head motion was analyzed quantitatively, by quantifying its movement with the signal of the rigid body set on the head front (as hand movement in section 4.1 of chapter 3). Facial motion was not investigated.

The fact that cycling movements did not have any effect on speech fluency or content recall is in line with the recent literature on the effect of light physical activity during a cognitive task. For instance, in Germany in 2017, desk-based employees reported to sit 73.0% of their working hours (Wallmann-Sperlich et al., 2017). To prevent the employees from health problems due to bad sitting positions, more and more companies invest in material enabling their workers to keep moving while working, such as mini-bikes or treadmills. Studying light exercise effects on cognitive performances became crucial to ensure that work efficiency is not impacted. Most of the findings support the idea that light physical activity at work (like desk-biking) does not disturb concurrent cognitive tasks (Commissaris et al., 2014; Torbeyns et al., 2016; Ruiter et al., 2019).

In the literature, it has been suggested that simple rhythmic movements can improve attention processes (Morillon et al., 2014). Some studies (Schmidt-Kassow et al., 2014; Plancher et al., 2019) mention this possibility (either pace arm movements or paced walking) to explain better working memory performances. This was not the case in our experiment. However, participants were biking during recall, not while watching the stories.

It would be interesting to make participants pedal when encoding the stories. This could have helped retention of information, perhaps only if the flow of information was synchronized with the pedaling frequency (with the acceleration peaks). We could also test whether pedaling between encoding and recall allows a better maintenance of information in working memory (Plancher et al., 2019).

The effects of limb movements on speech may strongly depend on the type of the ongoing tasks. Wagner and colleagues (2004) showed that co-speech gestures, but not mere hand-waving, supported the memorization of information such as letters. On the other hand, Morillon and colleagues (2014) showed that a mere rhythmic movement enhanced attention processes. The type of body movements and the type of learning and executive functions involved in a cognitive task have to be considered. Not all body movements have positive effects on cognition, and not all cognitive tasks can benefit from a particular movement. As Shebani and Pulvermüller (2013; 2018) suggest, at a neural level, the facilitation or interference effect between action and semantics also depends on the complexity of the motor task. Shebani and Pulvermüller (2013) showed that working memory for arm and leg related action words was impaired by a complex motor task from the corresponding effector. In a more recent study (2018), they showed that a simple motor task such as finger tapping facilitated working memory for arm-related words.

The speech parameters analyzed in this thesis are related to speech fluency (such as the number of syllables, speech rate, and the proportion of speech time), content recall and speech acoustics (F0 and amplitude envelope). These parameters are not exhaustive, and do no describe all the dimensions of speech. Effect of motions might be more visible in the analysis of the trajectories and coordination of speech articulators such as the tongue and the lips. Previous studies found such coordination patterns between vocalization and pointing (Rochet-Capellan et al., 2008) or finger movements (Zelic et al., 2015). It could be interesting to extend these investigations to natural speech, and non-communicative limb movements such as cycling. However, recording and analyzing the movement of the articulators during spontaneous speech may be challenging.

6.1.2 Effect of speech on limb movements and motor-respiratory coupling over time

Speech reshapes both breathing and cycling movements:

1. Biking cycles were longer during speech compared to being quiet. Biking movements may be entrained by the cognitive activity linked to the narration task. The decreased variability of biking cycle duration from day 1 to day 2 can be linked to the improvement of speech fluency and content recall from one day to another. The cog-

nitive demand may be reduced with time, and the variability of co-occurent motor tasks would decrease as well. This reduction in duration variability may also be due to a learning effect of the task.

2. The motor respiratory coupling is no longer present when speaking. Breathing is entrained by both leg and arm cycling movements when quiet, but not while speaking. This result is also coherent with the fact that speech breathing stays individually consistent across different limb movement conditions, while the individual consistency of quiet breathing is not maintained across the different limb movement conditions.

Analysis of the effect of speech on cyclic movements and on movement-breathing coupling revealed an increased sensitivity of the upper limb motor system to speech compared to the lower limbs. When biking with the arms, speaking increased the variability of biking cycle duration. No difference in the variability of leg cycle duration between quiet and speaking was found. These results suggest a stronger effect of speech on biking with the arms than biking with the legs. This stronger effect is also present in the motor respiratory coupling: synchronisation between biking and breathing is present when quiet for both legs and arms, but significantly more stable with the arms. Pouw et al. (2020a) have recently introduced the idea that arm motion may constrain changes in speech production via the breathing system (Pouw et al., 2020d; Pouw et al., 2020b), which may affect vocal output. They defend the idea that this link is anchored in phylogenesis: the bipedal state freed the upper limbs and respiratory system from being synchronized with locomotion. In return, the upper limbs may have participated in the complexity of respiratory control in speech. This is a parsimonious explanation for the mutual influences between hand movement and phonation in communication as compared with a “pure” cognitive explanation (Pouw et al., 2020b). This phylogenetic account can be extended to the link between ventilation and upper limbs in general. The potential link between ventilation and upper limbs outside a communication context may come from our previous quadrupedal state, in which the breathing apparatus was heavily constrained by upper limb activity.

6.1.3 Why would speech breathing be speaker specific?

This section is based on the paper by Serré et al. (2021).

In chapter 5, we found that the intra-individual distances between speech breathing patterns of different conditions and days are smaller than inter-individual distances, suggesting the existence of a respiratory individuality. Speech breathing patterns are consistent across limb movements and over days. These results raised the question of why should speech breathing be consistent or not for the same speaker over bodily contexts and days? Different constraints related to speech production might shape breathing cycles.

First, the cycle shape, as defined in the current work, may be heavily affected by segmental and prosodic aspects of the produced speech material as well as speaker-specific voice quality due to different use of laryngeal (and supralaryngeal) configurations. The measured kinematic output in speech production is, therefore, heavily modulated by the laryngeal configuration and upper vocaltract, which can also compensate for the loss of air (Fuchs et al., 2013; Zhang, 2016; Aare et al., 2018). The control of the larynx and upper articulators may be speaker specific and could account for the consistency of the cycle shape. Second, speech breathing rhythm is in general irregular: cycle durations, and especially exhalation durations, are related to the length of the utterance produced. The current study focused on the analysis of the breathing cycles only. It is, however, possible that consistency in breathing profiles — in particular for the duration parameter — could be explained by the fact that people are consistent in the duration of their utterances (interpausal unit). Further analyses of our dataset suggest that utterance duration is, indeed, correlated to cycle duration (Pearson's correlation:$r=0.9$, $t=138.83$, $df=4689$, $p < 2.2e-16$) and to a lesser extent anti-correlated to cycle symmetry ($r=-0.65$, $t=57.9$, $df=4689$, $p < 2.2e-16$). We applied the method used to assess the breathing individuality on utterance duration. This analysis gave the same results as for cycle duration: on each day, within-speaker distances were ranked significantly lower than random association distances for all the comparisons ($Pt<0.012$), as well as for the comparisons between days ($Pt<0.004$) except for armFree ($Pt=0.16$).The question arises of how to rule out whether utterance length determined breathing cycle duration or if utterance length is determined by individual lung volume. The fact that breathing profiles during spontaneous speech are different than during reading, and in particular, the fact that the link between inhalation profile and utterance properties is weaker in the former than in the latter (Winkworth et al., 1995; Rochet-Capellan and Fuchs, 2013) suggests that two levels of control are involved: a global speech planning and an online adaptation. The contribution of these two types of controls (that echo to some extent forward versus feed-forward models of motor control) (Wolpert and Kawato, 1998; Haith and Krakauer, 2013) might be specific to the speaker and may depend at least on three types of elements: bodily-related constraints, cognitive and linguistic constraints, and communicative constraints.

First, speech breathing might be individually consistent due to body properties. The first factor considered in the literature to affect breathing is lung volume capacity that is variable among speakers and depends on the size of the body (Heldner et al., 2019; Manifold and Murdoch, 1993; Hoit and Hixon, 1986). In the course of child development, the evolution of utterance length was observed in relation to an increase in lung volume (Boucher and Lalonde, 2015). For adults, one can expect speakers with a larger lung volume, taller people, for example, to produce longer utterances. However, this was not observed so far: a positive correlation between VC (the maximal amount of air one can inhale) and utterance duration has not been confirmed in the literature (Heldner et al., 2019). In their investigation of the implication of the forebrain in the breathing individuality, Shea et al. (1990) reported different cycleshapes for the same body types and variable body types for similar cycle shapes.

Second, cognitive functioning and abilities, such as working memory, attentional capacity, but also specific linguistic experience, might also affect speech breathing. The hypothesis of a link between cognitive processing and breathing profiles was, indeed, introduced early. Golla and Antonovitch (1929) suggested that people who rely more on visual mental imagery display more regular breathing profiles at rest than people who rely more on auditory imagery. Grassmann et al. (2016) show that "mentally demanding episodes are clearly marked by faster breathing and higher minute ventilation" (p.1). More recently, in a comprehensive study, Park et al. (2020) investigated the "readiness potential," an ERP component consistently found before initiating voluntary actions.Their findings show that voluntary actions frequently occur during exhalation and that the amplitude of readiness potential depends on the respiratory phase,that is, actions are phase coupled to breathing. While they do not report individual differences, they refer to the literature on different participant groups, showing, for example, differences among athletes and non-athletes. Finally, this study was run in a monologue situation, while speech is most often produced in a dialogue situation. Previous work shows that breathing adapts to dialogue constraints and, in particular, to the turn-taking process (Wlodarczak and Heldner, 2020; Rochet-Capellan and Fuchs, 2014; McFarland, 2001) but did not find overall coordination of breathing (Rochet-Capellan and Fuchs, 2014; McFarland, 2001) except in synchronous reading (Bailly et al., 2013). In conversational settings, speakers have to adapt to their interlocutor. The breathing individuality would be particularly interesting to assess during dialogue as a way to investigate (1) if speakers develop a specific way to control breathing in dialogue consistent across interlocutors, or (2) if they rather lose their individuality to adapt to their interlocutor.

Speech breathing is complex and determined by many variables in interaction with one another in such a way that disentangling speech from breathing mechanisms would require multifactorial analyses with cognitive and physiological parameters. We focused on only a few parameters, and we used a specific method to assess consistency. However, breathing profiles can be described by many other variables, such as physiological patterns (VO2max and VC), or even biomarkers present in the air expired (Zou et al., 2021). Relative contribution of abdominal versus thoracic breathing is also a well-known specific characteristic of speech breathing control (Thorpe et al., 2001). In a recent study (Zou et al., 2021), machine learning was applied to 308 biomarkers of patients' breathing profiles to detect early stages of lung cancer. These techniques allow to determine the most discriminant markers among the 308 features. Such methods could be promising for investigating the breathing individuality. Using wearable devices and conducting studies of speech breathing "in the wild" will probably develop in the near future. Individual consistency of speech breathing within and between conditions might be assessed on larger datasets using different methods. The possibility of characterizing respiratory profiles on the basis of large databases and linking these profiles to other data, such as speech content, will also have to be considered in the context of personal data protection and raises ethical questions which should not be lost from sight.

6.2 To what extent does breathing enlighten speech-movement coordination?

Integrating breathing to the analyses enlightens the complexity of the interactions between speech and limb movements: while cycling movements entrain breathing when quiet, cycling movements are themselves entrained by cognitive load. While arm cycling seems to be a stronger attractor for breathing compared to leg cycling when being quiet, leg cycling impacts speech intensity while arm cycling does not.

Since limb and speech movements frequently co-occur, the brain may have developed specific abilities to control them jointly. This control may rely on biomechanical patterns emerging from the speech-movement entanglement, these patterns being used to optimize energy expenditure (Cos et al., 2011; Shadmehr et al., 2016). One of the systems playing a crucial role in allocating energy resources between speech and movements is breathing (Pouw et al., 2020b). Breathing adapts flexibly to the production of speech (Fuchs and Rochet-Capellan, 2021) as well as to limb movements (Amazeen et al., 2001). When speech and limb movements co-occur, breathing becomes a shared resource shaped by the constraints of the two systems. Hence, breathing is increasingly studied as a central aspect of the relationship between speech and a range of limb movements such as manual gestures accompanying speech in communication (Pouw et al., 2019; Pouw and Dixon, 2019; Pouw et al., 2020d; Pouw et al., 2020b; Pouw et al., 2020a); arm movements for object manipulation (Mateika and Gordon, 2000; Hofstetter et al., 2021); or leg movements for locomotion (Fuchs et al., 2015; Weston et al., 2020). Trouvain and Truong (2015) introduced the 'Talk and Run' database involving 23 participants reading different texts before, during, and after running on a treadmill. They showed higher speech rate, f0 and longer averaged breath durations in text reading after running than before. Physical activity also impacts spectral properties of phonemes. Baker et al. (2008) observed lower breathing frequency and signs of dyspnea when speaking while exercising as compared to exercising alone. Ziegler (2014) observed that an increase in workload during treadmill walking induced a decrease in laryngeal resistance and an increase in airflow.

One of the contribution of this thesis was also to see that improvement in a cognitive task such as recalling a story over time can be reflected in the three actors - speech, breathing and movement. Content recall was better from one day to another in short-term as well as in long-term recalls. This improvement can be associated with the decrease of hand gestures in the arm free condition, the decrease of the biking cycle variability, the increase of breath group and number of syllables, the decrease of the breathing cycle symmetry, and the increase of the breathing cycle duration, over time. Participants may have less difficulty to retrieve with time. Their cognitive load decreases, so they need fewer gestures, their movements are less disturbed, and their breath group organisation is better.

Analysis of speech at the respiratory cycle level revealed that the increase in syllables over time goes hand in hand with a decrease in silent pauses within the respiratory group, and results in a greater number of syllables per group, not an increase in respiratory cycles per session. Even though participants have better recall performance over time, this does not translate into longer recalls, but better organized in terms of breath group.

6.3 Methodological considerations

6.3.1 The challenge of recording and analysing speech, breathing and motion at the same time

Recording and analysing speech, breathing and motion at the same time is a methodological challenge. All the signals have to be synchronized, which is has been a first step to solve. For breathing, we rely on a non-invasive recording method which could be less precise than methods of volume measurement such as masks. Inductance plethysmography is sensitive to trunck motion and arm motion, creating some artefacts in the signals. However, masks substantially limit the production of speech. For movement, we chose a motion capture equipment to record precise 3D movements. However, the equipment is heavy and may prevent participants from behavioring as in dailylife. Simpler systems such as kinect$^{\text{TM}}$ from Microsoft (Redmond, WA) or motion tracking using good quality videos (Pouw et al., 2020e) would have made the participant more at ease.

Besides, working on the interactions between three simultaneous non-stationary and really different time-series is complex. One of the contribution of this thesis is to use methodologies coming from field such as physiology and motor control and to apply them on data involving natural speech. We also applied statistical methods appropriate for describing a behaviour in time, such as generalized additive mixed models. But we may not have considered this temporal dimension enough, to describe the evolution and coordination of the signals (acoustic, movement, breathing) with time. For instance, we could have looked at the temporal coordination between speech and limb movements trough a cross-wavelet analysis between the signal movement and the speech amplitude envelope, enabling to highlight the shared periodicities between the two signals (Pouw and Dixon, 2019). A wavelet analysis is a spectral decomposition continuously across time to estimate how the spectrum changes for each point in time (Grinsted et al., 2004; Issartel et al., 2006; Schmidt et al., 2014). Wavelet transforms allow to consider the nested temporal scales such as breath group, breathing cycles and trials, and see at which scales synchronization occurs within complex and non-stationary time series such as natural speech and body movements.

It is relevant to examine the evolution of respiratory motor coupling with time, to have a clearer view of the periods of synchronization and the corresponding modes. This would also enable to analyze the MRC of each participant and not to put all the participants under an aggregate variable. Indeed, we did not consider enough the fact that the interactions between speech, breathing and limb movements may be highly speaker-specific. This specificity is considered into the linear mixed models, but is not analyzed per se. Individuals can diverge in their behavior regarding the effect of physical activity on speech acoustics (Godin and Hansen, 2015), or regarding MRC patterns (Loring et al., 1990; Mahler et al., 1991a). Such effects can be influenced by individual characteristics such a gender or movement expertise.

6.3.2 Between hypothesis-driven and data-driven approach: strength and limits

The goal in this thesis was to describe the interactions between speech, breathing and limb movements, and enable the emergence of unexpected processes that could appear during spontaneous speech and may not be considered by current theories on speech, breathing or limb control. The task chosen in our study also allowed to test some results of the literature with spontaneous speech. In work studying either MRC or the links between speech and breathing, experiments are usually highly controlled, using isolated word or simplified sentences for speech, and imposing a frequency on limb pace for movement. Controlled experiments are necessary to isolate and highlight the effect of one variable on a phenomenon. Once the effect has been brought out, controlled experiments should be generalized to more ecological experiments to validate such effects in everyday life. As Hamilton and Huth (2020) mention in their recent paper on the use of natural stimuli in speech neuroscience, a good illustration of the necessity of more ecological studies is the field of visual neuroscience over the past twenty years. Within this field, highly controlled experiments were set-up to investigate properties of the visual cortex. These experiments were crucial in understanding the functioning of the visual cortex, with retinototopic cells or directional neurons for instance. However, many mechanisms supposed to generalize to everyday life actually depended on the controlled stimuli and their size effect was significantly weaker in more ecological studies (David et al., 2004). Reporting these issues led to the use of more natural stimuli to model visual processing. Hamilton and Huth use this example in favor of the use of natural language in speech experiments. This statement is encouraged by the major technological advances to measure and record human behaviour continuously over time. Using motion capture systems or movement tracking algorithms, the trajectory of limbs over time is nowadays highly accurate and enable the use of sophisticated statistical tools, such as functional data analysis.

Our study is not purely exploratory, neither guided by haphazardness, but is rooted in the theoretical background of embodied cognition, and is justified by findings from the literature on the role of body movements in spoken communication. Our results inform theories about co-speech gesture as well as motor control, in an attempt to join these two fields together and replace speech in its bodily context. Despite of the density of the dataset recorded, with which an infinity of analyses could be done, the questions we chose to focus on and the analyses conducted were always based on previous work, and we took care of bringing a contribution to each one of them.

Hypothesis testing is inherent to research, but this pratice should not hinder the possibility of doing exploratory studies when they are justified in a proper way. This is a current burning issue in science. Rowbottom and Alexander (2012) investigated the frequency of papers presenting as hypothesis testing while the intent was to bring some answers to a specific question. The authors examined 100 papers about biomechanics published in Journal of Biomechanics and The Journal of Experimental Biology. None of the papers assumed an exploratory methodology, and 58% stated to be hypothesis driven. Out of this 58%, 31% were likely to hide an exploratory aim behind their hypothesis testing. Because of the tyranny of hypothesis-testing, some studies are making up their good question into fake hypothesis, even though the question at the origin of their work is interesting. Exploratory research should not be left aside and should be considered at the same level as hypothesis testing, because both are complementary.

6.4 Theoretical suggestions and perspectives

6.4.1 Suggestions about the joint control of speech and limb movements

The question is to know how the brain deal with speech and concurrent motor behaviors. Are these movements disruptive and independently controlled, or can the brain create synergies between the control of speech and these movements? Either speech is controlled independently of limb movements even when they occur at the same time, or we developped joint control modalities specific to the context, in which case coordination patterns would appear. To include co-speech gestures as well as non-communicative movements cooccuring with speech in the same experiment enabled to compare the effect of both types of movement on speech. We did not find any effect of limb movements on speech fluency and content recall. Aside from the fact that participants do no gesture a lot, another explanation can be put forward. Together with the impact of speech on non-communicative movements, this absence of effect also suggest that speech affects limb movements, but not the contrary (or to a lesser extent). One theory fitting well with such results is the theory of gestures as simulated action (Hostetter and Alibali, 2008; Hostetter and Alibali, 2019): speech is first conceived from mental imagery, which activates the premotor cortex. If this activation exceeds a certain threshold (this threshold being defined by many factors such as context or life experience), the activity of the premotor cortex spreads to the hand control area of the motor cortex (not only the co-articulation control) and triggers hand movements. This theory is also supported by studies on neural correlates of co-speech gestures. Marstaller and Burianova (2015) compared the brain areas activated during the production of speech, pantomime and co-speech gestures. They found that the neural network of co-speech gesture production is mainly driven by the language network.

Marstaller and Burianova (2015) also found that the co-speech gestures network is an overlap between the language network and the network of hand movement production. The dissociation of neural activation area between only-speech and only-gesture task does not support the integrated view of speech and co-speech gestures that McNeill proposes (McNeill, 1992; McNeill and Duncan, 2000). The co-speech gestures as well as the pantomime neural networks involve primary sensory areas responsible for processing sensory input, which suggests that co-speech gestures do not function independently of any sensory feedback loop, as it was suggested regarding the deafferented patient (McNeill, 1992). Motor areas also involved in pedalling, such as the pre-supplementary motor area, basal ganglia, or the cerebellum, are activated during speech alone, and co-speech gesture. The authors suggest that the BA 44 part of Broca's area mediates and integrates sensorimotor signals of speech and gestures, and is the principal actor of their coordination. It could be interesting to see if the BA 44 area fires in a dual-task situation involving speech and non-communicative forelimb movements. Some activation in this area, interfering (or being coordinated) with motor commands sent for pedalling, could explain the increased variability in forelimb kinematics caused by speech. The activation of BA 44 depending on a certain threshold of the activation in the premotor cortex during mental imagery, we can suggest that gesturing frequency increases with the decrease of this threshold. The lower this threshold, the higher BA 44 activation interferes with non-communicative movements. Put simply, the more participants gesture, the more their non-communicative movements may show variability when speaking. This could be tested by including more participants to have a greater range of co-speech gesture frequency.

Lower limb movements were also impacted by speech. This could be due to interference in motor areas shared by both speech and limb motor planning. The impact of speech on limb movements may not be due to the motor but to the cognitive part of speech. Natural speech is such a complex and multidertermined object. In his review on non-oral motor speech, Kent writes: ‘An implicit definition of speech generally is assumed, even though speech is not a monolithic behavior but rather subsumes a variety of sensory, motor, and cognitive skills that vary across behavioral tasks’(Kent, 2015, p.764). Our results suggest that speech fluency, breathing and motion behaviours are driven by the ongoing cognitive activity. One thing to do would be to disentangle the motor from the cognitive mechanisms involved in speech, to test which part does influence movement kinematics. The effect of arithmetic tasks on co-occurent bi-manual coordination task has been tested by Pellechia and colleagues (2010). The authors found that parameters of motor synchronization were impacted by splitting the attention between the motor and cognitive task. On the other hand, Zelic and colleagues (2015) found synchronization patterns between finger flexion/extension and simple nonspeech movements of the articulators. It would also be revelant to conduct cross-linguistic studies to determine which mechanisms are influenced by language and which common effects can be found over different languages. This comparison was one of the goal of the SALAMMBO project. We wanted to reproduce our experiment with French speakers, to compare the results between French and German language. The experiment could not take place because of Covid.

6.4.2 Toward a unified theory of motor control

The work of this thesis directed our reflexion towards the question of the comparison of the control between various types of movement. Schaal and colleagues (2004) pointed out that ’Approaches to understanding the neural and behavioral basis of arm movements have been traditionally divided into two separate fields: one focusing on rhythmic pattern generators and the other on visually guided trajectory formation of reach and grasp movements. [...] At present, rhythmic and discrete movements are investigated by largely distinct research communities using different experimental paradigms and theoretical constructs. As these two classes of movements are tightly interlinked in everyday behavior, a common theoretical foundation spanning across these two types of movements would be valuable.’ (p.1140,1141).

The same reasoning can be made about communicative and non-communicative movements. Neural correlates of co-speech gestures have not been widely studied. Gestures are movements of the body that can also have meaningful content. In what way gestures are different from non meaningful movements? What are the kinematic stakeholders in the production and perception of co-speech gestures ? Can gesture kinematics be studied in the light of current motor control theories of human movement ?

Co-speech gestures have been widely modeled. There is a growing research on how to implement gestures on avatars and robots to make them look human and spontaneous. Latest methods of co-speech gesture generation are data-driven: neural networks are trained on real data with multimodal inputs (speech and text) and can generate gestures and facial expressions on a new set of speech. But unlike non-communicative movements, co-speech gestures have not been modeled through their biomechanical or physiological properties, like pedalling has been for instance (Raasch and Zajac, 1999; Ting et al., 1999).

Runeson and Frykholm (1983) developed the following idea: human movement seems to be submitted to biomechanics laws and motor control principles, such as minimum energy expenditure. Based on these laws learned across perceiving and producing human movements every day, people can infer the trajectory of a human action, intentions, or identify the gender of the agent, and detect deception. Runeson and colleagues argue that these principles stay true for instrumental actions, but not for communicative movements such as conventional gestures, for which biomechanical and motor control principles would not apply.

Indeed, the goal of communicative gestures is not to reach a target in a 2D or 3D space as efficiently as possible, but to convey meaning, in some cases synchronized with co-occurent speech. In this respect, they may not follow general motor control principles. On the other hand, co-speech gestures are still constrained by the physical limits of the body, so laws of optimal control theories may also apply to such gestures (Gribble and Ostry, 1996).

One of the laws of optimal control theories is minimizing the jerk of a movement. The jerk is the derivative of the acceleration. It is a good indicator of the smoothness of the movement, and one of the most exploited metrics regarding human movement in the paradigm of optimal control theory: models minimizing the jerk of a movement have good results at predicting the trajectory of a movement as hand-gripping (Hogan and Flash, 1987). The minimum jerk is linked to the minimum of energy expenditure while realizing a task, characterizing efficient movements. In a recent study on synthesized gesture generation, Kucherenko and collaborators (2020) observed that the generated gestures had a lower jerk than the ground truth (so they were looking too smooth) and could adjust the steps of their model to be closer to the original gestures. They also noticed that jerk values were correlated with the participants preferences concerning the model of gesture generation. This result is a first insight on the relevance of metrics coming from optimal control theories to study communicative movements.

On the other hand, in several studies on production and comprehension of communicative movements, Trujillo et al (2018; 2020) observed that the kinematics of efficient movements such as hand-gripping are modulated by communicative intentions: when a movement is addressed to someone, its amplitude and complexity increases. Communicative movements may be different from other movements because of their kinematic 'inefficiency'. This inefficient property would let observers know that the intention of the gesturing person is to communicate. Indeed, humans are able to distinguish different intentions in the early kinematics of the movement: Cavallo and colleagues (2016) observed that participants were able, with a threshold signficantly above chance, to predict whether the hand in the videoclip was going to grasp the bottle to pour some water in a glass or to drink the water within it. People can also identify whether an actor is going to cooperate or compete (Manera et al., 2011), or whether an actor has a deceptive intention (Runeson and Frykholm, 1983). Cavallo and colleagues (2016) also found that the difference between grasping a bottle either for drinking water or pouring it in a glass lies in kinematic features such as the writs height or the grip aperture. Becchio and collaborators (2008) noticed that cooperative actions were slower, displayed a smaller grip aperture, and had a larger trajectory. The precision in perception of human movement also depends on the motor expertise of the perceiver. Sebanz and colleagues (2009) observed that real passes could be identified compared to fake passes by expert basketball players when observing another player's actions, but not by novice basketballers. This result was the same both when the actions were shown in videos and when they were shown as point-light displays. What emerges from previous studies is that many cues about agents' intentions are conveyed by kinematics, that observers can use to make predictions. Often, cues about the action itself and about intentions of the agent (either to communicate (Pezzulo et al., 2019), to teach (McEllin et al., 2018), or to deceive (Runeson and Frykholm, 1983)) are co-specified in the kinematics. The numerous degrees of freedom of our motor system induce redundancy in movement: a given task such as grasping a cup of coffee can be achieved by many different limb trajectories and kinematics. This redundancy enables intentions and emotions to affect kinematics without impacting the outcome. Investigating the influence of cognitive states such as emotions and intentions on body movements opens up promising research perspectives for theories about motor control.

6.5 Conclusion

This thesis provides new answers and methodological approaches, applied from different fields, to the challenge of understanding the complex and plural links between speech, breathing and limb motion. We found that the breath group evolves with performances over the repetition of a narrative task, and this evolution impacts speech breathing cycles. The speech breathing patterns are likely to be specific to the speakers, and the speaker consistency to be maintained across days and limb motion, unlike quiet breathing. Synchronization between cycling and breathing does not emerge during short narrative sessions, while it seems to be present in quiet conditions. Arm motion seems more sensitive to simultaneous speech than leg motion, while speech intensity seems more impacted by the legs than the arms. These results highlight the complexity of the interactions between the different body mechanisms simultaneously at work. In order to complete the picture and determine the underpinnings of such interactions, it would be relevant to continue studying co-occurrent speech, breathing and limb movements in other motor contexts and languages. The aim of this thesis was also to link different scientific fields between which the permeability can be improved to benefit the understanding of complex body processes and human behavior. It required to review a wide span of literature, tracing the border of a research field at the intersection of complex methods and rich theories. With the improvement of sensors and recording systems and the development of embodied cognition, transversal studies crossing different scientific fields enable to tackle all the dimensions which can emerge from mechanisms usually studied apart from each other. This work will raise new research perspectives and we hope, will encourage the development of more comprehensive studies.

Appendix A

General instructions

A.1 Instructions

You will hear stories about aliens coming to Earth to explore the possibility of settlement. They come from four different groups, have brought a typical vehicle from their planet and want to build a typical house. The stories are shown to you in short videos. At the beginning, you will be introduced to all the aliens, their vehicles and houses. After that you will see stories as videos. We ask you to listen carefully to the stories and their progress. After each story, you will recount what you have seen and heard in one of the following four conditions:

- You are not restricted in any way - You are to put your hands on the chair. - You are to ride a bicycle with your hands - You are to ride a bicycle with your legs and place your hands on the chair.

You are going to tell the story to a person who does not understand German very well. Please do everything you can to make sure that this person understands what you are saying.

Remember as many details of the story as possible, everything is valuable, but especially: - the names of the aliens, vehicles and houses - the description of the aliens, vehicles and houses and their special features - the course of the story

Feel free to talk as much as you can and explain everything to your counterpart. Only stop if you can't remember anything else. If you remember something you have said before, feel free to repeat it.

APPENDIX B

Generation of the pseudo-words

Several methods for generating pseudowords can be found in the literature:

The method WordGen (Duyck et al., 2004) consists in altering existing words by changing 1 phoneme and selecting some of these words by constraint evaluation. Their generation tool is fully orthographic and the generated words have a definite bias towards their starting word. Moreover, choosing the number of altered phonemes does not control the distance of a word to the dictionary words, if we take into account the similarity between phonemes in our notion of distance (e.g. /p/ is closer to /t/ than to /v/).

The method GenerateandTest (Hamed and Zesch,2015) involves generating words randomly and testing the validity of the results. We will use this method to control all necessary constraints.

B.1 Constraints to handle

B.1.1 Phonotactic probabilities

The phonotactic probability of a word represents the frequency with which its component sounds occur in a given language. It has been shown that words with higher phonotactic probabilities would be more typical of a given language (wordlike) than words with lower phonotactic probabilities (Large et al., 1998). It is possible to operationalize this concept by calculating the probability of phonemes, biphones and triphones at a particular place in the word (Vitevitch and Luce, 2004). Pseudowords should therefore have high phonotactic probabilities.

B.1.2 Lexical neighborhood

It has been shown that the size of the lexical neighborhood has a positive influence on the memorization of words and pseudowords (Guitard et al., 2017). The lexical neighborhood can be seen as the minimal distance between the considered word and a word in the base, or the number of new words that can be created by changing a single phoneme of the existing word (Agnello, 2018). This definition has a major drawback: its binary nature prevents it from taking into account the similarity between phonemes. Moreover, our perception is not binary: the words in the lexicon are all similar to some degree. For these 2 reasons, (Bailey and Hahn, 2001) define a new notion of distance, based on the idea of natural classes of phonemes: the similarity between 2 phonemes will be a function of the number of natural classes they share, as well as the number of natural classes of a phoneme not shared by the other phoneme. The distance between a word and a dictionary will then be the smallest distance found within that dictionary. According to Flemming (2005), natural classes are classes of sounds that appear together in phonetic rules. It has been observed that natural classes correspond to simple phonetic features. To define them, we can rely on the feature theory: a natural class is a set of sounds having characteristics in common, and contains all the sounds of the given language having these phonetic features. We have kept the natural classes corresponding to places, modes, sonorities and nasalities. Some simplifications were made in the distance calculation:

diphthongs(ai,au,oy) are considered as their first constituent vowel. The short or long value of the phoneme in German depends on the context (vowel pairs are allophone), so we have grouped them together. The same choice was made for open and closed vowels. After generation, the choice is made according to the context.

B.2 Generation algorithm

- The histogram of syllable distribution is calculated, followed by the distribution of phonemes, biphones and triphones at different positions in the word.
- A large number of words are randomly generated according to the natural distribution of syllable frequencies.
- Among these words, the ones with the highest phonotactic probability scores are selected. The retained words combine a high phonotactic probability score of phonemes and biphones, and an existing triphone sequence.
- We ensure a minimal distance between the retained words. Indeed, maximizing the phonotactic probabilities leads to the generation of words that look alike.
- Only words with a distance between 1 and 2 from the dictionary are selected, to avoid being either too close to an existing word, or too unwordlike. To obtain equivalent words, it was necessary to limit the size of the selection interval. Moreover, only

some phonemes allow to obtain a fixed distance, so choosing a point rather than an interval would have biased the generation.

- a real word of 4 or more phonemes is included in a generated pseudoword, the pseudoword is eliminated. The threshold of 4 phonemes is chosen for the inevitable case of words like "ma". We then retain 30 potential pseudowords per pattern.
- A selection allows us to reach the 15 desired words per syllabic pattern.

Appendix C

Test of association to select the best pseudo-words

Among the 15 words generated, it is necessary to verify that there is no systematic association between :

- the pseudo word and a word of the German language. Indeed, sometimes a prefix is enough to evoke a word, even if the distance between the two words seems to be enough. This would facilitate memory retention and bias the results.
- the pseudo-word and an image. This kind of bias could favor the memorization of one word over another.

For each potential pseudoword, we asked:


Telitik ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

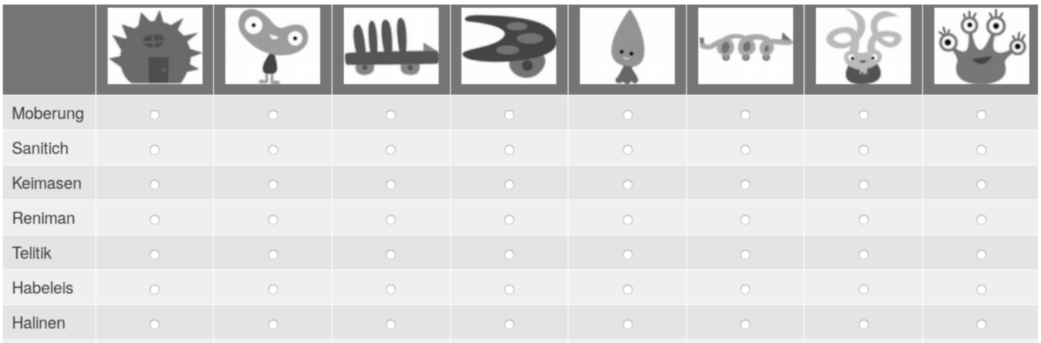
Keimasen ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Melaren ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Figure C.1: Subjective evaluation of wordlikeness. From 1 (does not look like a real word at all) to 5 (very close to a real word)

- Does it sound German? (see Figure C.1)
- What word does it make you think of?
- Which image (among the 12 proposed) does it make you think of? (see Figure C.2)

Words with one of these two associations more often than random are eliminated from the possibilities. We then kept the 12 words maximizing the wordlikeness score.



Moberung	○	○	○	○	○	○	○	○
Sanitich	○	○	○	○	○	○	○	○
Keimasen	○	○	○	○	○	○	○	○
Reniman	○	○	○	○	○	○	○	○
Telitik	○	○	○	○	○	○	○	○
Habeleis	○	○	○	○	○	○	○	○
Halinen	○	○	○	○	○	○	○	○

Figure C.2: Test for associating one word to one picture.

Appendix D

Questionnaire

D.1 General data

Breathing behaviour depends, among other things, on body size and mass. We need this information.

How much do you weigh?

.... kg

How tall are you?

.... cm

D.2 Activities

The following questions are about different activities. 1. do you play a musical instrument? Please tick: Yes No Which instrument(s)? ....

What level do you have? (only for your favourite instrument)? Please tick:

Beginner Intermediate Advanced

2. do you dance? Please tick:

Yes No

Which dance(s)? ....

What level do you have? (only for your favourite dance)? Please tick

Beginner Intermediate Advanced

3) Are you an actor (whether in your professional life or in your activities outside the

professional environment)?

Yes No

How many years have you been acting?

4) Do you do any sports? Please tick:

Yes No

Which sport(s)? ....

What level have you reached? (for your favourite sport only)? Please tick:

Beginner Intermediate Advanced

5) Are there any other activities you have not yet mentioned? Please tick:

Yes No

Which activity(ies)?

....

What level have you reached? (only for your favourite activity)?

Beginner Intermediate Advanced

6) Which languages do you speak at a good to high level?

7) Do you ride a bicycle? Please tick:

Yes No

If yes, how often do you cycle? Please tick:

- Two or three times a year
- Several times a month
- Once a week
- Several times a week
- Every day

D.3 Sporting activities in the last time

The following questions refer to the time you have been physically active in the last 7 days. Please answer these questions even if you do not consider yourself an active person. The questions refer to the physical activities you do at work or while studying, at home, while travelling and in your free time.

D.3.1 Block 1: Intense activities in the last 7 days

1. Please think of all the intensive activities you have done in the last 7 days. Intense physical activities refer to activities that require a lot of physical effort and make breathing much harder than normal. Think only of activities that you have done for at least 10 minutes in a row.

1-a. In the last 7 days, how many days were there when you did strenuous physical activities, such as picking up something heavy, extreme cycling or playing football?

.... Day(s)

I did not have any intense physical activit y==> Start block 2.

1-b. Overall, how much time did you spend on intense activities in the last 7 days?

.... hour(s) .... minute(s)

I don't know

D.3.2 Block 2: Moderate activities in the last 7 days

2. Think of all the moderate activities you have done in the last 7 days. Moderate physical activities refer to activities that require moderate physical effort and make you breathe a little more than normal. Think only of activities that you have done for at least 10 minutes in a row.

2-a. In the last 7 days, how many days were you moderately physically active, for example, carrying light loads or riding a bicycle quietly? Walking (running) is not considered here and will be asked separately later.

.... Day(s)

I did not have any moderate physical activity ==> Start block 3

2-b. In the last 7 days, how much time in total did you spend in moderate activity?

.... hour(s) .... minute(s)

I don't know

D.3.3 Block 3: Running (walking) in the last 7 days

3. Please recall the time you spent running (walking) for at least 10 minutes in a row in the last 7 days. This includes walking to get from one place to another and any other type of walking you did in your free time for relaxation, exercise or recreation.

3-a. In the last 7 days, how many days were there when you walked for at least 10 minutes in a row?

...........Day(s)

I did not walk ==> start block 4.

3. b. In the last 7 days, how many times did you walk for a total of at least ten minutes in a row?

.... time(s)

I don't know.

If you use a pedometer, approximately how many steps have you taken in the last 7 days?

.... step(s)

D.3.4 Block 4: Sitting in the last 7 days

4. The last question is about the amount of time you have sat in the last 7 days. This includes time spent at work, at home, while studying and in your free time. This may include time spent sitting at your desk, visiting friends, reading, sitting or lying down, watching TV. 4-a. In the last 7 days, how much time per day on average did you sit?

.... hour(s) .... minute(s)

Appendix E

Details of the calculation of the index stability

Let A and B be two time series describing two spontaneously coordinated oscillators. Let a and b respectively the series of the time onsets of each cycle of A and B. All the time onsets of the series a and b are concatenated in one vector t. For each component t_k of the vector t, the time onsets of the last cycle of A (t_i) and the last cycle of B (t_j) before or at the same time as t_k are respectively substracted to the time onsets of the first cycle of A (t_{i+1}) and the first cycle of B after or at the same time as t_k (t_{j+1}). The real quotient at the time onset t_k is the ratio between these two substractions.

$$RQ(t_k) = \frac{t_{j+1} - t_j}{t_{i+1} - t_i}$$

Therefore, there is a real quotient of each time onset, whether from A or from B.

Each real quotient $RQ(t_k)$ is assigned a rational ratio $RR_n(t_k)$ from each set F_n. Each set F_n contains rational ratios, named Farey quotients, that are present in a Farey Tree of n levels. Thus, each real quotient $RQ(t_k)$ is assigned n rational ratios (i.e. n rational ratio time series are created, one for each Farey Tree).

for each time onset of A t_i, the phase of A is

$$\theta_A(t_i) = 0 + i[2pi]$$

for each time onset of B t_j, the phase of B is

$$\theta_B(t_j) = 0 + j[2pi]$$

Indeed, at each time onset starts a new cycle, so the phase is a multiple of $2pi$

for each time onset of A t_i, the phase of B is

$$\theta_B(t_i) = \frac{t_i - t_j}{t_{j+1} - t_j}$$

for each time onset of B t_j, the phase of A is

$$\theta_A(t_j) = \frac{t_j - t_i}{t_{i+1} - t_i}$$

The relative phase can be deduced from the computation of the phases:

for each time onset of A t_i, the relative phase between A and B is

$$RP(t_i = q(t_i * \theta_A(t_i)) - p(t_i) * \theta_B(t_i)$$

for each time onset of B t_j, the relative phase between A and B is

$$RP(t_j = q(t_j * \theta_A(t_j)) - p(t_j) * \theta_B(t_j)$$

$p(t_i)$ and $q(t_i)$ are respectively the numerator and denominator of the rational ratio $RR_n(t_i) = p/q$ identified at the time onset t_i in the time series RR_n. Similarly, $p(t_j)$ and $q(t_j)$ are respectively the numerator and denominator of the rational ratio $RR_n(t_j) = p/q$ identified at the time onset t_j in the time series RR_n.

There are n time series of relative phase (one for each level of the Farey Tree).

For each time-series, the fluctuation of the relative phase at the time t_k within its attraction mode $RR_n(t_k) = p/q$ is estimated as follows:

$$IS = \frac{\sqrt{(\sum_{i=1}^{N} cosRP(t_i) + \sum_{j=1}^{M} cosRP(t_j))^2 + (\sum_{i=1}^{N} sinRP(t_i) + \sum_{j=1}^{M} sinRP(t_j))^2}}{N + M}$$

IS is defined on the $[0; 1]$ interval. if $IS = 0$, there is no coordination between A and B. If $IS = 1$, A and B are perfectly coordinated. each of the n time series of relative phase has its own score of stability. The highest stability score characterise the stability of the coordination. The most frequent mode in the corresponding RR_n time series is defined as the mode of coordination.

Appendix F

Naming and identification test of the new words

F.1 Naming test

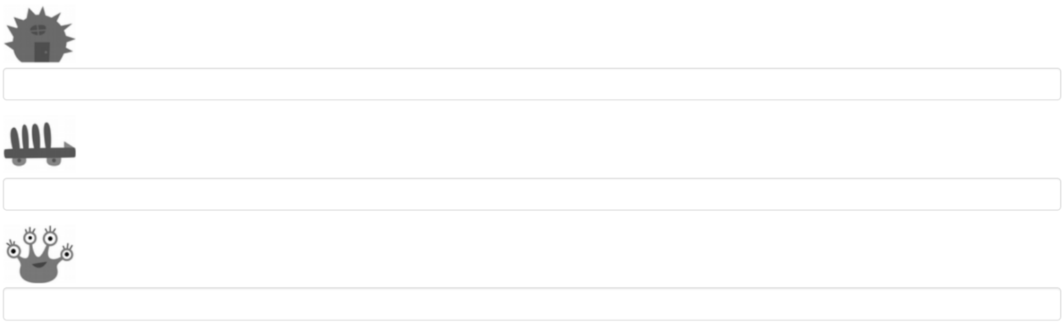
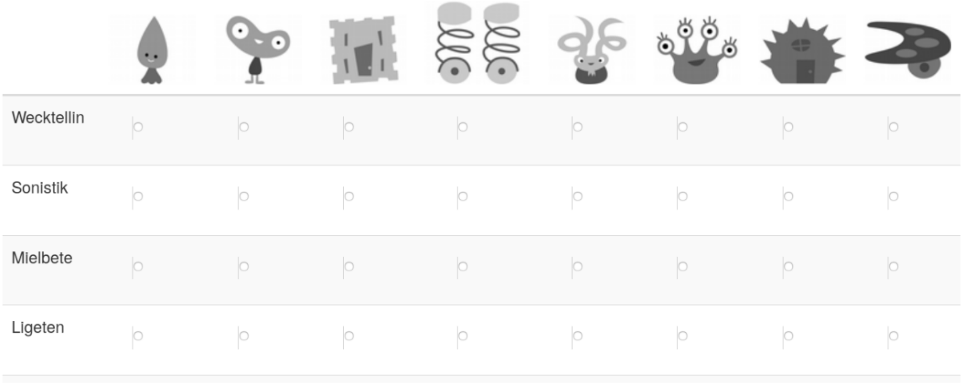


Figure F.1: Presentation of the naming test of the the pseudowords. Participants had to tick the word corresponding to the picture.

F.2 Identification test



Wecktellin	○	○	○	○	○	○	○	○
Sonistik	○	○	○	○	○	○	○	○
Mielbete	○	○	○	○	○	○	○	○
Ligeten	○	○	○	○	○	○	○	○

Figure F.2: Presentation of the identification test of the pseudowords. Participant had to fill in the spaces below the pictures.

Bibliography

Aare, K., Lippus, P., Włodarczak, M., and Heldner, M. Creak in the respiratory cycle. In: *Interspeech 2018, hyderabad, india, september 2-6, 2018*. The International Speech Communication Association (ISCA). 2018, 1408–1412 (cit. on p. 128).

Al-Yahya, E., Dawes, H., Smith, L., Dennis, A., Howells, K., and Cockburn, J. (2011). Cognitive motor interference while walking: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *35*(3), 715–728 (cit. on pp. 10, 11).

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and computation*, *5*(4), 307–331 (cit. on p. 8).

Allen, M., Varga, S., and Heck, D. H. (2022). Respiratory rhythms of the predictive mind. *Psychological Review* (cit. on p. 123).

Amazeen, P. G., Amazeen, E. L., and Beek, P. J. (2001). Coupling of breathing and movement during manual wheelchair propulsion. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1243 (cit. on pp. 32–34, 130).

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis* (Vol. 16). Oxford University Press. (Cit. on p. 1).

Bachy, J.-p., Eberhard, A., Baconnier, P., and Benchetrit, G. (1986). A program for cycle-by-cycle shape analysis of biological rhythms. application to respiratory rhythm. *Computer methods and programs in biomedicine*, *23*(3), 297–307 (cit. on pp. 52, 53).

Bailly, G., Rochet-Capellan, A., and Vilain, C. Adaptation of respiratory patterns in collaborative reading. In: *Interspeech 2013-14th annual conference of the international speech communication association*. 2013, 1653–1657 (cit. on p. 129).

Baker, S. E., Hipp, J., and Alessio, H. (2008). Ventilation and speech characteristics during submaximal aerobic exercise (cit. on pp. 31, 120, 130).

Banzett, R. B., Mahan, S. T., Garner, D. M., Brughera, A., and Loring, S. H. (1995). A simple and reliable method to calibrate respiratory magnetometers and respitrace. *Journal of Applied Physiology*, *79*(6), 2169–2176 (cit. on p. 52).

Bardy, B. G., Hoffmann, C. P., Moens, B., Leman, M., and Dalla Bella, S. (2015). Sound-induced stabilization of breathing and moving. *Annals of the New York Academy of Sciences*, *1337*(1), 94–100 (cit. on pp. 31, 33, 34).

Bardy, B. G., Warren, W. H., and Kay, B. A. (1999). The role of central and peripheral vision in postural control duringwalking. *Perception and psychophysics*, *61*(7), 1356–1368 (cit. on p. 33).

Baringhaus, L, and Henze, N. (2017). Cramér–von mises distance: Probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *Journal of Nonparametric Statistics*, *29*(2), 167–188 (cit. on p. 102).

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in cognitive science*, *2*(4), 716–724 (cit. on p. 5).

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48 (cit. on p. 61).

Becchio, C., Sartori, L., Bulgheroni, M., and Castiello, U. (2008). The case of dr. jekyll and mr. hyde: A kinematic study on social intention. *Consciousness and cognition*, *17*(3), 557–564 (cit. on p. 137).

Bechbache, R., and Duffin, J. (1977). The entrainment of breathing frequency by exercise rhythm. *The Journal of physiology*, *272*(3), 553–561 (cit. on p. 32).

Ben-Tal, A., Wang, Y., and Leite, M. C. (2019). The logic behind neural control of breathing pattern. *Scientific reports*, *9*(1), 1–19 (cit. on p. 29).

Benchetrit, G., Shea, S., Dinh, T. P., Bodocco, S, Baconnier, P, and Guz, A. (1989). Individuality of breathing patterns in adults assessed over time. *Respiration physiology*, *75*(2), 199–209 (cit. on pp. 27, 52, 102, 103).

Bernardet, U., Kanq, S.-H., Feng, A., DiPaola, S., and Shapiro, A. Speech breathing in virtual humans: An interactive model and empirical study. In: *2019 ieee virtual humans and crowds for immersive environments (vhcie)*. IEEE, 2019, 1–9 (cit. on p. 29).

Bernasconi, P., and Kohl, J. (1993). Analysis of co-ordination between breathing and exercise rhythms in man. *The Journal of physiology*, *471*(1), 693–706 (cit. on p. 32).

Bernstein, N. (1968). The co-ordination and regulation of movements. *Science(American Science Association for the Advancement of Science* (cit. on pp. 18, 33).

Besleaga, T., Blum, M., Briot, R., Vovc, V., Moldovanu, I., and Calabrese, P. (2016). Individuality of breathing during volitional moderate hyperventilation. *European journal of applied physiology*, *116*(1), 217–225 (cit. on p. 28).

Beurskens, R., Helmich, I., Rein, R., and Bock, O. (2014). Age-related changes in prefrontal activity during walking in dual-task situations: A fnirs study. *International journal of psychophysiology*, *92*(3), 122–128 (cit. on p. 11).

Binkofski, F., Buccino, G., Posse, S., Seitz, R. J., Rizzolatti, G., and Freund, H.-J. (2001). A fronto-parietal circuit for object manipulation in man: Evidence from an fmri-study. *European Journal of Neuroscience*, *11*(9), 3276–3286 (cit. on p. 22).

Blumenthal, A. L. (1970). Language and psychology: Historical aspects of psycholinguistics (cit. on p. 17).

Boersma, P. (2009). Praat: Doing phonetics by computer (version 5.1. 05). *http://www. praat. org/* (cit. on pp. 59, 60).

Borreggine, K. L., and Kaschak, M. P. (2010). The action–sentence compatibility effect: It's all in the timing. *Cognitive Science*, *30*(6), 1097–1112 (cit. on p. 12).

Boucher, V. J., and Lalonde, B. (2015). Effects of the growth of breath capacities on mean length of utterances: How maturing production processes influence indices of language development. *Journal of Phonetics*, *52*, 58–69 (cit. on p. 128).

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., and Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of cognitive neuroscience*, *18*(10), 1607–1615 (cit. on p. 12).

Bouvet, C. J., Varlet, M., Dalla Bella, S., Keller, P. E., Zelic, G., and Bardy, B. G. (2019). Preferred frequency ratios for spontaneous auditory-motor synchronization: Dynamical stability and hysteresis. *Acta Psychologica*, *196*, 33–41 (cit. on pp. 112, 113).

Bramble, D. M., and Carrier, D. R. (1983). Running and breathing in mammals. *Science*, *219*(4582), 251–256 (cit. on pp. 32, 33, 112, 123).

Bramble, D. M., and Jenkins Jr, F. A. (1993). Mammalian locomotor-respiratory integration: Implications for diaphragmatic and pulmonary design. *Science*, *262*(5131), 235–240 (cit. on p. 32).

Brandimore, A. E., Troche, M. S., Huber, J. E., and Hegland, K. W. (2015). Respiratory kinematic and airflow differences between reflex and voluntary cough in healthy young adults. *Frontiers in Physiology*, *6*, 284 (cit. on p. 30).

Broaders, S. C., Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550 (cit. on p. 7).

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400 (cit. on p. 61).

Brown, A. (1991). A review of the tip-of-the-tongue state experience. *Psychological Bulletin*, *109*(2), 339–347 (cit. on p. 20).

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined tms and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363 (cit. on p. 12).

Buck, J. B. (1938). Synchronous rhythmic flashing of fireflies. *The Quarterly Review of Biology*, *13*(3), 301–314 (cit. on p. 18).

Bunn, J. C., and Mead, J. (1971). Control of ventilation during speech. *Journal of Applied Physiology*, *31*(6), 870–872 (cit. on p. 30).

Butler, P., and Woakes, A. (1980). Heart rate, respiratory frequency and wing beat frequency of free flying barnacle geese branta leucopsis. *Journal of Experimental Biology*, *85*(1), 213–226 (cit. on p. 32).

Butterworth, B., and Beattie, G. Gesture and silence as indicators of planning in speech. In: *Recent advances in the psychology of language*. Springer, 1978, pp. 347–360 (cit. on p. 17).

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54 (cit. on p. 61).

Calbet, J. A., Holmberg, H.-C., Rosdahl, H., van Hall, G., Jensen-Urstad, M., and Saltin, B. (2005). Why do arms extract less oxygen than legs during exercise? *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *289*(5), R1448–R1458 (cit. on p. 89).

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive neuropsychology*, *7*(3), 161–189 (cit. on p. 6).

Caretti, D. M., Pullen, P. V., Premo, L. A., and Kuhlmann, W. D. (1994). Reliability of respiratory inductive plethysmography for measuring tidal volume during exercise. *American Industrial Hygiene Association Journal*, *55*(10), 918–923 (cit. on p. 46).

Casile, A., and Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current biology*, *16*(1), 69–74 (cit. on p. 6).

Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and cognition*, *7*(1), 1–34 (cit. on p. 39).

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., and Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific reports*, *6*(1), 1–8 (cit. on p. 137).

Chawla, P., and Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology*, *30*(6), 580–601 (cit. on pp. 55, 86).

Cherdieu, M., Palombi, O., Gerber, S., Troccaz, J., and Rochet-Capellan, A. (2017). Make gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. *Frontiers in psychology*, *8*, 1689 (cit. on p. 9).

Cherdieu, M., Versace, R., Rey, A. E., Vallet, G. T., and Mazza, S. (2018). Sleep on your memory traces: How sleep effects can be explained by act–in, a functional memory model. *Sleep Medicine Reviews*, *39*, 155–163 (cit. on p. 86).

Chu, M., and Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*(4), 1726–1741 (cit. on p. 21).

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116 (cit. on pp. 8, 86).

Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709 (cit. on p. 8).

Church, R. B., and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71 (cit. on p. 7).

Clarenbach, C. F., Senn, O., Brack, T., Kohler, M., and Bloch, K. E. (2005). Monitoring of ventilation during exercise by a portable respiratory inductive plethysmograph. *Chest*, *128*(3), 1282–1290 (cit. on p. 46).

Clark, A., and Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7–19 (cit. on p. 4).

Commissaris, D. A., Könemann, R., Hiemstra-van Mastrigt, S., Burford, E.-M., Botter, J., Douwes, M., and Ellegast, R. P. (2014). Effects of a standing and three dynamic workstations on computer task performance and cognitive function tests. *Applied ergonomics*, *45*(6), 1570–1578 (cit. on p. 125).

Conrad, B, and Schoenle, P. (1979). Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, *226*(4), 251–268 (cit. on pp. 30, 52).

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*(4), 465–475 (cit. on p. 8).

Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and cognitive processes*, *27*(4), 594–610 (cit. on pp. 8, 86).

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and brain sciences*, *26*(2), 199–208 (cit. on p. 16).

Corballis, M. C. (2009). The evolution of language. (cit. on p. 1).

Correa, Á., Lupiáñez, J., Madrid, E., and Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain research*, *1076*(1), 116–128 (cit. on p. 13).

Cos, I., Bélanger, N., and Cisek, P. (2011). The influence of predicted arm biomechanics on decision making. *Journal of neurophysiology*, *105*(6), 3022–3033 (cit. on p. 130).

Crapse, T. B., and Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587–600 (cit. on p. 13).

Cravotta, A., Busà, M. G., and Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3204–3219 (cit. on pp. 20, 39).

Daffertshofer, A., Huys, R., and Beek, P. J. (2004). Dynamical coupling between locomotion and respiration. *Biological cybernetics*, *90*(3), 157–164 (cit. on p. 112).

David, S. V., Vinje, W. E., and Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, *24*(31), 6991–7006 (cit. on p. 132).

Davie, K., Cardy, J. O., Holmes, J., Gagnon, M, Hyde, A, Jenkins, M., and Johnson, A. M. (2012). The effects of word length, articulation, oral-motor movement, and lexicality on gait: A pilot study. *Gait and posture*, *35*(4), 691–693 (cit. on p. 23).

Dayalu, V., Teulings, H.-L., Bowers, A., Crawcour, S., and Saltuklaroglu, T. (2013). Manual disfluency in drawing while producing and listening to disfluent speech. *Human Movement Science*, *32*(4), 677–690 (cit. on pp. 23, 68, 87, 123).

de Vega, M., Moreno, V., and Castillo, D. (2013). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychological research*, *77*(1), 20–30 (cit. on p. 12).

De Lucca, L., de Oliveira, F. R., Foster, C., and Carminatti, L. J. (2021). Talk test: A simple alternative to identify lactate thresholds during progressive cycling exercise. *Kinesiology*, *53*(1), 20–27 (cit. on p. 30).

Dejours, P. (1966). *Respiration*. Oxford University Press. (Cit. on p. 27).

Del Negro, C. A., Funk, G. D., and Feldman, J. L. (2018). Breathing matters. *Nature Reviews Neuroscience*, *19*(6), 351–367 (cit. on p. 29).

Doust, J., and Patrick, J. (1981). The limitation of exercise ventilation during speech. *Respiration Physiology*, *46*(2), 137–147 (cit. on pp. 30, 31, 120).

Dreyer, F. R., Picht, T., Frey, D., Vajkoczy, P., and Pulvermueller, F. (2020). The functional relevance of dorsal motor systems for processing tool nouns–evidence from patients with focal lesions. *Neuropsychologia*, *141*, 107384 (cit. on pp. 5, 6).

Ebert, D., Raßler, B., and Hefter, H. (2000). Coordination between breathing and forearm movements during sinusoidal tracking. *European journal of applied physiology*, *81*(4), 288–296 (cit. on p. 32).

Eisele, J., Wuyam, B, Savourey, G, Eterradossi, J, Bittel, J., and Benchetrit, G. (1992). Individuality of breathing patterns during hypoxia and exercise. *Journal of Applied Physiology*, *72*(6), 2446–2453 (cit. on pp. 28, 103, 120).

Ejiri, K., and Masataka, N. (1999). Synchronization between preverbal vocal behavior and motor action in early infancy: Ii. an acoustical examination of the functional significance of the synchronization. *Japanese Journal of Psychology*, *69*(6), 433–440 (cit. on p. 19).

Eldridge, F. L., Millhorn, D. E., and Waldrop, T. G. (1981). Exercise hyperpnea and locomotion: Parallel activation from the hypothalamus. *Science*, *211*(4484), 844–846 (cit. on pp. 29, 33).

Ellis, R., and Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British journal of psychology*, *91*(4), 451–471 (cit. on p. 4).

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A tms study. *European journal of Neuroscience*, *15*(2), 399–402 (cit. on p. 12).

Fadiga, L., Craighero, L., and D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York academy of sciences*, *1169*(1), 448–458 (cit. on p. 22).

Farley, B. G., and Koshland, G. F. (2005). Training big to move faster: The application of the speed–amplitude relation as a rehabilitation strategy for people with parkinson's disease. *Experimental brain research*, *167*(3), 462–467 (cit. on p. 87).

Faulkner, K. A., Redfern, M. S., Rosano, C., Landsittel, D. P., Studenski, S. A., Cauley, J. A., Zmuda, J. M., Simonsick, E. M., Kritchevsky, S. B., Newman, A. B., et al.

(2006). Reciprocal influence of concurrent walking and cognitive testing on performance in older adults. *Gait and posture*, *24*(2), 182–189 (cit. on p. 11).

Feher, J. 6.6 - control of ventilation (J. Feher, Ed.; Second Edition). In: In *Quantitative human physiology (second edition)* (J. Feher, Ed.; Second Edition). Ed. by Feher, J. Second Edition. Boston: Academic Press, 2017, pp. 672–681 (cit. on p. 27).

Finlayson, S., Forrest, V., Lickley, R., and Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. *Proceedings of Diss, Gothenburg Papers in Theoretical Linguistics*, *90*, 21–24 (cit. on p. 125).

Fodor, J. A. (1983). *The modularity of mind*. MIT press. (Cit. on p. 6).

Foster, C., P Porcari, J., Ault, S., Doro, K., Dubiel, J., Engen, M., Kolman, D., and Xiong, S. (2018). Exercise prescription when there is no exercise test: The talk test. *Kinesiology*, *50*(Supplement 1.), 33–48 (cit. on p. 30).

Fuchs, S., Hoole, P., Vornwald, D., Gwinner, A., Velkov, H., and Krivokapic, J. The control of speech breathing in relation to the upcoming sentence. In: *Proceedings of the 8th international seminar on speech production*. 2008, 77–80 (cit. on p. 46).

Fuchs, S., Petrone, C., Krivokapić, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in german. *Journal of Phonetics*, *41*(1), 29–47 (cit. on pp. 30, 128).

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. D., and Koenig, L. L. (2015). Assessing respiratory contributions to f0 declination in german across varying speech tasks and respiratory demands. *Journal of Phonetics*, *52*, 35–45 (cit. on pp. 24, 31, 130).

Fuchs, S., and Rochet-Capellan, A. (2021). The respiratory foundations of spoken language. *Annual Review of Linguistics*, *7*, 13–30 (cit. on pp. 30, 52, 123, 130).

Gallagher, S. (2006). *How the body shapes the mind*. Clarendon Press. (Cit. on p. 4).

Garlando, F, Kohl, J, Koller, E., and Pietsch, P. (1985). Effect of coupling the breathing- and cycling rhythms on oxygen uptake during bicycle ergometry. *European journal of applied physiology and occupational physiology*, *54*(5), 497–501 (cit. on p. 32).

Gentilucci, M., Benuzzi, F., Gangitano, M., and Grimaldi, S. (2001). Grasp with hand and mouth: A kinematic study on healthy subjects. *Journal of Neurophysiology*, *86*(4), 1685–1699 (cit. on p. 21).

Gentilucci, M., and Volta, R. D. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology*, *61*(6), 944–957 (cit. on pp. 21, 87).

Gerardin, E., Sirigu, A., Lehéricy, S., Poline, J.-B., Gaymard, B., Marsault, C., Agid, Y., and Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cerebral cortex*, *10*(11), 1093–1104 (cit. on p. 22).

Gillespie, M., James, A. N., Federmeier, K. D., and Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, *132*(2), 174–180 (cit. on p. 8).

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin and review*, *9*(3), 558–565 (cit. on pp. 5, 11).

Godin, K. W., and Hansen, J. H. Analysis and perception of speech under physical task stress. In: *Ninth annual conference of the international speech communication association*. 2008 (cit. on p. 25).

Godin, K. W., and Hansen, J. H. (2011). Analysis of the effects of physical task stress on the speech signal. *The Journal of the Acoustical Society of America*, *130*(6), 3992–3998 (cit. on p. 25).

Godin, K. W., and Hansen, J. H. (2015). Physical task stress and speaker variability in voice quality. *EURASIP Journal on Audio, Speech, and Music Processing*, *2015*(29), 1–13 (cit. on pp. 25, 132).

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 595–607 (cit. on p. 8).

Goldin-Meadow, S., Cook, S. W., and Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological science*, *20*(3), 267–272 (cit. on pp. 7, 123).

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological science*, *12*(6), 516–522 (cit. on p. 8).

Goldin-Meadow, S., and Singer, M. A. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental psychology*, *39*(3), 509–520 (cit. on p. 7).

Goldin-Meadow, S., and Wagner, S. M. (2005). How our hands help us learn. *Trends in cognitive sciences*, *9*(5), 234–241 (cit. on pp. 5, 7, 8).

Golla, F. L., and Antonovitch, S. (1929). The respiratory rhythm in its relation to the mechanism of thought. *Brain*, *52*(4), 491–509 (cit. on p. 129).

Graf, H. P., Cosatto, E., Strom, V., and Huang, F. J. Visual prosody: Facial movements accompanying speech. In: *Proceedings of fifth ieee international conference on automatic face gesture recognition*. IEEE. 2002, 396–401 (cit. on p. 125).

Graham, J. A., and Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*(2), 189–195 (cit. on p. 20).

Grassmann, M., Vlemincx, E., Von Leupoldt, A., Mittelstädt, J. M., and Van den Bergh, O. (2016). Respiratory changes in response to cognitive load: A systematic review. *Neural plasticity* (cit. on p. 129).

Gribble, P. L., and Ostry, D. J. (1996). Origins of the power law relation between movement velocity and curvature: Modeling the effects of muscle mechanics and limb dynamics. *Journal of Neurophysiology*, *76*(5), 2853–2860 (cit. on p. 136).

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics*, *11*(5/6), 561–566 (cit. on p. 132).

Guertin, P. A. (2013). Central pattern generator for locomotion: Anatomical, physiological, and pathophysiological considerations. *Frontiers in neurology*, *3*, 183 (cit. on p. 28).

Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology*, *8*(3-4), 221–228 (cit. on p. 17).

Hadar, U., Steiner, T. J., Grant, E., and Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Human Movement Science*, *2*(1-2), 35–46 (cit. on p. 125).

Haith, A. M., and Krakauer, J. W. (2013). Theoretical models of motor control and motor learning. *Routledge handbook of motor control and motor learning*, 1–28 (cit. on p. 128).

Hamilton, L. S., and Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, cognition and neuroscience*, *35*(5), 573–582 (cit. on p. 132).

Hartig, F., and Hartig, M. F. (2017). Package 'dharma'. *Vienna, Austria: R Development Core Team* (cit. on p. 62).

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307 (cit. on p. 12).

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579 (cit. on p. 1).

Heck, D. H., Kozma, R., and Kay, L. M. (2019). The rhythm of memory: How breathing shapes memory function. *Journal of Neurophysiology*, *122*(2), 563–571 (cit. on p. 14).

Heldner, M., Carlsson, D., and Wlodarczak, M. Does lung volume size affect respiratory rate and utterance duration? In: *Fonetik 2019, stockholm, sweden, june 10–12, 2019.* 2019, 97–102 (cit. on p. 128).

Hessler, E. E., Gonzales, L. M., and Amazeen, P. G. (2010). Displays that facilitate performance of multifrequency ratios during motor-respiratory coordination. *Acta psychologica*, *133*(1), 96–105 (cit. on p. 33).

Hill, A., Adams, J., Parker, B., and Rochester, D. (1988). Short-term entrainment of ventilation to the walking cycle in humans. *Journal of Applied Physiology*, *65*(2), 570–578 (cit. on pp. 34, 107, 120).

Hoetjes, M., Krahmer, E., and Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257–267 (cit. on pp. 20, 74, 88, 125).

Hoffmann, C. P., and Bardy, B. G. (2015). Dynamics of the locomotor–respiratory coupling at different frequencies. *Experimental brain research*, *233*(5), 1551–1561 (cit. on pp. 33, 34, 119).

Hoffmann, C. P., Torregrosa, G., and Bardy, B. G. (2012). Sound stabilizes locomotor-respiratory coupling and reduces energy cost (cit. on pp. 31–33).

Hofstetter, E., Keevallik, L., and Löfgren, A. (2021). Suspending syntax: Bodily strain and progressivity in talk. *Frontiers in Communication*, 96 (cit. on p. 130).

Hogan, N., and Flash, T. (1987). Moving gracefully: Quantitative theories of motor coordination. *Trends in neurosciences*, *10*(4), 170–174 (cit. on p. 136).

Hoit, J. D., and Hixon, T. J. (1986). Body type and speech breathing. *Journal of Speech, Language, and Hearing Research*, *29*(3), 313–324 (cit. on p. 128).

Hoit, J. D., and Hixon, T. J. (1987). Age and speech breathing. *Journal of Speech, Language, and Hearing Research*, *30*(3), 351–366 (cit. on p. 31).

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin and review*, *15*(3), 495–514 (cit. on pp. 17, 134).

Hostetter, A. B., and Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic bulletin and review*, *26*(3), 721–752 (cit. on p. 134).

Hostetter, A. B., Alibali, M. W., and Kita, S. Does sitting on your hands make you bite your tongue? the effects of gesture prohibition on speech during motor descriptions. In: *Proceedings of the annual meeting of the cognitive science society*. *29*. (29). 2007 (cit. on pp. 20, 74, 88).

Hostetter, A. B., Alibali, M. W., and Kita, S. (2007b). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and cognitive processes*, *22*(3), 313–336 (cit. on p. 20).

Hostetter, A. B., and Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, *82*(1), 22–29 (cit. on p. 39).

Hostetter, A. B., and Skirving, C. J. (2011). The effect of visual vs. verbal stimuli on gesture production. *Journal of Nonverbal Behavior*, *35*(3), 205–223 (cit. on p. 39).

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363 (cit. on p. 62).

Huber, J. E. (2007). Effect of cues to increase sound pressure level on respiratory kinematic patterns during connected speech. *Journal of Speech, Language, and Hearing Research* (cit. on p. 31).

Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory physiology and neurobiology*, *164*(3), 323–330 (cit. on pp. 30, 31, 46).

Huber, J. E., Chandrasekaran, B., and Wolstencroft, J. J. (2005). Changes to respiratory mechanisms during speech as a result of different cues to increase loudness. *Journal of Applied Physiology*, *98*(6), 2177–2184 (cit. on p. 31).

Huys, R., Daffertshofer, A., and Beek, P. J. (2003). Learning to juggle: On the assembly of functional subsystems into a task-specific dynamical organization. *Biological cybernetics*, *88*(4), 302–318 (cit. on p. 33).

Hyman Jr, I. E. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. *Applied Cognitive Psychology*, *8*(1), 49–66 (cit. on p. 60).

Issartel, J., Marin, L., Gaillot, P., Bardainne, T., and Cadopi, M. (2006). A practical guide to time—frequency analysis in the study of human motor behavior: The contribution of wavelet transform. *Journal of motor behavior*, *38*(2), 139–159 (cit. on p. 132).

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, *37*(2), 229–261 (cit. on pp. 8, 24).

Iverson, J. M., and Goldin-Meadow, S. (1997). What's communication got to do with it? gesture in children blind from birth. *Developmental psychology*, *33*(3), 453–467 (cit. on p. 16).

Iverson, J. M., and Thelen, E. (1999). Hand, mouth and brain. the dynamic emergence of speech and gesture. *Journal of Consciousness studies*, *6*(11-12), 19–40 (cit. on pp. 4, 19).

Jacobs, N., and Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303 (cit. on p. 86).

Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C. A., Thomas, J. L., Adler, A. B., and Gerzer, R. (2007). Non-linear function model of voice pitch dependency on physical and mental load. *European journal of applied physiology*, *101*(3), 267–276 (cit. on p. 25).

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological science*, *13*(4), 313–319 (cit. on p. 13).

Kawakami, Y., Yamamoto, H., Yoshikawa, T., and Shida, A. (1984). Chemical and behavioral control of breathing in adult twins. *American Review of Respiratory Disease*, *129*(5), 703–707 (cit. on p. 28).

Kelso, J. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *246*(6), R1000–R1004 (cit. on p. 18).

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press. (Cit. on p. 112).

Kelso, J. A., Saltzman, E. L., and Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*(1), 29–59 (cit. on p. 18).

Kemper, S., Herman, R. E., and Lian, C. H. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, *18*(2), 181–192 (cit. on pp. 23, 24, 88, 123).

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. (Cit. on pp. 15, 125).

Kent, R. D. (2015). Nonspeech oral movements and oral motor disorders: A narrative review. *American Journal of Speech-Language Pathology*, *24*(4), 763–789 (cit. on pp. 24, 135).

Kim, J., Cvejic, E., and Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, *57*, 317–330 (cit. on p. 16).

Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*(1), 16–32 (cit. on pp. 17, 20).

Konno, K., and Mead, J. (1967). Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of applied physiology*, *22*(3), 407–422 (cit. on p. 46).

Kontra, C., Goldin-Meadow, S., and Beilock, S. L. (2012). Embodied learning across the life span. *Topics in cognitive science*, *4*(4), 731–739 (cit. on p. 6).

Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of memory and language*, *57*(3), 396–414 (cit. on p. 21).

Kramer, A. F., Erickson, K. I., and Colcombe, S. J. (2006). Exercise, cognition, and the aging brain. *Journal of applied physiology*, *101*(4), 1237–1242 (cit. on p. 14).

Krauss, R. M. (1998). Why do we gesture when we speak? *Current directions in psychological science*, *7*(2), 54–54 (cit. on p. 17).

Krivokapic, J., Tiede, M. K., Tyrone, M. E., and Goldenberg, D. Speech and manual gesture coordination in a pointing task. In: *Speech prosody*. 2016 (cit. on p. 21).

Kucherenko, T., Jonell, P., van Waveren, S., Henter, G. E., Alexandersson, S., Leite, I., and Kjellström, H. Gesticulator: A framework for semantically-aware speech-driven gesture generation. In: *Proceedings of the 2020 international conference on multimodal interaction*. 2020, 242–250 (cit. on p. 136).

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in cognitive science*, *4*(4), 773–785 (cit. on p. 5).

Lamberg, E. M., and Muratori, L. M. (2012). Cell phones change the way we walk. *Gait and posture*, *35*(4), 688–690 (cit. on p. 10).

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1), 119 (cit. on p. 12).

Lausberg, H., and Sloetjes, H. (2016). The revised neuroges–elan system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior research methods*, *48*(3), 973–993 (cit. on pp. 55, 56).

Le Gal, J.-P., Juvin, L., Cardoit, L., and Morin, D. (2016). Bimodal respiratory–locomotor neurons in the neonatal rat spinal cord. *Journal of Neuroscience*, *36*(3), 926–937 (cit. on p. 29).

Lee, H., and Banzett, R. B. (1997). Mechanical links between locomotion and breathing: Can you breathe with your legs? *Physiology*, *12*(6), 273–278 (cit. on p. 32).

Leone, C., Feys, P., Moumdjian, L., D'Amico, E., Zappia, M., and Patti, F. (2017). Cognitive-motor dual-task interference: A systematic review of neural correlates. *Neuroscience and Biobehavioral Reviews*, *75*, 348–360 (cit. on p. 10).

Levelt, W. J., Richardson, G., and La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*(2), 133–164 (cit. on pp. 17, 21).

Li, K. Z., Lindenberger, U., Freund, A. M., and Baltes, P. B. (2001). Walking while memorizing: Age-related differences in compensatory behavior. *Psychological science*, *12*(3), 230–237 (cit. on pp. 11, 87).

Liuzzi, G., Ellger, T., Flöel, A., Breitenstein, C., Jansen, A., and Knecht, S. (2008). Walking the talk—speech activates the leg motor cortex. *Neuropsychologia*, *46*(11), 2824–2830 (cit. on p. 12).

Loring, S. H., Mead, J., and Waggener, T. B. (1990). Determinants of breathing frequency during walking. *Respiration physiology*, *82*(2), 177–188 (cit. on pp. 35, 132).

Lyons, I. M., Mattarella-Micke, A., Cieslak, M., Nusbaum, H. C., Small, S. L., and Beilock, S. L. (2010). The role of personal experience in the neural processing of action-related language. *Brain and language*, *112*(3), 214–222 (cit. on p. 12).

Macintyre, A. D. (2022). *The analysis of breathing and rhythm in speech* (Doctoral dissertation). UCL (University College London). (Cit. on p. 29).

Mahler, D. A., Hunter, B., Lentine, T., and Ward, J. (1991a). Locomotor-respiratory coupling develops in novice female rowers with training. *Medicine and science in sports and exercise*, *23*(12), 1362–1366 (cit. on pp. 32, 132).

Mahler, D. A., Shuhart, C. R., Brew, E., and Stukel, T. A. (1991b). Ventilatory responses and entrainment of breathing during rowing. *Medicine and science in sports and exercise*, *23*(2), 186–192 (cit. on pp. 32, 33, 35).

Malcolm, B. R., Foxe, J. J., Butler, J. S., and De Sanctis, P. (2015). The aging brain shows less flexible reallocation of cognitive resources during dual-task walking: A mobile brain/body imaging (mobi) study. *Neuroimage*, *117*, 230–242 (cit. on p. 10).

Mandler, J. M., and Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology*, *9*(1), 111–151 (cit. on pp. 40, 62).

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., and Perrey, S. (2013). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: A near-infrared spectroscopy study. *Neuroscience research*, *76*(3), 156–162 (cit. on p. 11).

Manera, V., Becchio, C., Cavallo, A., Sartori, L., and Castiello, U. (2011). Cooperation or competition? discriminating between social intentions by observing prehensile movements. *Experimental brain research*, *211*(3), 547–556 (cit. on p. 137).

Manifold, J. A., and Murdoch, B. E. (1993). Speech breathing in young adults: Effect of body type. *Journal of Speech, Language, and Hearing Research*, *36*(4), 657–671 (cit. on p. 128).

Marstaller, L., and Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin and Review*, *20*(3), 496–500 (cit. on p. 8).

Marstaller, L., and Burianová, H. (2015). A common functional neural network for overt production of speech and gesture. *Neuroscience*, *284*, 29–41 (cit. on p. 134).

Mateika, J. H., and Gordon, A. M. (2000). Adaptive and dynamic control of respiratory and motor systems during object manipulation. *Brain Research*, *864*(2), 327–337 (cit. on p. 130).

MATLAB version 9.10.0.1613233 (R2021a). (2021a). The Mathworks, Inc. Natick, Massachusetts. (Cit. on p. 58).

MATLAB version 9.10.0.1613233 (R2021a). (2021b). The Mathworks, Inc. Natick, Massachusetts. (Cit. on p. 60).

Mavilidi, M.-F., Okely, A. D., Chandler, P., Cliff, D. P., and Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational psychology review*, *27*(3), 413–426 (cit. on p. 9).

McCaig, C. M., Adams, S. G., Dykstra, A. D., and Jog, M. (2016). Effect of concurrent walking and interlocutor distance on conversational speech intensity and rate in parkinson's disease. *Gait and posture*, *43*, 132–136 (cit. on p. 24).

McEllin, L., Sebanz, N., and Knoblich, G. (2018). Identifying others' informative intentions from movement kinematics. *Cognition*, *180*, 246–258 (cit. on p. 137).

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* (cit. on pp. 31, 129).

McFarland, D. H., Fortin, A. J., and Polka, L. (2020). Physiological measures of mother–infant interactional synchrony. *Developmental psychobiology*, *62*(1), 50–61 (cit. on p. 46).

McKenna, V. S., and Huber, J. E. (2019). The accuracy of respiratory calibration methods for estimating lung volume during speech breathing: A comparison of four methods across three adult cohorts. *Journal of Speech, Language, and Hearing Research*, *62*(8), 2632–2644 (cit. on pp. 46, 52).

McNeill, D. (1992). *Hand and mind: What gestures reveal abouth thought*. (Cit. on pp. 16, 17, 22, 39, 134).

McNeill, D., and Duncan, S. D. (2000). Growth points in thinking-for-speaking. *Language and gesture*, (1987), 141–161 (cit. on pp. 17, 134).

McNeill, D., Duncan, S. D., Cole, J., Gallagher, S., and Bertenthal, B. (2008). Growth points from the very beginning. *Interaction Studies*, *9*(1), 117–132 (cit. on p. 16).

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., and Töpper, R. (2003). Motor cortex hand area and speech: Implications for the development of language. *Neuropsychologia*, *41*(4), 401–406 (cit. on p. 12).

Melnychuk, M. C., Robertson, I. H., Plini, E. R., and Dockree, P. M. (2021). A bridge between the breath and the brain: Synchronization of respiration, a pupillometric marker of the locus coeruleus, and an eeg marker of attentional control state. *Brain sciences*, *11*(10), 1324 (cit. on p. 13).

Metropolis, N., and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, *44*(247), 335–341 (cit. on p. 102).

Mitchell, H. L., Hoit, J. D., and Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech, Language, and Hearing Research*, *39*(1), 93–104 (cit. on pp. 31, 52).

Mitra, S., Amazeen, P. G., and Turvey, M. T. (1998). Intermediate motor learning as decreasing active (dynamical) degrees of freedom. *Human movement science*, *17*(1), 17–65 (cit. on p. 33).

Mohler, J. (1982). Quantification of dyspnea confirmed by voice pitch analysis. *Bulletin Europeen de Physiopathologie Respiratoire*, *18*(6), 837–850 (cit. on p. 24).

Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature communications*, *5*(1), 1–9 (cit. on pp. 13, 125, 126).

Morin, D., and Viala, D. (2002). Coordinations of locomotor and respiratory rhythms in vitro are critically dependent on hindlimb sensory inputs. *Journal of Neuroscience*, *22*(11), 4756–4765 (cit. on p. 29).

Nyberg, L. (1993). The enactment effect: Studies of a memory phenomenon (cit. on p. 9).

Oppezzo, M., and Schwartz, D. L. (2014). Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of experimental psychology: learning, memory, and cognition*, *40*(4), 1142 (cit. on p. 14).

Oudgenoeg-Paz, O., Volman, M. C. J., and Leseman, P. P. (2012). Attainment of sitting and walking predicts development of productive vocabulary between ages 16 and 28 months. *Infant Behavior and Development*, *35*(4), 733–736 (cit. on p. 24).

Park, H.-D., Barnoud, C., Trang, H., Kannape, O. A., Schaller, K., and Blanke, O. (2020). Breathing is coupled with voluntary action and the cortical readiness potential. *Nature communications*, *11*(1), 1–8 (cit. on pp. 14, 129).

Parrell, B., Goldstein, L., Lee, S., and Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of phonetics*, *42*, 1–11 (cit. on pp. 21, 22, 124).

Pasternak, T., and Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), 97–107 (cit. on p. 5).

Paterson, D. J., Wood, G. A., Morton, A. R., and Henstridge, J. D. (1986). The entrainment of ventilation frequency to exercise rhythm. *European journal of applied physiology and occupational physiology*, *55*(5), 530–537 (cit. on p. 119).

Pellecchia, G. L., Shockley, K., and Turvey, M. T. (2010). Concurrent cognitive task modulates coordination dynamics. *Cognitive science*, *29*(4), 531–557 (cit. on p. 135).

Perry, K. E., and Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist*, *33*(4), 177–194 (cit. on p. 7).

Perry, M., Church, R. B., and Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*(4), 359–400 (cit. on p. 5).

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., and Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of life reviews*, *28*, 1–21 (cit. on p. 137).

Phillipson, E., McClean, P., Sullivan, C., and Zamel, N. (1978). Interaction of metabolic and behavioral respiratory control during hypercapnia and speech. *American Review of Respiratory Disease*, *117*(5), 903–909 (cit. on p. 30).

Pikovsky, A, Rosenblum, M, Kurths, J, and Strogatz, S. (2003). Books-synchronization: A universal concept in nonlinear sciences. *Physics Today*, *56*(1), 47 (cit. on pp. 18, 112).

Pine, K. J., Bird, H., and Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754 (cit. on p. 20).

Pinheiro, J., Bates, D., and R Core Team. (2022). *Nlme: Linear and nonlinear mixed effects models* [R package version 3.1-158]. (Cit. on p. 61).

Plancher, G., Mazeres, F., and Vallet, G. T. (2019). When motion improves working memory. *Memory*, *27*(3), 410–416 (cit. on pp. 13, 125, 126).

Plummer, P., Eskes, G., Wallace, S., Giuffrida, C., Fraas, M., Campbell, G., Clifton, K.-L., Skidmore, E. R., et al. (2013). Cognitive-motor interference during functional mobility after stroke: State of the science and implications for future research. *Archives of physical medicine and rehabilitation*, *94*(12), 2565–2574 (cit. on p. 10).

Pouw, W., De Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., Dixon, J. A., et al. (2020a). Gesture-speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, *1491*(1), 89–105 (cit. on pp. 24, 39, 60, 74, 79, 80, 88, 127, 130).

Pouw, W., and Dixon, J. A. (2019). Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. *Cognitive Science*, *43*(3), e12721 (cit. on pp. 17, 22, 130, 131).

Pouw, W., Harrison, S. J., and Dixon, J. A. (2020b). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391 (cit. on pp. 24, 25, 38, 127, 130).

Pouw, W., Harrison, S. J., and Dixon, J. A. (2020c). The mystery of gesture-speech synchrony in the case of deafferentation: The role of visual control and biomechanics (cit. on p. 16).

Pouw, W., Harrison, S. J., Esteve-Gibert, N., and Dixon, J. A. (2020d). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America*, *148*(3), 1231–1247 (cit. on pp. 24, 127, 130).

Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. Acoustic specification of upper limb movement in voicing. In: *The 6th gesture and speech in interaction conference*. Universitaetsbibliothek Paderborn. 2019, 68–74 (cit. on pp. 24, 25, 38, 130).

Pouw, W., Trujillo, J. P., and Dixon, J. A. (2020e). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior research methods*, *52*(2), 723–740 (cit. on pp. 87, 131).

Pouw, W. T., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., and Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in psychology*, *5*, 359 (cit. on p. 9).

Primov-Fever, A., Lidor, R., Meckel, Y., and Amir, O. (2013). The effect of physical effort on voice characteristics. *Folia Phoniatrica et Logopaedica*, *65*(6), 288–293 (cit. on p. 25).

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature reviews neuroscience*, *6*(7), 576–582 (cit. on pp. 5, 12).

Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797 (cit. on p. 12).

Raasch, C. C., and Zajac, F. E. (1999). Locomotor strategy for pedaling: Muscle groups and biomechanical functions. *Journal of neurophysiology*, *82*(2), 515–525 (cit. on p. 136).

Ragsdale, J. D., and Fry Silvia, C. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. *Language and Speech*, *25*(2), 185–190 (cit. on pp. 16, 20).

Raichlen, D. A., and Alexander, G. E. (2017). Adaptive capacity: An evolutionary neuroscience model linking exercise, cognition, and brain health. *Trends in neurosciences*, *40*(7), 408–421 (cit. on p. 14).

Rassler, B, Nietzold, I, and Waurick, S. (1999). Phase-dependence of breathing and finger tracking movements during normocapnia and hypercapnia. *European journal of applied physiology and occupational physiology*, *80*(4), 324–332 (cit. on p. 33).

Raßler, B., and Kohl, J. (1996). Analysis of coordination between breathing and walking rhythms in humans. *Respiration physiology*, *106*(3), 317–327 (cit. on p. 32).

Rassler, B., and Raabe, J. (2003). Co-ordination of breathing with rhythmic head and eye movements and with passive turnings of the body. *European journal of applied physiology*, *90*(1), 125–130 (cit. on pp. 33, 123).

Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological science*, *7*(4), 226–231 (cit. on p. 20).

Ravizza, S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval? *Psychonomic bulletin and review*, *10*(3), 610–615 (cit. on p. 19).

Recalde, J. P., Pedro, T, Foster, C., Skemp-Arlt, K. M., Fater, D. C., Neese, C. A., Dodge, C., et al. (2002). The talk test as a simple marker of ventilatory threshold. *South African Journal of Sports Medicine*, *9*(2), 5–8 (cit. on p. 1).

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, *26*(6), 867–891 (cit. on pp. 18, 33).

Rimé, B., Schiaratura, L., Hupet, M., and Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*(4), 311–325 (cit. on pp. 20, 125).

Rochet-Capellan, A., and Fuchs, S. The interplay of linguistic structure and breathing in german spontaneous speech. In: *Interspeech 2013-14th annual conference of the international speech communication association*. 2013, 1228 (cit. on pp. 30, 31, 128).

Rochet-Capellan, A., and Fuchs, S. (2014). Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130399 (cit. on pp. 31, 46, 129).

Rochet-Capellan, A., Laboissière, R., Galván, A., and Schwartz, J.-L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task (cit. on pp. 21, 126).

Romaniuk, J., Kasicki, S, Kazennikov, O., and Selionov, V. (1994). Respiratory responses to stimulation of spinal or medullary locomotor structures in decerebrate cats. *Acta neurobiologiae experimentalis*, *54*(1), 11–17 (cit. on p. 33).

Rotstein, A, Meckel, Y, and Inbar, O. (2004). Perceived speech difficulty during exercise and its relation to exercise intensity and physiological responses. *European Journal of Applied Physiology*, *92*(4), 431–436 (cit. on pp. 31, 120).

Roustan, B., and Dohen, M. Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In: *Speech prosody 2010-5th international conference on speech prosody*. 2010, 100110–1 (cit. on p. 21).

Rowbottom, D. P., and Alexander, R. M. (2012). The role of hypotheses in biomechanical research. *Science in Context*, *25*(2), 247–262 (cit. on p. 133).

Ruiter, J. P. d. (1998). *Gesture and speech production*. MPI Series in Psycholinguistics. (Cit. on p. 18).

Ruiter, M., Eielts, C., Loyens, S., and Paas, F. (2019). Comparing cognitive control performance during seated rest and self-paced cycling on a desk bike in preadolescent children. *Journal of Physical Activity and Health*, *16*(7), 533–539 (cit. on p. 125).

Ruiter, M., Loyens, S., and Paas, F. (2015). Watch your step children! learning two-digit numbers through mirror-based observation of self-initiated body movements. *Educational Psychology Review*, *27*(3), 457–474 (cit. on p. 9).

Runeson, S., and Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of experimental psychology: general*, *112*(4), 585 (cit. on pp. 136, 137).

Schaal, S., Sternad, D., Osu, R., and Kawato, M. (2004). Rhythmic arm movement is not discrete. *Nature neuroscience*, *7*(10), 1136–1143 (cit. on p. 135).

Schäfer, C., Rosenblum, M. G., Kurths, J., and Abel, H.-H. (1998). Heartbeat synchronized with ventilation. *Nature*, *392*(6673), 239–240 (cit. on pp. 18, 107).

Schmidt, R., Nie, L., Franco, A., and Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in human neuroscience*, *8*, 633 (cit. on p. 132).

Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of experimental psychology: human perception and performance*, *16*(2), 227 (cit. on p. 33).

Schmidt-Kassow, M., Heinemann, L. V., Abel, C., and Kaiser, J. (2013). Auditory–motor synchronization facilitates attention allocation. *Neuroimage*, *82*, 101–106 (cit. on p. 13).

Schmidt-Kassow, M., Zink, N., Mock, J., Thiel, C., Vogt, L., Abel, C., and Kaiser, J. (2014). Treadmill walking during vocabulary encoding improves verbal long-term memory. *Behavioral and Brain Functions*, *10*(1), 1–9 (cit. on pp. 11, 125).

Schwartz, D. L., and Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 116–136 (cit. on p. 5).

Scott Kelso, J., Tuller, B., and Harris, K. S. A “dynamic pattern” perspective on the control and coordination of movement. In: *The production of speech*. Springer, 1983, pp. 137–173 (cit. on pp. 21, 124).

Sebanz, N., and Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic bulletin and review*, *16*(1), 170–175 (cit. on p. 137).

Serré, H., Dohen, M., Fuchs, S., Gerber, S., and Rochet-Capellan, A. (2021). Speech breathing: Variable but individual over time and according to limb movements. *Annals of the New York Academy of Sciences*, *1505*(1), 142–155 (cit. on pp. 53, 91, 102–106, 127).

Serré, H., Fuchs, S., Dohen, M., and Rochet-Capellan, A. Speaking while moving: Does the head compensate for the hands not being able to move? In: *2020-12th international seminar on speech production*. 2020 (cit. on pp. 124, 125).

Shadmehr, R., Huang, H. J., and Ahmed, A. A. (2016). A representation of effort in decision-making and motor control. *Current biology*, *26*(14), 1929–1934 (cit. on p. 130).

Shea, S., Benchetrit, G, Dinh, T. P., Hamilton, R., and Guz, A. (1989). The breathing patterns of identical twins. *Respiration physiology*, *75*(2), 211–223 (cit. on p. 28).

Shea, S., Walter, J, Murphy, K, and Guz, A. (1987). Evidence for individuality of breathing patterns in resting healthy man. *Respiration physiology*, *68*(3), 331–344 (cit. on pp. 28, 121).

Shea, S. A. (1996). Behavioural and arousal-related influences on breathing in humans. *Experimental Physiology: Translation and Integration*, *81*(1), 1–26 (cit. on pp. 27, 30, 119, 123).

Shea, S. A., Horner, R. L., Benchetrit, G., and Guz, A. (1990). The persistence of a respiratory 'personality'into stage iv sleep in man. *Respiration physiology*, *80*(1), 33–44 (cit. on pp. 28, 30, 129).

Shebani, Z., and Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm-and leg-related action words. *Cortex*, *49*(1), 222–231 (cit. on pp. 12, 126).

Shebani, Z., and Pulvermüller, F. (2018). Flexibility in language action interaction: The influence of movement type. *Frontiers in Human Neuroscience*, *12*, 252 (cit. on pp. 12, 126).

Shevtsova, N. A., Marchenko, V., and Bezdudnaya, T. (2019). Modulation of respiratory system by limb muscle afferents in intact and injured spinal cord. *Frontiers in neuroscience*, *13*, 289 (cit. on pp. 28, 120, 123).

Singer, M. A., and Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85–89 (cit. on p. 7).

Smith, J. C., Abdala, A. P., Rybak, I. A., and Paton, J. F. (2009). Structural and functional architecture of respiratory networks in the mammalian brainstem. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1529), 2577–2587 (cit. on p. 29).

Sperry, E. E., and Klich, R. J. (1992). Speech breathing in senescent and younger women during oral reading. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1246–1255 (cit. on pp. 30, 31).

Sporer, B. C., Foster, G. E., Sheel, A. W., and McKenzie, D. C. (2007). Entrainment of breathing in cyclists and non-cyclists during arm and leg exercise. *Respiratory physiology and neurobiology*, *155*(1), 64–70 (cit. on p. 119).

Srygley, J. M., Mirelman, A., Herman, T., Giladi, N., and Hausdorff, J. M. (2009). When does walking alter thinking? age and task associated findings. *Brain research*, *1253*, 92–99 (cit. on p. 10).

Stevanoni, E., and Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, *29*(4), 217–233 (cit. on pp. 8, 39).

Strack, F., Martin, L. L., and Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology*, *54*(5), 768–777 (cit. on p. 4).

Team, R. C. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. (Cit. on p. 61).

Temprado, J., Milliex, L, Grélot, L, Coyle, T, Calvin, S, and Laurent, M. (2002). A dynamic pattern analysis of coordination between breathing and rhythmic arm movements in humans. *Neuroscience letters*, *329*(3), 314–318 (cit. on pp. 18, 33, 123).

Thomas, S. P., and Suthers, R. A. (1972). The physiology and energetics of bat flight. *Journal of Experimental Biology*, *57*(2), 317–335 (cit. on p. 32).

Thorpe, C. W., Cala, S. J., Chapman, J., and Davis, P. J. (2001). Patterns of breath support in projection of the singing voice. *Journal of Voice*, *15*(1), 86–104 (cit. on p. 129).

Ting, L. H., Kautz, S. A., Brown, D. A., and Zajac, F. E. (1999). Phase reversal of biomechanical functions and muscle activity in backward pedaling. *Journal of Neurophysiology*, *81*(2), 544–551 (cit. on p. 136).

Torbeyns, T., de Geus, B., Bailey, S., De Pauw, K., Decroix, L., Van Cutsem, J., and Meeusen, R. (2016). Cycling on a bike desk positively influences cognitive performance. *PLoS One*, *11*(11), e0165510 (cit. on p. 125).

Treffner, P., and Peter, M. (2002). Intentional and attentional dynamics of speech–hand coordination. *Human Movement Science*, *21*(5-6), 641–697 (cit. on pp. 18, 124).

Treffner, P. J., and Turvey, M. T. (1993). Resonance constraints on rhythmic movement. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(6), 1221 (cit. on p. 112).

Trouvain, J., and Truong, K. P. Prosodic characteristics of read speech before and after treadmill running. In: *Sixteenth annual conference of the international speech communication association*. 2015 (cit. on pp. 25, 130).

Trujillo, J. P., Simanova, I., Bekkering, H., and Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A kinect study. *Cognition*, *180*, 38–51 (cit. on p. 136).

Trujillo, J. P., Simanova, I., Bekkering, H., and Özyürek, A. (2020). The communicative advantage: How kinematic signaling supports semantic comprehension. *Psychological research*, *84*(7), 1897–1911 (cit. on p. 136).

Tucker, M., and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 830–846 (cit. on p. 4).

Tuite, K. (1993). The production of gesture (cit. on p. 18).

Van Rossum, G., and Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace. (Cit. on p. 59).

Varela, F. J., Thompson, E., and Rosch, E. (2016). *The embodied mind, revised edition: Cognitive science and human experience*. Massachusetts Institute Technology press. (Cit. on p. 4).

Varlet, M., and Richardson, M. J. (2011). Computation of continuous relative phase and modulation of frequency of human movement. *Journal of biomechanics*, *44*(6), 1200–1204 (cit. on pp. 108, 109).

Verghese, J., Kuslansky, G., Holtzer, R., Katz, M., Xue, X., Buschke, H., and Pahor, M. (2007). Walking while talking: Effect of task prioritization in the elderly. *Archives of physical medicine and rehabilitation*, *88*(1), 50–53 (cit. on pp. 23, 68, 87, 88, 123).

Villard, S., Casties, J.-F., and Mottet, D. (2005). Dynamic stability of locomotor respiratory coupling during cycling in humans. *Neuroscience letters*, *383*(3), 333–338 (cit. on p. 32).

von der Malsburg, T. (2015). Py-span-task: A software for testing working memory span. *Zenodo. doi*, *10* (cit. on p. 49).

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: An overview. (Cit. on pp. 15, 16, 123).

Wagner, S. M., Nusbaum, H., and Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395–407 (cit. on pp. 8, 126).

Walle, E. A., and Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental psychology*, *50*(2), 336–348 (cit. on p. 10).

Wallmann-Sperlich, B., Chau, J. Y., and Froboese, I. (2017). Self-reported actual and desired proportion of sitting, standing, walking and physically demanding tasks of office employees in the workplace setting: Do they fit together? *BMC Research Notes*, *10*(1), 1–6 (cit. on p. 125).

Wells, G. L., and Petty, R. E. (1980). The effects of over head movements on persuasion: Compatibility and incompatibility of responses. *Basic and applied social psychology*, *1*(3), 219–230 (cit. on p. 4).

Weston, H., Fuchs, S., and Rochet-Capellan, A. Speech during light physical activity: Effect on f0 and intensity. In: *12th international seminar on speech production*. 2020, 246–249 (cit. on pp. 25, 130).

Wexler, M., Kosslyn, S. M., and Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77–94 (cit. on p. 4).

Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129 (cit. on p. 5).

Whipp, B. J., Ward, S. A., Lamarra, N., Davis, J. A., and Wasserman, K. (1982). Parameters of ventilatory and gas exchange dynamics during exercise. *Journal of Applied Physiology*, *52*(6), 1506–1513 (cit. on p. 29).

Whitfield, J. A., Kriegel, Z., Fullenkamp, A. M., and Mehta, D. D. (2019). Effects of concurrent manual task performance on connected speech acoustics in individuals with parkinson disease. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2099–2117 (cit. on pp. 23, 88).

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. (Cit. on p. 62).

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*, *70*, 86–116 (cit. on p. 80).

Wilder, C. N. (1974). Respiratory patterns in infant cry. *Developmental psychology*, *6*, 293–301 (cit. on p. 30).

Wilke, J. T., Lansing, R. W., and Rogers, C. A. (1975). Entrainment of respiration to repetitive finger tapping. *Physiological Psychology*, *3*(4), 345–349 (cit. on p. 32).

Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., and Casasanto, D. (2011). A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. *Psychological science*, *22*(7), 849–854 (cit. on p. 12).

Winkworth, A. L., Davis, P. J., Adams, R. D., and Ellis, E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech, Language, and Hearing Research*, *38*(1), 124–144 (cit. on pp. 30, 128).

Winkworth, A. L., Davis, P. J., Ellis, E., and Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. *Journal of Speech, Language, and Hearing Research*, *37*(3), 535–556 (cit. on pp. 30, 119, 123).

Wlodarczak, M., and Heldner, M. (2020). Breathing in conversation. *Frontiers in Psychology*, 2574 (cit. on pp. 31, 129).

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks*, *11*(7-8), 1317–1329 (cit. on p. 128).

Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman; Hall/CRC. (Cit. on p. 61).

Zelano, C., Jiang, H., Zhou, G., Arora, N., Schuele, S., Rosenow, J., and Gottfried, J. A. (2016). Nasal respiration entrains human limbic oscillations and modulates cognitive function. *Journal of Neuroscience*, *36*(49), 12448–12467 (cit. on pp. 13, 14).

Zelic, G., Kim, J., and Davis, C. (2015). Articulatory constraints on spontaneous entrainment between speech and manual gesture. *Human Movement Science*, *42*, 232–245 (cit. on pp. 21, 124, 126, 135).

Zelic, G., Varoqui, D., Kim, J., and Davis, C. (2018). A flexible and accurate method to estimate the mode and stability of spontaneous coordinated behaviors: The index-of-stability (is) analysis. *Behavior Research Methods*, *50*(1), 182–194 (cit. on pp. 107, 108, 112).

Zhang, Z. (2016). Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation. *Journal of Voice*, *30*(6), 760–e7 (cit. on p. 128).

Ziegler, A. S. (2014). *Effects of vocal intensity and physical activity levels on phonatory and respiratory function* (Doctoral dissertation). University of Pittsburgh. (Cit. on p. 130).

Zivotofsky, A. Z., and Hausdorff, J. M. (2007). The sensory feedback mechanisms enabling couples to walk synchronously: An initial investigation. *Journal of neuroengineering and rehabilitation*, *4*(1), 1–5 (cit. on p. 18).

Zou, Y., Wang, Y., Jiang, Z., Zhou, Y., Chen, Y., Hu, Y., Jiang, G., and Xie, D. (2021). Breath profile as composite biomarkers for lung cancer diagnosis. *Lung Cancer*, *154*, 206–213 (cit. on p. 129).

Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11 (cit. on p. 12).