Working memory differences in prosodic imitation

Caterina Petrone, Daria d'Alessandro, Simone Falk

Version of Record: https://www.sciencedirect.com/science/article/pii/S0095447021000759
Manuscript_542842666575a01864196674d7b09f21

Working memory differences in prosodic imitation

Caterina Petrone[1], Daria D'Alessandro[2, 3] and Simone Falk[1,2, 4]

[1]CNRS, Aix-Marseille Université, LPL, UMR 7309, Aix-en-Provence, France

[2] Université Paris 3 Sorbonne Nouvelle, Laboratoire de Phonétique et Phonologie, UMR 7018, Paris, France

[3]University "Federico II", Naples, Italy

[4] University of Montreal, Department of Linguistics and Translation, Montreal, Canada

Running Head: Prosodic imitation in French

Corresponding author:

Caterina Petrone

Laboratoire Parole et Langage

5, av. Pasteur

13100 Aix-en-Provence

FRANCE

Email: caterina.petrone@lpl-aix.fr

Declarations of interest: none

Abstract

Speakers strongly vary in their imitation abilities, but the factors underlying this variation are still unclear. This study examined whether individual differences in working memory affect the accuracy of imitation of phonological and phonetic aspects of French prosody. Thirty-six French native speakers were asked to listen to twenty sentences extracted from a read and a spontaneous speech corpus, and to repeat the words and the way the utterances were said. Overall, obligatory phonological events (boundary tones and the H* tone of LH* rises) were more accurately reproduced than optional phonological ones (the Hi tone of LHi rises) and their speaker-specific phonetic details. Speakers with higher working memory capacities were more accurate in phonological imitation of both obligatory and optional phonological events, possibly because of their increased capacity in retaining the prosodic characteristics of the utterances. Imitating read speech, which was richer in terms of number of LHi rises, was slightly more difficult for speakers with low working memory capacities. There was no relation between working memory and imitation of phonetic aspects, which showed more idiosyncratic patterns of imitation. Our findings indicate that working memory constraints should be taken into account in modelling prosodic imitation, along with linguistic and task-specific factors.

Number of words : 200 (max 200)

Keywords : Prosodic imitation, working memory, speaking style, French

1. Introduction

In conversation, speakers can adapt to their interlocutors on a variety of linguistic levels, including syntax, lexicon, phonology and phonetics (e.g., Pardo, 2006; Garrod and Pickering, 2009; Fusaroli et al., 2012; Nguyen and Delvaux, 2015; Garrod et al., 2018). One important mechanism that potentially underlies adaptation is imitation (e.g., Goldinger, 1998; Pickering and Garrod, 2013). In particular, phonetic imitation, i.e., the tendency for a speaker to reproduce the phonetic aspects of the speech of another speaker, has been widely studied at the segmental level (e.g., for vowel quality, e.g., Babel, 2009; Sato et al., 2013; for Voice Onset Time, e.g., Nielsen, 2011).

In the field of prosody, a number of studies have focused on imitation of overall phonetic aspects of utterances, such as speech rate, fundamental frequency (f0) mean, median and f0 range (see De Looze et al., 2014 for a review). Few studies have examined the imitation of intonation. Here, it was found that speakers can imitate phonological events (e.g., pitch accents and edge tones), within the same language or across different languages or varieties (e.g., Michelas and Nguyen, 2011; German, 2012; D'Imperio et al., 2014; German and D'Imperio, 2015; Petrone et al., 2017). For example, in a preliminary cross-language study, Petrone et al. (2017) foundthat Neapolitan speakers can accurately imitate the phonological form of German polar questions (an utterance-final H- H% rise) and even the form's phonetic details (the f0 slope of such a rise). However, the degree of imitation may depend on the constraints of the native phonological system (German, 2012; D'Imperio et al., 2014; German and D'Imperio, 2015). In a cross-variety study on Bari and Neapolitan Italian, D'Imperio et al. (2014) asked participants to listen to a speaker using an "unfamiliar dialect" and to imitate the speaker's pronunciation as closely as possible. They found that both Neapolitan and Bari speakers imitated within-category differences in tonal alignment of the model speaker of the other variety. However, Neapolitan speakers produced less accurate alignment than Bari speakers. In fact, Neapolitan speakers showed an overshoot relative to the model speaker, i.e., they moved tonal alignment much earlier than the model speaker when

reproducing L+H* Bari Italian nuclear accents. The overshoot is interpreted as a strategy to preserve a native phonological contrast between two rising accents (L+H* vs. L*+H).

Another line of work underscores the role of the native phonological system in imitation. Using natural English utterances, Braun et al. (2006) randomly generated stimuli containing f0 contours not normally found in English. British English participants were requested to imitate the f0 contours in different sessions, with the productions of a previous session becoming the input for the next one. Speakers produced f0 contours converging towards a limited set of distinct tunes or 'attractors' which were typical of English. The authors also reported strong individual variation, as participants differed in the number of iterations they needed to converge towards the attractors (i.e., some of the participants were more adept in imitating the phonetic characteristics of the non-native contours over the sessions, while others shifted almost immediately towards the attractors from the first session on). Cole and Shattuck-Hufnagel (2011) found that American English participants imitated the phonological patterns in American English target utterances, but they did not reproduce the specific phonetic cues to prosodic structure employed by the model speaker, such as the duration of silent pauses and the occurrence of irregular pitch periods. The authors speculate that imitators may reproduce a phonological pattern idiosyncratically, that is, by using different phonetic cues from the model speaker or by weighting the same cues in an individual way.

In this paper, we contribute to a further understanding of the processes underlying prosodic imitation by evaluating how individual variability modulates the ability to reproduce phonological and phonetic aspects of prosody in French. While individuals strongly vary in their imitation abilities, the factors underlying this variation are still unclear. A few studies have found that prosodic imitation at both phonological and phonetic levels can be impacted by individual differences, such as auditory preferences (Postma-Nilsenová and Postma, 2013) or rhythmic abilities (Cason et al., 2019). For instance, Postma-Nilsenová and Postma (2013) found that 'fundamental listeners' (listeners who primarily focus on fundamental frequency) are more capable of imitating f0 than 'spectral listeners' (listeners who focus primarily on individual harmonics).

Here, we will deal with an additional source of individual differences: working memory. Though working memory has been already linked to speech imitation, as yet no studies have examined its impact on the imitation of prosody.

Working memory (WM) refers to the memory sub-system which allows people to temporarily store information during an ongoing cognitive task (Baddley and Hitch, 1974; Baddeley, 2000). Working memory capacity appears limited and varies across individuals, constraining the way stimuli are stored and processed by different people (Miller, 1956; Luck and Vogel, 1997; Cowan, 2001; Cowan et al., 2008). For example, people with high working memory capacity perform a variety of real-world tasks (such as reading comprehension or problem solving) better than people with low working memory capacity, as they can store a higher amount of information to help them perform such tasks (Engle and Kane, 2004; Hartsuiker and Barkhuysen, 2006). Working memory constraints could explain individual differences in syntactic parsing or speech production planning, for instance (cf. Swets et al., 2007, 2014 and references therein). Individual differences in working memory capacities could also partially explain differences in prosodic strategies. Swets et al. (2007) argue, based on their findings in listener comprehension of syntactically ambiguous sentences, that working memory capacities affect prosodic structure, such that readers with low WM are more likely to chunk a text into smaller prosodic phrases than readers with high WM. Petrone and colleagues (2011) found that speakers with higher working memory capacities have larger scopes of prosodic planning compared to speakers with low working memory capacities, as evidenced by measures of f0 declination.

Some theories of WM have revealed that higher working memory capacities not only provide a greater capacity to retain information but also greater efficiency (Swets et al., 2014). Given that working memory capacity is limited, such theories assume that attentional selection mechanisms can be used to ensure that only task-relevant information is stored in WM (e.g., Cowan, 1988, Oberaurer, 2013, 2019; Kane et al., 2007; Swets et al., 2014). Broadly speaking, people with high

WM capacity may be more capable of retaining only the features which are currently relevant to the goals of the task and of keeping task-irrelevant information out of their WM so that it does not consume part of their storage resource. Conversely, people with low WM capacity may be less capable of filtering out irrelevant information, resulting in a higher burden for working memory (see Oberaurer, 2019 and Swets et al., 2014 for a more detailed discussion about the relationship between storage and efficiency). In phonetics/phonology, it is unclear to what extent working memory can help an individual to retain phonetic/phonological information, and what the impact of individual working memory differences on imitation may be (Poll et al., 2013; Yu et al., 2013; Nguyen and Delvaux, 2015; Christiner and Reiterer, 2013, 2018). It is often assumed that phonetic details of a speech sound fade away from working memory within a few seconds, unless refreshed by articulatory rehearsal, for instance (Baddeley, 2001). However, speakers can differ widely in their ability to retain and reproduce phonetic patterns. Christener and Reiterer (2013) evaluated the effects of working memory capacities on phonetic imitation in German speakers. They asked their participants to listen to utterances in a foreign language that they either knew (English) or did not know (Hindi) and to imitate the pronunciation of the model speaker. Non-expert raters listened to the utterances produced by the imitators and judged overall accuracy of imitation, based on the intonation, speech rate, fluency and intelligibility of the utterances. Speakers with higher working memory capacity were judged as more accurate in phonetic imitation than speakers with lower working memory capacity. This effect was stronger for Hindi than for English, possibly because of their increased capacity to remember and repeat the acoustic characteristics of utterances in absence of the knowledge of the language (thus, in absence of the influence of L2 phonological representations). Yu et al. (2013) looked at the relationship between individual working memory capacities and imitation of phonetic detail at the segmental level, i.e., in Voice Onset Time (VOT) imitation. In their experiment, they exposed American English speakers to auditory sentences in which target words starting with a plosive were manipulated as for their VOT duration. After exposure, the speakers read the same target words aloud, and these words were then evaluated on

their degree of imitation. No correlation was found between VOT imitation and speakers' working memory capacities. The authors suggested that executive functioning (and not working memory capacity per se) could affect phonetic imitation. Lewandowski's (2012, 2019) exemplar-based accounts also explore the effects of individual differences in working memory on phonetic imitation. In exemplar-based approaches, every perceived spoken item leaves a unique 'episodic' trace in memory which contains detailed information, including phonetic information about speakers' voices (Goldinger, 1998; Pardo, 2006). During phonetic imitation, listeners use episodic traces to shift their production of overall phonetic features and phonetic details in the direction of what they have heard (Nielsen, 2011). Lewandowski (2012) assumed that the richness of the exemplars varies among individuals depending on the amount of phonetic information that they can retain in working memory. In this view, individuals with higher working memory capacities would be more capable of constructing richer-indexed exemplars including more phonetic detail than individuals with lower working memory capacities.

In sum, individual differences in working memory capacity affect both speech perception and production, and they may explain variability in prosodic imitation. To our knowledge, only one study (Yu et al., 2013) has been carried out on the correlation between working memory and phonetic detail, which failed to find such a correlation at the segmental level.

Therefore, the aim of the present study is to further elucidate the link between working memory and the imitation of phonological aspects as well as phonetic detail of prosody. We refer here to 'phonetic detail' as systematic phonetic variation excluded from abstract representations (Cangemi, 2014). For instance, within the traditional Autosegmental-Metrical model of intonation (Pierrehumbert, 1980; Ladd, 2008, among others), the intonation contour is decomposable into a sequence of abstract L and H tones whose tonal targets are solely defined in their phonetic alignment and scaling. f0 slope, for example, is considered a phonetic detail, as it does not play any role in this model in defining phonological contrasts (but see Cangemi, 2014 and references

therein). We chose French as a test language because of its peculiarities at the prosodic level, which allows us to verify to what extent previous findings can be generalized across languages. In the following sections, we first introduce some basic notions involving the prosody and intonation of French (Section 1.1). After detailing our hypotheses (Section 1.2), we present our imitation experiment (Sections 2 and 3). The pattern of the results and their theoretical implications are discussed in the final section (Section 4).

1.1. Basics of French prosody

Models of intonation greatly differ in the number and types of prosodic constituents they posit for French (e;g., Di Cristo, 2000; Jun and Fougeron, 2000, 2002; Post, 2000; Astésano, 2001; Michelas and D'Imperio, 2010; Delais-Roussarie et al., 2015; Garnier et al., 2016). Two levels are widely agreed upon: the Intonational Phrase and, at a lower level, a constituent containing at least one content word, plus any associated function word, whose label differs depending on the theoretical account (e.g. '*intonème mineur'*, Delattre, 1966; Rossi, 1985, 1999; 'prosodic word', Vaissière, 1992; 'Rhythmic Unit', Di Cristo and Hirst, 1993, 'Accentual Phrase', Jun and Fougeron, 2000). This prosodic constituent is characterized by a final rise (also called '*accent primaire*', 'primary stress', 'late rise', or 'final accent'), which defines its right edge, and an initial rise (also called '*accent sécondaire*', 'secondary stress', 'early rise', or 'initial accent', see Welby, 2003 for a review), which marks its left edge (e.g., Pasdeloup, 1990; Hirst and Di Cristo, 1996; Di Cristo, 1998; Welby, 2003; Jun and Fougeron, 2000, 2002; German and D'Imperio, 2015). Traditional accounts of French intonation consider the late rise as obligatory, ie., it is always realized in the intonation contour. The late rise marks the primary stress of a word at a phrasal level. In fact, differently from lexical-stress languages such as English or Italian, primary stress in French is always word final and its realization depends on the position of the word within the phrase. On the other hand, the early rise is optional, i.e. it may be not realized in the intonation contour, and it has only a rhythmic function (see German and D'Imperio, 2015 and references therein). We note

though that the status of the early rise is controversial in the literature. For instance, it has been suggested that the early rise is part of the lexical entry of a word (Di Cristo, 1999; Astésano, 2001) and that might also signal pragmatic functions, such as highlighting a semantic unit (e.g., Vaissière, 1991; Di Cristo, 1999; Astésano, 2001).

In this paper, we adopt Jun and Fougeron's (2000, 2002) Autosegmental-Metrical (AM) model of French intonation in which the basic tonal unit of French is the Accentual Phrase (AP). The default phonological pattern for the AP is /LHiLH*/. The late rise (annotated as LH*) is a bitonal pitch accent and occurs at the right edge of the AP, which is also the location corresponding to the primary stressed syllable. Hence, the H* tone has a double association with the end of the AP (as it marks the end of the constituent) and with the AP-final (full) syllable (as it marks primary stress; see Jun and Fougeron, 2000, 2002). The early rise (annotated as LHi) is also part of the underlying /LHiLH*/ tonal structure and it is a bitonal edge tone (phrase accent) marking the left edge of the AP. At the phonetic level, the /LHiLH*/ can be implemented in different ways, leading to different variants of the basic AP pattern such as, e.g., LLH* (without the realization of the Hi) or LH* (without the whole LHi). Similar to the traditional accounts, Jun and Fougeron (2000, 2002) claim that H* is obligatory, and it is always realized in non-final IP position (except in specific cases, such as in tonal clash contexts). On the other hand, Hi is optional: the phonetic occurrence depends on many factors such as, e.g., speech rate, AP length, speaking style or syntactic constituency; see also Welby, 2006, Astésano et al., 2007; Michelas and D'Imperio, 2012; German and D'Imperio, 2015). Following Jun and Fougeron (2000, 2002), we will assume a distinction between obligatory H* and optional Hi in our paper. The LH* and LHi rises also differ in their phonetic properties. Phonetically, the f0 rise from L to H* in LH* is usually very prominent, and it reaches its f0 peak around the end of the associated syllable. Such a syllable is also characterized by longer duration (especially of the rhyme) and increased intensity (e.g., Pasdeloup, 1990; Welby, 2003). On the other hand, the f0 rise from L to Hi in LHi is less marked than that for H*, with the H peak for Hi being

lower than that for H*, and variably realized in the first syllables of the AP (Jun and Fougeron, 2000; Welby, 2003, 2006; German and D'Imperio, 2015).

Higher in the prosodic hierarchy, there is the Intonational Phrase (IP), which is marked by a H% or by a L% at its right edge. The boundary tones are always realized on the last syllable of the IP. Hence, when an AP is IP-final, both the H tone of the LH* rise and the boundary tones L% or H% might be realized on the same syllable, resulting in a situation of tonal crowding. Jun and Fougeron (2002) claimed that, when the boundary tone is H%, a condensed realization is expected, that is, a single f0 rise, that does not allow for distinguishing the H target of the LH* rise and the H target of the H% boundary tone; when the boundary tone is L%, the H* will be "pre-empted", that is, replaced entirely by L% (Figure 1). The IP is also signaled by non-tonal cues such as final lengthening, with the duration of the last syllable of the IP being longer than the duration of the last syllable of the AP (Jun and Fougeron, 2000).

In languages like English, nuclear accents are obligatory and prenuclear ones are optional, whereas in French, the distinction between prenuclear and nuclear contours is not straightforward (Post, 2000; Jun and Fougeron, 2002; but see Di Cristo, 1998 and D'Imperio et al., 2007 for a different account). Here, we will adopt though this terminology as for a better comparison with work on other languages. We define the 'nuclear accent' as the last LH* accent in the IP (e.g., in the Accentual Phrase in IP-final position) and 'prenuclear accents' all preceding LH* accents (i.e., in APs in IP non-final position). Given the particular realization of nuclear patterns involving tonal crowding, we define 'nuclear contour' as the section of the contour for the Accentual Phrase in IP-final position containing either a nuclear L followed by L% (with pre-empting of the nuclear H*) or a nuclear LH* followed by H%, and the 'prenuclear contour' as the stretch of the contour preceding the nuclear contour. The prenuclear contour includes both obligatory pitch accents (LH* in IP non-final APs) and optional edge tones (LHi in both final and non-final IPs). In our paper, we will focus on the nuclear boundary tones, and on the prenuclear LH* and LHi rises.

Such language-specific characteristics appear to constrain phonological imitation. Michelas and Nguyen (2011) asked French listeners to imitate relatively short APs (containing a bysillabic noun plus a monosyllabic function word), that could either include or which differed for the presence vs. absence of a Hi tone. They found that listeners reproduced Accentual Phrases more accurately when they did not include a Hi tone (i.e., when the APs contained only a final H*) than when they additionally included the Hi tone (i.e., when the APs contained only both a Hi tone and a final H*). This indicates that imitation of optional Hi is partially driven by native speakers expectations and prior knowledge (since shorter APs in French tend to be produced more frequently with a LH* pattern, while the realization of Hi is more likely in longer APs). Note that, in other languages, the obligatory vs. optional status also impacts imitation. Cole and Shattuck-Hufnagel (2011) asked their American English participants to "repeat the words and the way the utterance was said" (Cole and Shattuck-Hufnagel, 2011, p. 2). Their instructions were intended to not explicitly encourage phonetic imitation of the model speaker's voice (unlike in D'Imperio et al.'s (2014) study). Target utterances were selected from a corpus of spontaneous speech, the American English Maptask Corpus (Shattuck-Hufnagel and Veilleux, 2007). Among pitch accents, nuclear accents (i.e., the last accents in the Intonational Phrase) were more accurately reproduced than prenuclear accents (i.e., a pitch accent preceding the nuclear one). Given that, in English, nuclear accents and boundary tones are obligatory elements in the intonation contour, they might be less prone to omissions or distortions than optional elements such as prenuclear accents.

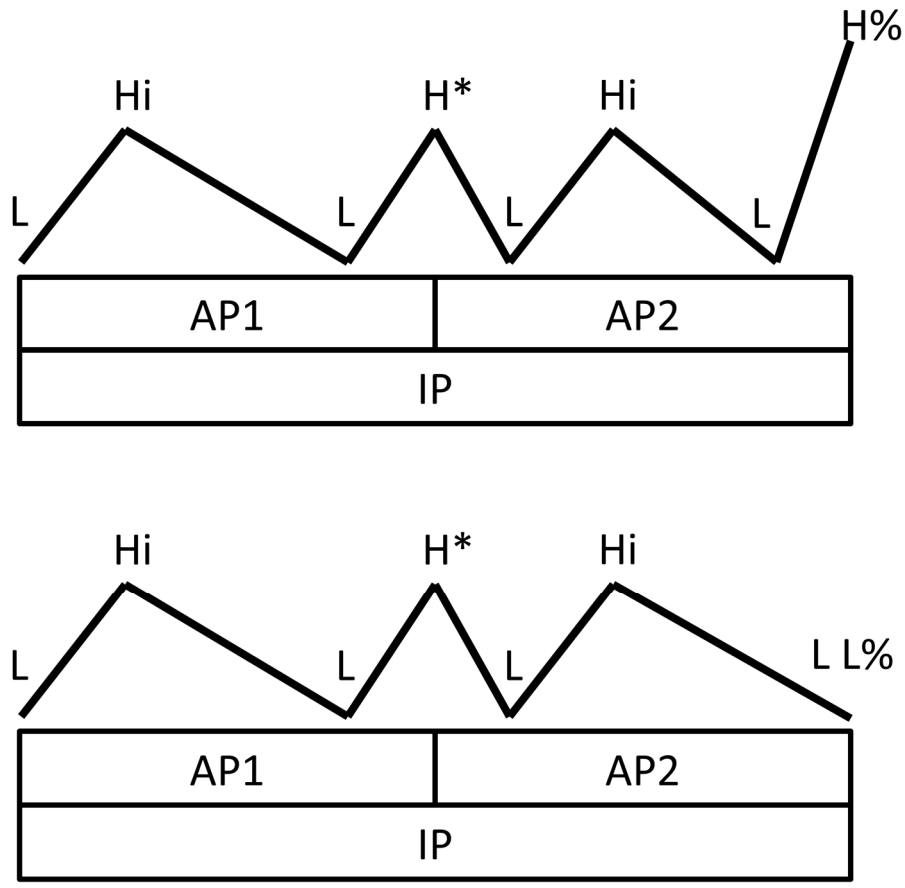


Figure 1. Schematic f0 contours of two Accentual Phrases (AP1 and AP2) illustrating an example of tonal annotation for AP in IP-non-final (AP1) and IP-final (AP2) position. AP2 can end either with a H% (top) or with a L% (bottom) boundary tone. Note that the H* in the nuclear LH* rise is condensed with the boundary tone in the case of H%, and it is deleted in the case of L%.

1.2. Research goals and hypotheses

The main goal of this paper is to investigate whether individual differences in working memory capacities affect accuracy in the imitation of phonological events and phonetic detail at the prosodic level.

In an imitation task, participants were instructed to repeat the words and the way the utterance was said (such as in Cole and Shattuck-Hufnagel, 2011). These instructions were chosen to avoid

drawing participants' attention to specific aspects of the stimuli (as is the case in explicit phonetic imitation). At the phonological level, 'accuracy' was defined as the correct imitation of phonological events which were present in the original stimulus. 'Phonological events' were detected by measuring the occurrence of phonetic targets signalling boundary tones in the nuclear contour, obligatory H* tone and optional Hi tone in the prenuclear contour. At the level of phonetic detail, we focused on durational (amount of final lengthening) and f0 details (rise slope of the H% boundary tones and of the LH* rise), representing how well participants imitate the speaker-specific acoustic implementation of the phonological events in the utterances. Note that a clear-cut division between phonological and phonetic aspects is rather schematic for the purpose of our analysis, as the division is operationalized in terms of certain acoustic-phonetic patterns that signal phonological elements (tonal targets) and other patterns that do not (e.g., fine-grained speaker specific aspects).

We predicted that participants with higher working memory would generally produce more accurate imitations. Two scenarios were possible. If speakers imitate both the phonological and the phonetic aspects of the utterance, individuals with high working memory capacities should be more accurate in phonological imitation and closer to the model speaker's phonetic implementation than individuals with low working memory capacities because of their increased skills in retaining auditory information. However, if imitation of phonetic detail is idiosyncratic or emerges only in explicit phonetic imitation, a positive correlation between working memory capacity and accuracy should only appear in the case of phonological imitation.

Within phonological imitation, we expected the degree of accuracy to vary depending on the obligatory vs. optional status of phonological events (Michelas and Nguyen, 2011; Cole and Shattuck-Hufnagel, 2011). An exploratory analysis tested whether phonological imitation for the prenuclear Hi tones is less accurate than imitation of boundary tones and of the prenuclear H* tones. These latter obligatory elements signal the prosodic structure of a sentence, and they should

thus be retained more often during imitation. In contrast, the realization of Hi depends on other (e.g., stylistic, rhythmical) factors, and their omission may be more acceptable to speakers.

The effects of working memory on imitation were compared across speaking styles, i.e., read and spontaneous speech. The distinction between read and spontaneous speech is often linked to a distinction in the degree of formality (Evans and Grabe, 1999), the read speaking style being more formal than the spontaneous speaking style, as is the case for our stimuli. In this respect, spontaneous speech is phonetically characterized by a higher articulation rate, lower f0 mean and smaller f0 range than read speech (e.g., Laan, 1997; Blaauw, 1992). We thus expected speakers to imitate read speech more easily than spontaneous speech.

2. Methods

2.1. Stimuli

The stimuli consisted of ten spontaneous and ten read speech utterances produced by the same speaker and extracted from two different corpora, the Corpus of Interactional Data (CID, Bertrand et al. 2008) and the TYPology, Adaptation, LOCalisation Corpus of French Dysarthric and Healthy speech (TYPALOC, Meunier et al., 2016).

The CID corpus consists of eight hours of spontaneous speech from 16 native speakers of (Southern) French. Each recording session consisted in a task-oriented interaction between two speakers in which they were instructed to relate a professional conflict or an unusual situation that they had been involved in. The speech was annotated at different (e.g., orthographic, phonetic, morphosyntactic, prosodic, discourse) levels. Prosodic annotation included marking Intonational and Accentual Phrase boundaries, which was carried out by an expert in French prosody (who is not among the authors of this paper; see Bertrand et al. 2008).

The TYPALOC corpus is a collection of recordings selected from different databases and aimed at comparing phonetic variations in healthy and dysarthric speech across different speech conditions.

All speakers read aloud *Le cordonnier* ('The shoemaker'), a 172-word French children's story. The corpus was provided with orthographic and syllabic transcriptions.

The twenty utterances used in our study were all produced by a female French speaker from South-West France who was in her forties at the time of the recordings. This speaker was chosen because she spoke French with a standard accent, and her speech was recorded for both corpora to facilitate a straightforward comparison between the two speech styles. The stimuli were analyzed both in terms of their phonological structure and their phonetic properties.

The target utterances were transcribed phonologically by an experienced ToBI labeler (the first author, CP) using the ToBI standard for French (Jun and Fougeron, 2000); another expert annotator checked and agreed upon the first labeler's transcriptions. The transcriptions for spontaneous speech were compared with those already existing in the CID corpus.

The twenty utterances were each four to eight syllables long. They had the same prosodic structure: each was composed of one Intonational Phrase (IP) which contained two Accentual Phrases (AP). The second AP was also IP-final, and it was characterized either by a final IP rise or by a final IP fall. In line with Jun and Fougeron (2002), we interpret the IP-final rise as resulting from a tonal crowding situation, by which the H* of nuclear LH* rise on the final AP syllable is condensed with the H% boundary tone, which is also realized on the final AP syllable. On the other hand, the final IP fall indicates a L L% sequence, with complete deletion of the H* of the nuclear LH* rise. For the sake of simplicity, we will refer to these intonational events as H% and L%. Furthermore, the first AP in each utterance contained a H* in the AP-final syllable in 19 out of 20 cases; in one read utterance, an L* was realized instead. This was due to the fact that H* can be replaced by L* as a strategy of avoiding tonal clash in the context of three consecutive H tones [Hi H* Hi] (Jun and Fougeron, 2002). The presence of Hi was more variable across utterances and APs. In read speech, Hi occurred 11 times, eight times in the first AP and three times in the second AP; in spontaneous speech, it occurred only five times in the second AP. An example of f0 tracks with prosodic

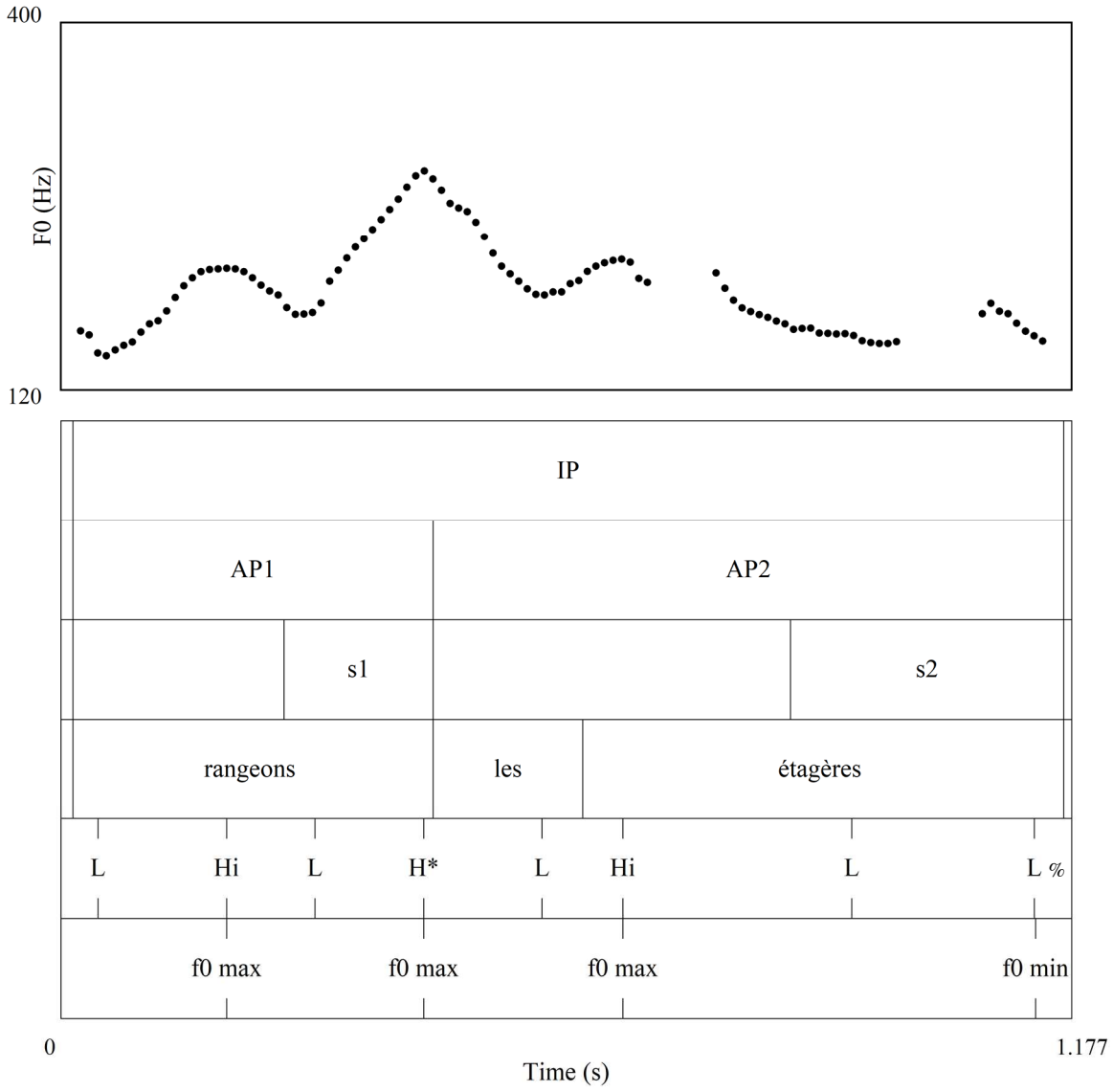
transcription is shown in Figure 2. The intonational profile of the twenty stimuli with the tones under investigation is summarized in Table 1.

Additionally, we focused on durational (amount of final lengthening) and f0 (f0 slope) details reflecting the speaker-specific implementation of the phonological patterns of the utterances. Final lengthening is a robust cue used for signaling prosodic structure, and it maximally affects the segments or syllables immediately adjacent to the prosodic boundaries (Turk and Shattuck-Hufnagel, 2007). Major prosodic boundaries lengthen more than minor prosodic boundaries, but the amount of final lengthening of a specific boundary varies widely across speakers (e.g., for American English, Byrd et al., 2006; Kim, 2019; for French, Nakata & Meynadier , 2008). Here, the prosodic structure of the stimuli was kept the same: each stimulus was composed of two APs within an IP. We thus focused on whether the imitators reproduced the speaker-specific details of the phonetic implementation of that structure. To do so, we measured the duration of the final full syllables in the first AP (s1) and in the second AP (s2) and calculated the s2/s1 ratio. We took this ratio to be an indicator of final lengthening produced by the model speaker to indicate the specific prosodic structure (D'Imperio and Michelas, 2009). In the model speaker, s2 (290 ms) was significantly longer than s1 (193 ms) [t=3.09, p=.006]. This is in line with previous studies on French (e.g., Jun and Fougeron, 2000), which showed that IP-final syllables are longer than AP-final syllables in French. The resulting mean s2/s1 ratio was not significantly different in spontaneous (mean ratio = 1.49) and read speech (mean ratio = 1.54).

The second phonetic detail under investigation is f0 slope. In the traditional AM model, tonal targets are defined along their scaling and alignment, with dynamic properties (such as the f0 slope between the targets) being considered as phonologically irrelevant f0 dimensions (see Ladd, 2008 and references therein). In particular, the f0 rise slopes for H% and for the LH* rise were calculated by subtracting the preceding f0 minimum from the f0 maximum divided by the excursion time (e.g., Welby, 2003). The model speaker's rise slope was steeper for H% (slope coefficient = 0.54 Hz/ms)

than for the LH* rise (slope coefficient = 0.44 Hz/ms), but in both cases there were no differences in read and spontaneous speech.

Additional phonetic analyses confirmed that spontaneous and read speech differed in terms of overall temporal and f0 features. Articulation rate (the number of syllable per second) was faster in spontaneous (6.6 syll/s, *SD* = 1.2) than in read speech (5.3 syll/s, *SD* = 1.2) [*t*=2.44, *p*=.025]. We focused on two overall measures of f0 register, the f0 median and the f0 range (e.g., Delooze et al., 2014). We chose the f0 median instead of the f0 mean because it provides a more robust measure against microprosodic perturbations and errors in f0 detection. f0 median was measured in Hertz, while f0 range was measured in octaves, as this takes into account speaker-specific differences more accurately (De Looze and Hirst, 2014). Median f0 was lower for spontaneous (177.9 Hz, *SD* = 10.6 Hz) than for read speech (196.7 Hz, *SD* =13.7 Hz) [*t*=-3.42, *p*=.003]. Similarly, f0 range was significantly smaller in spontaneous speech, with a mean of 0.77 o (*SD* = 0.20 o), than in read speech, with a mean of 1.08 o (*SD* = 0.25 o) [*t*=-3.04, *p*=.007]. Results for articulation rate and f0 variations confirm differences across speaking styles already found in the literature (e.g., Blaauw, 1992; Laan, 1997).


400
F0 (Hz)
120
IP
AP1
AP2
s1
s2
rangeons
les
étagères
L
Hi
L
H*
L
Hi
L
L %
f0 max
f0 max
f0 max
f0 min
0
1.177
Time (s)

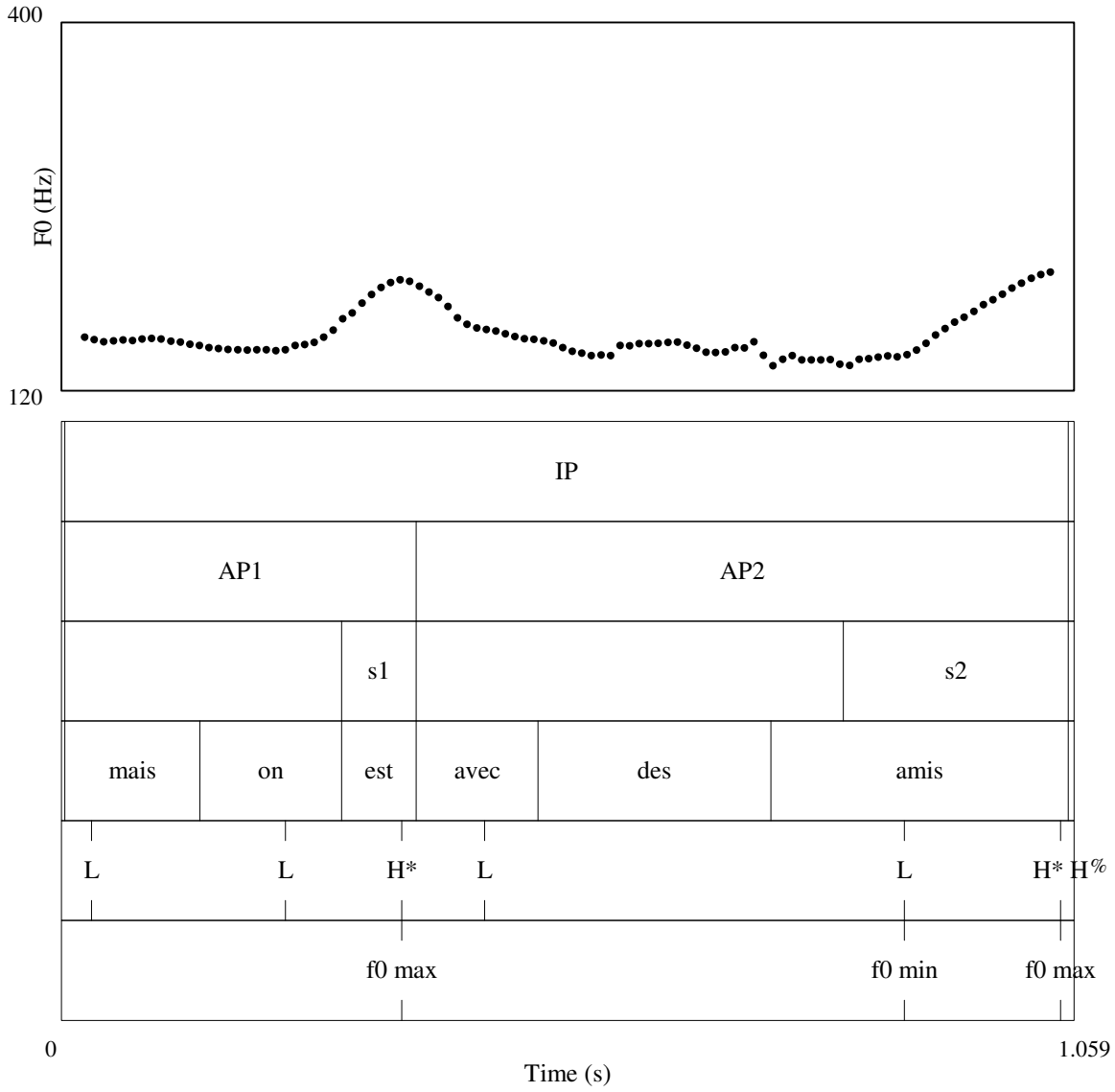


Figure 2. Examples of f0 tracks and Textgrids for the target sentences: (top) *Rangeons les étagères* ('Let's organize the shelves') and (bottom) *Mais on est avec des amis* ('But we are with some friends'), produced by the model speaker as statements. The utterances are extracted from the TYPALOC read speech corpus and from the CID spontaneous corpus, respectively. Annotation includes: IP boundaries (tier 1), AP boundaries for the first (AP1) and second (AP2) Accentual Phrases (tier 2); syllable boundaries for the AP final syllables (tier 3); orthographic transcription (tier 4); and the phonological (tier 5) and phonetic (tier 6) annotation for the f0 contour. In the first utterance, the H* of LH* rise in the second AP (AP2) is pre-empted, thus resulting in the pattern

LHiLL%. In the second utterance, the H target of the LH* rise and of the H% are condensed in a single f0 maximum.

Table 1. Number of occurrences of the tonal events under investigation in the twenty target stimuli.

Corpus	Hi	H*	H%	L%
Read speech	11	9	5	5
Spontaneous speech	5	10	5	5

2.2. Participants

Thirty-six monolingual native French speakers (31 F and 5 M, mean age 21.9 y.o., *SD* 5.28) were recruited for the experiment in the city of Aix-en-Provence, France, and surrounding area. Five of the speakers were amateur choir singers (three of whom had played a musical instrument non-professionally either in middle or high school).

2.3. Procedure

Participants were asked to perform two tasks, an imitation task and a working memory reading span task. In the imitation task, participants heard each target sentence once and repeated it after it ended. They were instructed to "repeat the words and the way the utterance was said" (Cole and Shattuck-Hufnagel, 2011). The stimuli were presented three times in blocks with a short pause between the blocks. Each block contained all the 20 stimuli presented randomly. Each subject was seated in front of a computer with a head-mounted microphone and professional-quality headphones. Recordings were made in the anechoic room at the *Laboratoire Parole and Langage* (Aix-en-Provence, France), or in quiet rooms. The imitation task lasted about 30 minutes for each participant. A total of 2160 sound files were obtained from the acoustic productions of the 36 speakers (20 stimuli X 3 times X 36 speakers). However, 71 productions were excluded (3.3% of the dataset), either because of technical errors occurring during the recording sessions, or because of

production errors (e.g., insertion of dysfluencies, mispronunciations). A total of 2089 files were acoustically analyzed.

After the imitation task, participants performed an automated version of the Working Memory reading span task (Daneman and Carpenter, 1980; French version of Swets et al., 2016). The task, which measures verbal working memory, included both a semantic verification and a word recall task on 36 items. Each item consisted of a written sentence with a semantically unrelated word presented underneath it. During semantic verification, participants silently read each sentence and judged whether it made sense or not. Half of the sentences were semantically plausible, and half were implausible. At the same time, participants were asked to remember the unrelated words. The semantic verification task was aimed at minimizing rehearsal strategies. A trial was scored "1" if the participant answered the semantic verification task correctly and remembered the word correctly; the score was "0" if the participant did not answer the semantic verification task correctly or could not remember the word. Thus, for the whole task, a participant's potential score ranged from 0 to 36. The Working Memory reading span task lasted on average 10 minutes for each participant. Based on this task, our participants averaged a working memory score of 16.3 (min = 3, max = 31).

2.4. Acoustic analysis

Each utterance produced by the imitators was processed with Praat (Boersma, 2001). The annotation (see examples in Figure 2) was conducted by a research assistant with an extensive training in phonetic segmentation and prosodic transcription with the ToBI system. Based on the model speaker, we expected each utterance to be produced as a single IP containing two APs. The expected prosodic structures were annotated by marking the boundaries at the beginning and the end of the APs and at the beginning and the end of the IP.

Our phonological analyses involved measurements of specific tonal targets of the imitators' utterances rather than a detailed phonological analysis based on ToBI transcription, which would

have been very time-consuming due to the large amount of data. The acoustic detection of tonal targets is relatively easy, and they could be taken as indicators of the phonological patterns of the utterances. Both rising and falling patterns were found at the right edge of the second AP (i.e., at the end of the IP), signaling the presence of boundary tones H% and L%, respectively. We measured the starting point (i.e., the f0 minimum) and endpoint (i.e., the f0 maximum) of the boundary tone H% as well as the utterance final L (L%). The f0 contour was often characterized by a rise at the end of the first AP, whose f0 maximum was reached in the final stressed syllable of the AP, and by a rise at the beginning of both the first and the second AP, whose f0 maximum was more variably aligned at the left edge of the AP. Based on the literature (e.g., Jun and Fougeron, 2000), the phonological analysis of the f0 rise assumes a late rise (i.e., a LH* pitch accent at the end of the first AP) and an early rise (a LHi phrase accent at the beginning of both the first and the second AP). We took the f0 maximum at the right edge of the first AP as an indicator of the presence of the H* tone of the LH* rise, and the f0 maximum at the left of each AP as an indicator of the presence of the Hi tone of the LHi rise.

Our analysis of phonetic imitation focused on durational (amount of final lengthening) and f0 (f0 slope) phonetic details. We examined final lengthening by annotating the duration of the final full syllables in the first AP (s1) and in the second AP (s2) and calculating the s2/s1 ratio. In the original stimuli, the last syllable in the second AP is also located at the end of the IP; accordingly, the s2/s1 ratio in imitators' productions was expected to be higher than 1. However, the exact value of the ratio may depend on the imitators' abilities to closely track the phonetic implementation of the model speaker. Possible extra-metrical syllables created by schwa insertions in word-final position (as typical of a stronger Southern French accent) were excluded from the annotation (Selkirk, 1977; D'Imperio et al., 2015).

A f0 rise slope for H% and for the LH* rise was calculated by subtracting the preceding f0 minimum from the f0 maximum divided by the excursion time. As the presence of the Hi tones was

much more variable in the original stimuli and few stimuli presented a Hi particularly in the second AP, we did not conduct detailed phonetic analyses of the f0 slopes connecting the two targets of the LHi rises, or of the f0 slope connecting the Hi rise to the L% (resulting in a utterance-final f0 fall).

Finally, we also measured overall phonetic patterns, such as articulation rate, f0 median and f0 range over each imitated token; we expected these patterns to vary depending on whether imitators reproduced stimuli from read vs. spontaneous speech.

2.5. Statistics.

All statistical tests were performed in R (R Core Team, 2018, v. 3.5.0). A series of mixed models was run on both the phonological and the phonetic variables of the imitations of prosodic events. We focused on the utterance-final f0 maximum as an indicator of the presence of the boundary tone H% and on the utterance-final f0 minimum for the boundary tone L%. We also examined the utterance non-final f0 maxima as indicators of the H* tone of the LH* rises in the first AP and of the Hi tone of the LHi rises in the first and second APs. In the original stimuli, the occurrence of the Hi tone in spontaneous speech is scarce and strongly unbalanced across APs. Hence, the statistical results for the imitation of Hi are preliminary, and they will be reported only for read speech.

For the phonological variables, logit models with mixed effects were run separately on the production scores of (1) the boundary tones at the end of the second AP (H% and L%); (2) the H* tone at the end of the first AP; (3) the Hi tone in both the first and the second APs. The production scores were expressed in numbers representing the correct imitation of a specific intonational event which was present in the original stimulus. Productions were considered incorrect when an intonational event present in the original stimulus was omitted by the imitators. Correct imitations were assigned the value of "1" and incorrect imitations the value of "0". The logit models are based on binomial distributions (z-scores, Generalized Linear Model, GLM) which can be used to model binary variables such as the presence ("1") vs. absence ("0") of an intonational event (Baayen,

2008). Linear mixed models were run separately on the s2/s1 ratio and on the f0 rise slopes for the phonetic variables.

The logit and linear models were applied to test the effects of the factors WORKING MEMORY (WM, 3-31) and CORPUS (read vs. spontaneous) on production scores, s2/s1 ratio and f0 slopes. In our corpus, WORKING MEMORY ranged from a score of 3 to a score of 31 (no participant obtained the lowest or the highest possible values, i.e., 0 and 36). This factor was entered as a numerical variable for each dependent variable and centered at the mean value for statistical analysis. For the analysis of the production scores of boundary tones, we also included a third fixed factor, BOUNDARY TONE (L% vs. H%), in the mixed models. This addition aimed at assessing possible differences in the accuracy of imitation across boundary tone type. We included POSITION (in the first vs. the second AP) as a third fixed factor for the analysis of the production scores of the Hi tone, which allowed us to test for possible differences in the accuracy of imitation depending on whether the Hi tones in the original read stimuli were located at the left edge of the first or the second AP.

SPEAKER (1-43), ITEM (1-20) and REPETITION (1-3) were included as random intercepts. The factor ITEM corresponded to the 20 original stimuli. REPETITION was considered as a random term since it was not controlled in a systematic way during the experimental session (e.g., stimuli were presented at irregular temporal intervals). We started the statistical analysis by fitting each model with all three intercepts, and by including by-speaker random slopes for working memory and corpus. By-subject random slopes for boundary type and position slopes were added separately for models on production scores of boundary tones and Hi tones. Backward elimination based on likelihood-ratio tests was used to decide which components should be retained in the models (Pinheiro and Bates, 2000). Likelihood-ratio tests were run comparing full models (e.g., which contained a random component) with simpler ones (e.g., without that component). For production scores, we report the *p*-value of the logit models. For the s2/s1 ratio and for the f0 slopes, *p*-values

were obtained through the *LmerTest* package. To better understand possible interactions (e.g., effects of WORKING MEMORY across CORPUS), we run the models twice changing the reference level (intercept) for CORPUS. Thus, the cut-off point for significance was set at 0.025 [p = 0.05 divided by the number of models (2) run]. Full model outputs are given in the Appendix.

3. Results

In following section, we describe the results of the imitation task for obligatory intonational events (boundary tones and H* tones), both at the phonological level (correct imitation of events which were present in the original stimuli) and at the phonetic level (final lengthening and f0 rise slopes for phonetic details; articulation rate, f0 range and f0 mean for overall phonetic aspects). This allowed us to test whether possible sources of variation in imitation across individuals are present at a phonological or phonetic level. A secondary aim was to explore the accuracy with which speakers imitate different elements of the phonological structure of a tune, such as the distinction between obligatory and optional intonational events. Hence, a preliminary analysis of the phonological imitation of the (optional) Hi tones is also reported.

Phonological imitation

3.1.1. Boundary tones

The contrast between L% and H% was significant [$\beta = 2.17$, $SE = 0.45$, $t = 4.78$, $p < .001$]. The L% boundary tone was missing more often than the H% boundary tone from locations where it was originally present (mean L% production = 82.2%, mean H% production = 97.6 %; see Figure 3). In addition, the degree of L% imitation accuracy progressively increased with increasing WORKING MEMORY [$\beta = 0.18$, $SE = 0.03$, $t = 5.18$, $p < .001$]. The relationship between working memory score and accuracy in L% imitation was slightly stronger for stimuli extracted from spontaneous than from read speech [$\beta = 0.11$, $SE = 0.03$, $t = 3.78$, $p < .001$]. On the other hand, the effects of

WORKING MEMORY and CORPUS as well as their interaction were not significant for H% imitation.

Figure 3 shows the accuracy of speaker imitations across the two different boundary tones. Figure 4 shows the relationship between working memory scores and L% productions (the data are split by corpus). Although the by-speaker plot shows a lot of variability, it appears that, with regard to L% imitation, speakers with high working memory tended to be more accurate than speakers with low working memory. For read speech, at the lowest (WM = 3) and highest (WM = 31) WM scores, the mean correct imitation scores were 33% and 100%, respectively. On the other hand, imitation scores were very similar for H% regardless of WM; at the lowest (WM = 3) and highest (WM = 31) WM scores, the mean correct imitation scores were 100 and 93.3%, respectively.

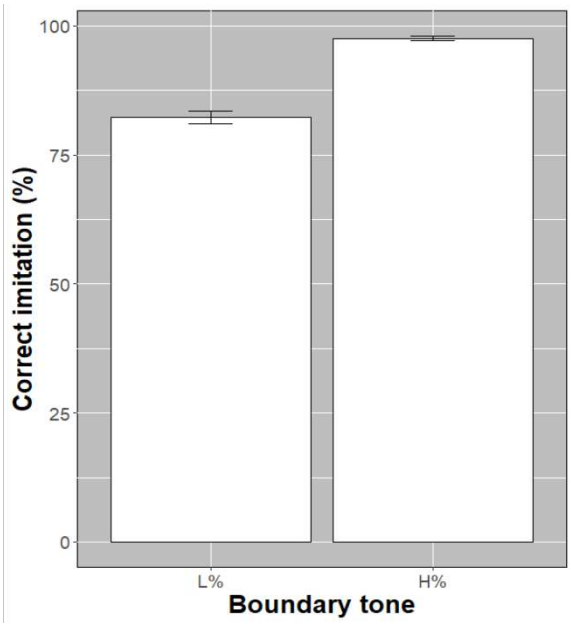


Figure 3. Mean scores of correct imitation (%) of the boundary tones (y-axis) for each boundary tone type (x-axis). Error bars represent the standard error of the mean.

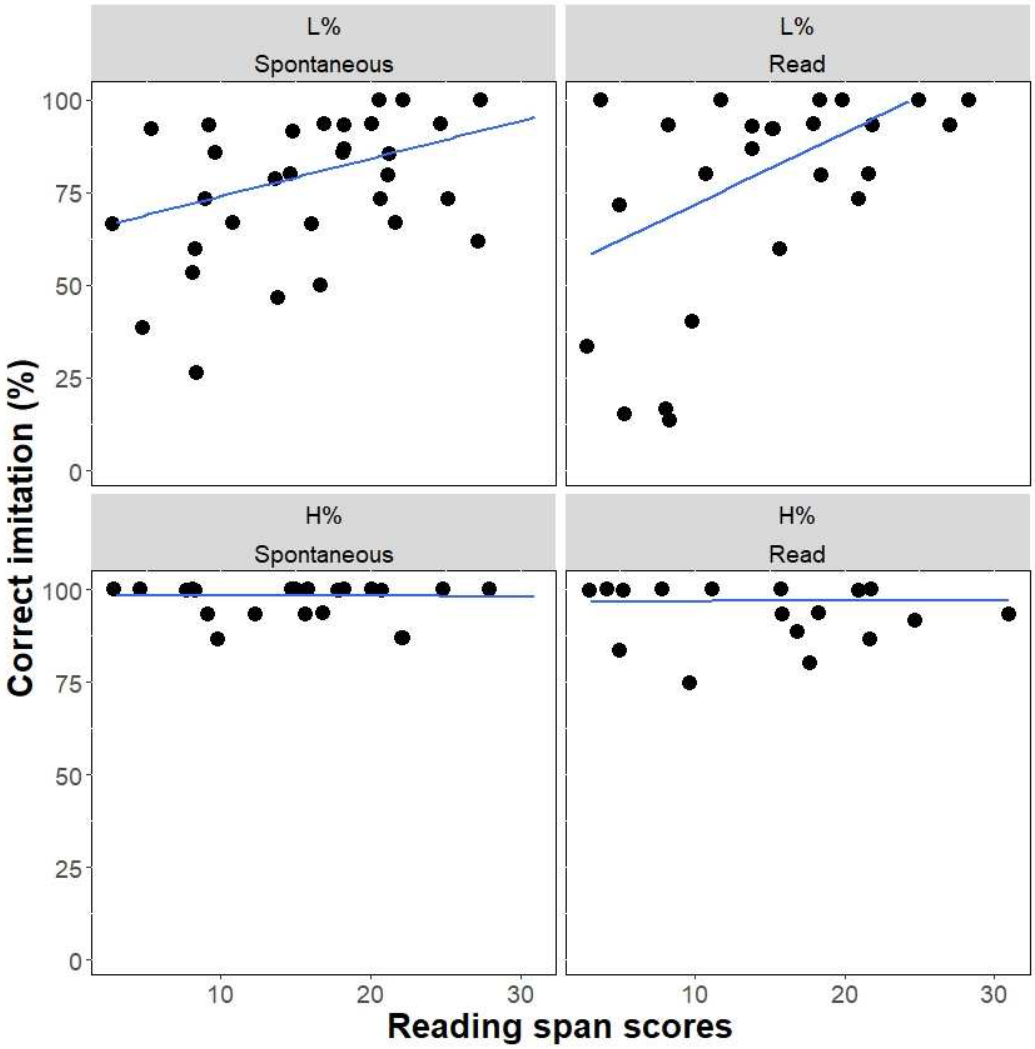


Figure 4. Relationship between correct imitation and working memory scores for each speaker, for the L% (left) and the H% (right) boundary tones. Data are shown separately for the imitation of stimuli from the spontaneous and read speech corpora. The regression line is superposed on each panel of the scatterplot.

3.1.2. Obligatory H tones*

The effects of WORKING MEMORY on the imitation of H* tones of the prenuclear LH* rises for the first AP were not significant, neither in read nor in spontaneous speech. Indeed, the statistical analysis revealed that CORPUS is the only significant effect [β = -1.87, *SE* = 0.18, *t* = -10.23, *p* < .001]. This is illustrated in Figure 5, where the average percentage of correct productions is

92.5% and 79% for the imitation of stimuli from the spontaneous and read speech corpora, respectively.

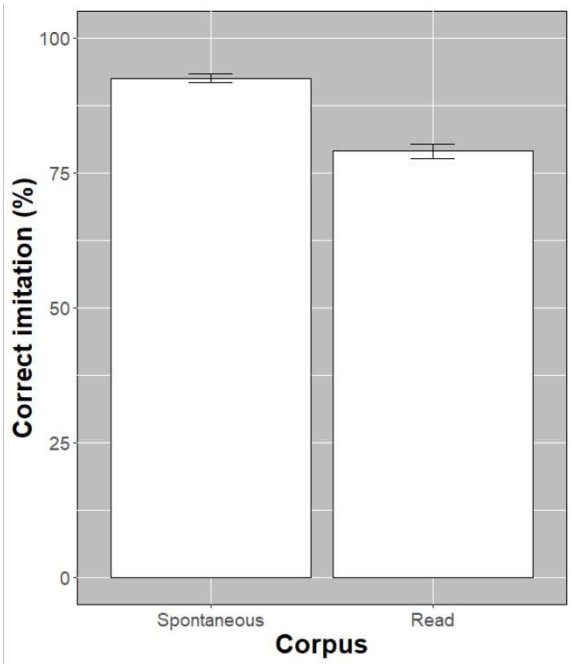


Figure 5. Mean scores of correct imitation (%) of the H* tone in the first AP (y-axis) for the imitation of original stimuli from the spontaneous and read speech corpora, shown separately (x-axis). Error bars represent the standard error of the mean.

3.1.3 Optional Hi tones

Correct imitation of the Hi tones depended on POSITION [β = 2.07, *SE* = 0.19, z = 10.76, p < 0.001]. That is, the Hi tone in the first AP occurred in 86.1% of imitators' productions, while in the second AP it occurred in only 56.2% of their productions (Figure 6). Moreover, the effect of WORKING MEMORY was significant for the imitation of the Hi tone in the second AP [β = 0.06, *SE* = 0.02, z = 3.04, p = 0.002] but not in the first AP (p > 0.05). Accordingly, Figure 7 shows that, despite the huge inter-speaker variability, the degree in imitation accuracy increased with increasing WORKING MEMORY for the Hi in the second AP.

We also qualitatively explored the number of insertions of Hi tones in locations where the target utterances lacked them. The number of insertions of optional Hi tones was very low both for read (16.7%) and spontaneous speech (8.6%; percentages are averaged between AP positions, i.e., regardless of whether the insertion was found in AP1 or AP2).

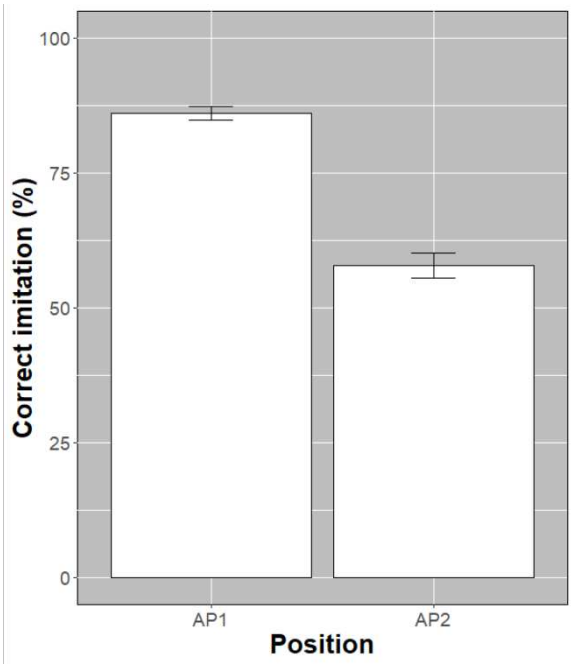


Figure 6. Mean scores of correct imitation (%) of the Hi tone (y-axis) for the first (AP1) and the second (AP2) as shown separately (x-axis) in read speech. Error bars represent the standard error of the mean.

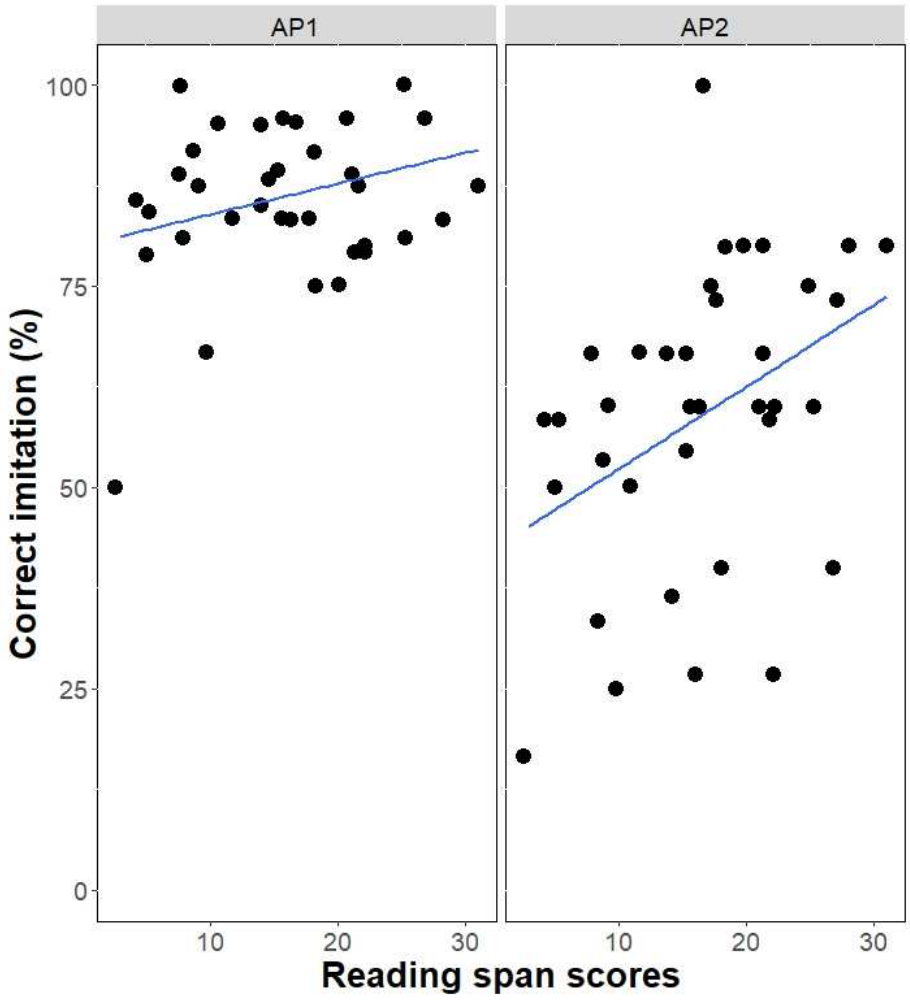


Figure 7. Relationship between correct imitation and working memory scores for LHi rises in the first (left panel) and in the second (right panel) AP. The regression line is superposed on each panel of the scatterplot.

3.1.Phonetic imitation

3.1.1. Amount of final lengthening

Imitation of final lengthening (expressed in terms of the s2/s1 duration ratio) was unaffected by WORKING MEMORY, CORPUS and their interaction. The s2 syllable was naturally longer than the s1 syllable since it occurs in IP-final position, resulting in an s2/s1 ratio over 1. The s2/s1 ratio was on average 1.78 for imitations of stimuli from spontaneous speech (mean s2 duration = 270 ms, mean s1 duration = 173 ms) and 1.66 for imitations of stimuli from read speech (mean s2 duration = 327 ms, mean s1 duration = 200 ms). Note that for the target stimuli, the s2/s1 ratio was 1.49 for

spontaneous speech and 1.54 for read speech. The imitations thus showed an overshoot relative to the model speaker, especially for spontaneous speech.

3.1.2. F0 rise slopes

The effect of WORKING MEMORY was not significant, neither in read nor in spontaneous speech. No significant differences were found in the f0 slope of H% across CORPUS, similar to the model speaker. However, the mean slope coefficient was 0.41 Hz/ms, which is much shallower than the slope for H% in the model speaker (0.54 Hz/ms).

As for LH*, the effect of WORKING MEMORY was again not significant for read nor for spontaneous speech. There was an effect of CORPUS, as the f0 rise slope was shallower when imitating stimuli from spontaneous (slope coefficient = 0.27 Hz/ms) than read speech (slope coefficient = 0.33 Hz/ms) [β = -0.33, SE = 0.01, t = -2.56, p = .013]. Note that this contrasts with the model speaker, who produced no difference in the f0 slope across read and spontaneous speech. Furthermore, both coefficients indicate that the f0 slopes for LH* were much shallower than those produced by the model speaker (0.44 Hz/ms). We note the uninteresting effect that the regression line for the relationship between WORKING MEMORY and f0 slope was flatter (with a slope coefficient around zero) when imitating read speech compared to spontaneous speech [β = -0.005, SE = 0.001 t = -3.38, p = .001]

3.1.3. Overall phonetic aspects

There was no effect of WORKING MEMORY for articulation rate, f0 median and f0 range, neither in read nor in spontaneous speech. However, a significant effect of CORPUS was found for the three variables. Articulation rate was significantly faster for spontaneous (6.2 syll/sec) than for read (5.23 syll/sec) speech [β = -0.94, SE = 0.04, t = -20.05, p < .001]. f0 median and f0 range were respectively lower [β = -6.93, SE = 0.66, t = -10.38, p < .001] and smaller [β = -1.05, SE = 0.24, t =

-4.28, $p < .001$] in spontaneous (f0 median = 193 Hz; f0 range = 9.54 o) than in read speech (f0 median = 200 Hz; f0 range = 10.44 o). More generally, differences in the f0 median across CORPUS were much smaller in all the imitators (7 Hz) than in the model speaker (18.8 Hz). On the other hand, differences in f0 range across CORPUS (0.90) were larger compared to the model speaker (0.31 o). Finally, we only note that people with high working memory capacities had slightly higher f0 median in spontaneous than in read speech [$\beta = 0.21$, $SE = 0.08$, $t = -2.51$, $p = .015$].

4. Discussion.

The present results suggest that obligatory phonological events (boundary tones in the nuclear contour and H* tones in the prenuclear one) are more accurately reproduced than optional phonological events (Hi tones in the prenuclear contour) Crucially, phonological imitation is affected by individual differences in working memory capacity. Speakers with high working memory capacity reproduced phonological events such as the L% boundary tone and the Hi tones in the second AP more often than speakers with low working memory capacity. The effect of working memory for L% was further modulated by the speaking style employed in the target utterances, i.e., whether they were extracted from read or spontaneous speech. No relationship was found between working memory and imitation of some acoustic detail (degree of final lengthening and f0 slopes). However, imitation of some speaker-specific phonetic aspects (articulation rate, f0 median and range) varied depending on whether participants were listening to spontaneous or read speech.

The results for the phonological and phonetic imitation are schematized in Tables 2 and 3.

Table 2. Schematized results for phonological variables. “WM” = working memory, “+” = significant effect, “-” = not significant effect; “na” = not available.

	L%	H%	H*	Hi_AP1	Hi_AP2

WM	+	-	-	-	+
CORPUS	-	-	+	na	na
WM:CORPUS	+	-	-	na	na

Table 3. Schematized results for phonetic variables (amount of FL = amount of final lengthening; f0 RS = f0 rise slope; AR = articulation rate). "WM" = working memory, "+" = significant effect, "-" = not significant effect; "na" = not available.

	Amount of FL	f0 RS (H%)	f0 RS (LH*)	AR	f0 median	f0 range
WM	-	-	-	-	-	-
CORPUS	-	-	+	+	+	+
WM:CORPUS	-	-	-	-	-	-

Our study found that phonological imitation of boundary tones was modulated by the tonal specification of the tones. While H% was almost always reproduced irrespective of working memory differences, speakers failed to imitate L% more often. Crucially, the percentage of correct imitations of the L% tone progressively increased with working memory score. We think that the change from L% (in the model speaker) to H% (in the imitators) is due to the use of a continuation rise (or '*continuation majeure*' in French; Delattre, 1966). Continuation rises have been debated as for their functional and formal status in French (e.g., Delattre, 1966; Rossi, 1999; Marandin et al., 2004; Delais-Rousserie, 2005, Portes et al., 2005; 2007). Portes et al. (2007) proposed that continuation rises link together "different chunk of text that could otherwise function as separated utterances" (p. 6) and they may be used as "a turn-holding cue" (p. 6). Along these lines, speakers could have produced continuation rises as to signal that their production task was not completed yet. This could lead to the production of a H% at the end of the intonational phrases (even when L%

was used by the model speaker). Another possibility is that the H% is due to the use of list intonation. That is, the task could be perceived as a list of intonational phrases to be imitated, signaled with an H% at the end of non-final list items. However, this explanation seems unlikely in our study. In French, rising contours in lists are a subclass of continuation rises with shallower f0 slope and more mid-plateau-like f0 shape than standard continuation rises (e.g., Jun and Fougeron, 2000; Portes et al., 2007). From informal acoustic and perceptual observations, we believe that the rising contour produced in our study does not correspond to list intonation, rather it is more similar to the standard continuation contour. Furthermore, each stimulus was presented in isolation, and repeated trials of stimuli were presented in random order, which minimized the possibility of perceiving the stimuli as a list. Working memory also modulated the accuracy of imitation of optional Hi tones, but not of obligatory phonological events, such as H% and H*. As was the case for L%, working memory scores correlated positively with the correct imitation of Hi tones in the second AP. The Hi tone in the second Accentual Phrase was in IP-medial position. IP-medial material, unlike IP-initial material such as in the first AP, may be more difficult to remember as it does not benefit from primacy effects on memory (Postman and Philips, 1965), and higher working memory capacities may be needed to correctly reproduce non-obligatory items in this position.

Our finding that French obligatory boundary tones and H* tones were better imitated than optional Hi tones reflects results from Cole and Shattuck-Hufnagel (2011) for American English showing that (obligatory) boundary tones and nuclear accents were more successfully reproduced than (optional) prenuclear accents. Obligatory elements such as boundary tones and the H* tone in French signal the prosodic and syntactic structure of the sentences, as well as word stress (given that in French primary stress is realized at the AP level), so deleting them can impact the meaning of the sentence. On the other hand, the function of optional LHi rises in French is less clear, and their realization depends on many different factors, such as speech rate, speaking style or phrase length (e.g., Jun and Fougeron, 2000, 2002). In languages like English, it has been suggested that

imitation accuracy is higher for the nuclear contour than for the prenuclear one (Cole and Shattuck-Hufnagel, 2011). However, this distinction cannot be applied to French, where we found that imitation accuracy was higher for obligatory H* in prenuclear position than for optional Hi tones in the same prenuclear position. Remember that the prenuclear contour includes both obligatory pitch accents (LH* in IP non-final Accentual Phrases) and optional edge tones (LHi in both final and non-final IPs). Hence, we argue that the criterion guiding imitation accuracy in French is the distinction between the obligatory vs. optional status of phonological events. Note also that our test sentences varied, e.g., in phonological length, with APs being minimally containing two syllables (see Figure 1). It is possible that the imitation of optional Hi is partially driven by top-down factors, such as native speakers' expectations concerning the realization of the early rise in short vs. long phrases (e.g., shorter APs in French are preferentially produced with a LH* pattern, while Hi rises are more likely to be realized in longer APs; see Welby, 2006; Michelas and Nguyen, 2011).

Phonological imitation was also modulated by speaking style in our data. L% was more accurately imitated in spontaneous speech, especially by individuals with low working memory. Similarly, the H* tones were more accurately reproduced for spontaneous speech than for read speech, suggesting that speakers reproduced a similar prosodic structure to that of the model speaker (as the location of the H* tone always corresponds to an AP boundary). The read speech stimuli displayed richer intonational structure possibly, as a function of factors of variability such as AP length or speech rate. In fact, they contained a much larger number of Hi tones, indicating a higher number of occurrences of initial rises (in eight out of ten sentences). The number of Hi tones was lower in the spontaneous speech stimuli, however. It is an interesting question as to whether such factors of variability and/or the richer intonational structure (as resulting from the higher number of initial rises at the beginning of the Accentual Phrases) could have affected correct imitation of H* tones or L% at the end of the first Accentual Phrase. Note also that the number of Hi insertions, though very low, mirrors the differences in speaking style found for Hi imitation accuracy, as speakers tended to

insert more Hi tones when imitating read speech.Our results on phonetic imitation show that phonetic details in prosody were only partially reproduced. In particular, speakers showed an overshoot for final lengthening when imitating spontaneous speech, and a marked undershoot for f0 slopes (both for H% and LH*) relative to the model speaker. Interestingly, in the case of LH* rises, speakers produced f0 slopes differently across speaking styles, in spite of the fact that no differences were present in the original stimuli. Our data highlighted overall phonetic patterns similar to previous findings: speakers imitated speech rate differences across speaking styles produced by the model speaker, as their productions were faster for spontaneous than for read speech (Jacewicz et al., 2010). Furthermore, they were reproducing overall aspects of the f0 contour, such as f0 median and range. However, differences in imitations of read and spontaneous speech were much smaller than those found for the model speaker. Our results on f0 are in line with, e.g., Babel and Boulatov (2012) that, in a single-word shadowing task, listeners imitated f0 mean, but the effect of imitation was rather limited in size.

The limited accuracy in phonetic imitation can be interpreted in light of the debate concerning the role of phonetic detail in phonological representations. According to one view, phonological representations and phonetic detail at the prosodic level are strongly related; phonetic detail is said to be encoded in memory and part of phonological representations (Post et al., 2007; Cangemi, 2014; D'Imperio et al., 2014). This view is consistent with exemplar-based approaches (Goldinger, 1998; Pardo, 2006) which posit that stored exemplars contain detailed phonetic information of every perceived spoken item, including phonetic detail at the prosodic level (D'Imperio et al., 2014). This perspective is also compatible with more hybrid approaches, claiming that abstract phonological categories (like AM phonological categories) may be enriched with additional phonetic information (Cangemi, 2014). According to another view, speech perception includes a normalization process in which idiosyncratic elements of how a word or sentence is produced by the speaker are filtered out in order to extract phonological features (i.e., Gaskell and Marslen-

Wilson, 1996). A phonological representation (including at the prosodic level) thus only contains contrastive features which distinguish it from other phonological representations (Cangemi, 2014). If we follow the latter approach, our result that phonetic details were only partially imitated could be explained by the fact that the phonetic implementation of prosody varies widely across speakers and/or because such details are not linguistically contrastive. Hence, a speaker is capable of perceiving and imitating the phonology of prosody produced by another speaker, but the first speaker implements the phonological categories and prosodic structure in an idiosyncratic way. In other words, in this view, phonological representations of prosody are encoded separately from the phonetic cues that signal them (Cole and Shattuck-Hufnagel, 2011).

A problem with this explanation is that it does not seem to generalize across different findings in the literature. At a first sight, our data seems to support a separate encoding of phonetics and phonology. However, our data contrast with previous research showing that imitation of phonetic details such as within-category differences in tonal alignment and f0 slope can be highly accurate (e.g., German, 2012; D'Imperio et al., 2014; Petrone et al., 2017). This line of research supports the idea that phonetic detail is part of the phonological representation of intonation (Post et al., 2007; Cangemi, 2014; D'Imperio et al., 2014). Our differing results may be due to another possibility: that phonetic imitation is modulated by the specific instructions employed in a task (see also D'Imperio et al. 2014 for a similar explanation). Prosodic studies reporting imitation of phonetic details often use explicit instructions for phonetic imitation. For instance, D'Imperio et al. (2014) asked their participants to imitate the model speaker's *pronunciation*, focusing their attention on the phonetic characteristics of the stimuli.

In our study, as in Cole and Shattuck-Hufnagel (2011), we asked our participants to "repeat" the model speaker's utterances, rather than explicitly asking for phonetic imitation. Our instructions may have led speakers to focus more on structurally relevant aspects (syntax, lexicon and phonology) rather than on the phonetic implementation of prosody produced by the model speaker.

Hence, contrary to the segmental level (e.g., Delaney et al., 2010), the imitation of phonetic detail at the prosodic level may only reliably emerge with explicit imitation instructions. Prosodic categories are conveyed by multiple cues, such as f0, intensity, duration, voice quality or even segmental properties (e.g., Welby and Niebuhr, 2019), just to mention a few. Even for the same cue, there are individual differences in the way people weight specific dimensions. For instance, f0 tonal alignment and f0 slope can be produced in an idiosyncratic way to signal a pitch accent; e.g., some speakers mark intonational contrasts with stronger tonal alignment differences, while others show weaker alignment differences enhanced by differences in f0 shape (Niebuhr et al., 2011). Similarly, in prosody perception, listeners can select or weight various cues to prosody in an individual way (Cangemi et al., 2015; Roy et al., 2017). Given the redundancy of phonetic cues and cue dimensions as well as the complex trading-relations they may enter into, it is possible that listeners just did not care to imitate a particular cue as changes on this cue could be offset by changes in another cue. In follow-up studies, it would be interesting to investigate to what extent phonetic imitation of phonological categories is influenced by individual cue-preferences, by looking at the way participants perceive and produce phonological categories under a variety of phonetic cue manipulations within the same categories.

Task instructions may have also played a role in explaining differences in working memory capacity in phonological imitation. In the context of our experiment, the reading span scores used to measure working memory reflected the imitators' ability to actively maintain goal-relevant information in memory while performing another concurrent task (Daneman and Carpenter, 1980). If our task instructions led the speakers to focus more on structural elements, individuals with high working memory would have been more capable of focusing and retaining phonological events compared to individuals with low working memory. Hence, individuals with high working memory capacity would be more capable of retaining task-relevant information and inhibiting irrelevant information (e.g., Conway et al., 2005). Further investigations could clarify how effects of working

memory on phonological vs. phonetic imitation depend on task instructions by directly comparing the impact of implicit vs. explicit instructions for phonetic imitation. Another line of research could investigate the relationship between working memory, selective attention and inhibition (e.g., via dichotic listening tasks), and how these different cognitive components interact in the process of prosodic imitation (Lewandowski, 2012, Lewandowski and Jilka, 2019). Finally, in our paper we used phonetic measurements to establish the presence of phonological events (e.g., a H target for H* and Hi tones) as this method was relatively fast given the large amount of acoustic data. It is important in the future to investigate the effects of working memory on phonological imitation through a combination of detailed prosodic transcription and phonetic measurements.

In sum, the results of our study help to reconcile results in the literature concerning speakers' abilities to imitate prosody by taking into account sources of individual variability in phonological and phonetic imitation. Our study is the first (to our knowledge) to show a link between working memory and the phonological imitation of prosody. Additional linguistic factors modulate this link, such as the obligatory vs. optional status of the phonological events and their tonal specification. More limited accuracy in the imitation of phonetic detail, on the other hand, may be the result of a task-specific strategy to ignore speaker-specific, irrelevant information. Taken together, these results suggest that prosodic imitation is mediated by speaker-specific cognitive factors, along with constraints of the native phonological system as well as by situational (e.g., task-specific) factors.

Acknowledgements

This research was supported by a grant from the *Agence Nationale de la Recherche* (ANR) to Caterina Petrone for the project 'Representation and Planning of Prosody' (ANR-14-CE30-0005-01) and by a grant from the European Union Seventh Framework Program (FP7/2007-2013; FP7-PEOPLE-2012-IEF, grant agreement n◦ 327586) to Simone Falk. An Erasmus Plus grant from the University of Naples allowed Daria D'Alessandro to relocate to the LPL for part of the duration of this project. The research was further supported by grant ANR-16-CONV-0002 (ILCB), and the

Excellence Initiative of Aix-Marseille University (A*MIDEX). We would like to thank the Editor, Marianne Pouplier, and three anonymous reviewers for their valuable comments on earlier drafts of this paper. Thanks also to Roxane Bertrand, Amandine Michelas and Cristel Portes for useful discussions and feedbacks. We thank Baptistine Marcel for her help with recordings, Elena Maslow and Marie-Charlotte Cuartero for their help with data processing, and Daniel Hirst for his precious help with PRAAT. We also thank the Leibniz-ZAS, Berlin, where C. Petrone is an associated scholar. The first author dedicates this work to Giovanni Petrone.

References

Astésano, C. (2001). *Rythme et Accentuation en Français : Invariance et Variabilité stylistique*. Paris: L'Harmattan. ISBN: 2-7475-0603-7

Astésano, C., Bard, E. G., and Turk, A. (2007). Structural influences on initial accent placement in French. *Language and Speech*, *50*(3), 423446. doi: 10.1177/00238309070500030501

Baayen, H. *(*2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press. ISBN: 978-0521709187

Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. (Doctoral dissertation, University of California, Berkeley). Available at: https://escholarship.org/uc/item/1mb4n1mv

Babel, M. and Boulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231–248. doi: 10.1177/0023830911417695

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56*(11), 851–864. doi: 10.1037/0003-066X.56.11.851

Baddeley, A. and Hitch, G.J. (1974) Working memory. In: *Recent Advances in Learning and Motivation* (Vol. VIII) (Bower, G., ed), pp. 47–89, Academic Press. doi: 10.1016/S0079-7421(08)60452-1

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. 2008. Le CID-Corpus of Interactional Data-Annotation et exploitation multimodale de parole conversationnelle. *Traitement automatique des langues*, *ATALA 49*(3), 1–30. Available at: https://hal.archives-ouvertes.fr/hal-00349893

Blaauw, E. (1992). Phonetic differences between read and spontaneous speech. In *Second international conference on spoken language processing*, 751–754. Available at: https://www.isca-speech.org/archive/archive_papers/icslp_1992/i92_0751.pdf

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345. Available at: http://hdl.handle.net/11245/1.200596

Braun B., Kochanski G., Grabe E., and Rosner B.S. (2006). Evidence for attractors in English intonation. *Journal of the Acoustical Society of America*, *119*, 4006–4015. doi: 10.1121/1.2195267

Byrd, D., Krivokapić, J., and Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–1599. doi: 10.1121/1.2217135

Cangemi, F. (2014). *Prosodic detail in Neapolitan Italian*. Berlin: Language Science Press. ISBN: 978-3-944675-01-5

Cason, N., Marmursztejn, M., D'Imperio, M., and Schön, D. (2019). Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages. *Language and Speech*, *63*(1), 149–165. doi: 10r.1177/00238309198263344

Christiner, M., and Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, *4*, 874. doi: 10.3389/fpsyg.2013.0087

Christiner, M., and Reiterer, S. M. (2018). Early influence of musical abilities and working memory on speech imitation abilities*:* Study with pre-school children. *Brain Sciences*, *8*(9), 169 doi: 10.3390/brainsci8090169

Cole, J., and Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? *Proceedings of Interspeech*, Florence, Italy, 969–972. Available at: https://www.isca-speech.org/archive/archive_papers/interspeech_2011/i11_0969.pdf

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, *12*(5), 769–786. doi: 10.3758/BF03196772

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163–191.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185. doi: 10.1017/S0140525X01003922

Cowan, N., Morey, C. C., Chen, Z., Gilchrist, A. L., and Saults, J. S. (2008). Theory and measurement of working memory capacity limits. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (p. 49–104). Elsevier Academic Press. doi: 10.1016/S0079-7421(08)00002-9

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. doi: 10.1016/S0022-5371(80)90312-6

Delais-Roussarie, E. (2005). *Phonologie et grammaire : études et modélisation des interfaces prosodiques*. Thesis for « Habilitation à Diriger des Recherches ». Université de Toulouse Le Mirail.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., ... & Sichel-Bazin, R. (2015). Intonational Phonology of French: Developing a ToBI system for French. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 63-100). Oxford, UK: Oxford University Press.

Delaney, M., Savji, S., and Babel, M. (2010). An acoustic and auditory comparison of implicit and explicit phonetic imitation. *Canadian Acoustics*, *38*(3), 132–133. Available at: https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2280

Delattre, P. (1966). Les Dix Intonations de base du français. *The French Review*, *40*(1), 1–14. Available at: http://www.jstor.org/stable/385000

De Looze, C. and Hirst, D. (2014) The OMe (Octave-Median) scale: a natural scale for speech melody. *Proceedings of Speech Prosody*, Dublin, Ireland, 910–914. doi: 10.21437/SpeechProsody.2014-170.

De Looze C., Scherer S., Vaughan B., and Campbell N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, *58*, 11–34. doi: 10.1016/j.specom.2013.10.002

Di Cristo, A. (1998). Intonation in French. In Hirst, D. and Di Cristo, A. (Ed.) *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge*:* Cambridge University Press.

Di Cristo, A. (1999). Vers une modelisation de l'accentuation du francais: Première partie. *Journal of French Language Studies, 9*, 143–180. doi: 10.1017/S0959269500004671

Di Cristo, A. (2000). Vers une modélisation de l'accentuation en français. Deuxième partie: Le modèle. *Journal of French Language Studies*, *10*, 27–44. doi. 10.1017/S0959269500000120

Di Cristo, A. and Hirst, D. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 9–24.

Di Cristo, A. (2016). *Les musiques du français parlé*, Berlin, Boston: De Gruyter. doi: 10.1515/9783110479645

D'Imperio, M., Bertrand, R., Di Cristo, A., and Portes, C. (2007). Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents? In J. Camacho, V. Deprez, N. Flores, and L. Sanchez (Eds.), *Selected Papers from the 36th Linguistic Symposium on Romance Languages* (pp. 97–110). Amsterdam/Philadelphia: John Benjamins Publishing Company. Available at: https://hal.archives-ouvertes.fr/hal-00265188

D'Imperio, M. and Michelas, A. (2009). Interface entre structure syntaxique et structure prosodique : le syntagme intermédiaire en français. *Actes IDP 2009: Interfaces discours et prosodie*, 145–156. Available at: http://makino.linguist.jussieu.fr/idp09/docs/IDP_actes/Articles/D%27imperio.pdf

D'Imperio, M., and Michelas, A. (2010). Embedded register levels and prosodic phrasing in French. *Proceedings of Speech Prosody*, Chicago, IL, US. Available at: https://hal.archives-ouvertes.fr/hal-00463076/

D'Imperio, M., Cavone, R., and Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, *5*, 1229. doi: 10.3389/fpsyg.2014.01226

D'Imperio M., and German, J. (2015) Phonetic detail and the role of exposure in dialect imitation. *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, United Kingdom. Available at: https://hal.archives-ouvertes.fr/hal-01191881/

D'Imperio, M., Petrone, C., and Graux-Czachor, C. (2015). The influence of metrical constraints on direct dialect imitation across French varieties. *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, United Kingdom. Available at: https://hal.archives-ouvertes.fr/hal-01191875

Engle, R. W., and Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. 44 (p. 145–199). Elsevier Science. ISBN: 9780125433440

Evans, B., and Grabe, E. (1999). Connected speech processes in intonation. *Proceedings of the International Conference of Phonetic Sciences*, San Francisco, CA, US, 33–36. Available at: www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0033.pdf

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., and Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science, 23*(8), 931–939. doi: 10.1177/0956797612436816

Garnier, L., Baqué, L., Dagnac, A., and Astésano, C. (2016). Perceptual investigation of prosodic phrasing in French. *Proceedings of Speech Prosody*, Boston, United States. Avalaible at: https://hal.archives-ouvertes.fr/hal-01330866/document

Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. doi: 10.1111/j.1756-8765.2009.01020.

Garrod, S., Tosi, A., and Pickering, M. J. (2018). Alignment during interaction. In S. Rueschemeyer and M. Gareth Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (2nd ed.). doi: 10.1093/oxfordhb/9780198786825.013.24

Gaskell M. G., Marslen-Wilson W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 144–158. doi: 10.1037/0096-1523.22.1.144

German, J. S. (2012). Dialect adaptation and two dimensions of tune. *Proceedings of Speech Prosody,* 430–433. Available at: https://hal.archives-ouvertes.fr/hal-01510666/document

German, J. S., and D'Imperio, M. (2015). The status of the initial rise as a marker of focus in French. *Language and Speech*, *59*(2), 165–195. doi: 10.1177/0023830915583082

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251. doi: 10.1037/0033-295X.105.2.251

Hartsuiker, R. J., and P. N. Barkhuysen. 2006. Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes 21*, 181–204. doi: 10.1080/01690960400002117

Hirst, D., and Di Cristo, A. (1996). Y at-il des unités tonales en français. *Proceedings of XXIèmes Journées d'Etude sur la Parole*, Avignon, France, 223–226.

Jacewicz, E., Fox, R. A., and Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839–850. doi: 10.1121/1.3459842

Jun, S. A., and Fougeron, C. (2000) A phonological model of French intonation. In A. Botinis (Ed.), *Intonation. Text, Speech and Language Technology*, 15. Dordrecht, GE: Springer. doi: 10.1007/978-94-011-4317-2_10

Jun, S. A., and Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1), 147–172. doi: 10.1515/prbs.2002.00

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., and Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane (Eds.) and A. Miyake and J. N. Towse (Ed.), *Variation in working memory* (p. 21–46). Oxford University Press. ISBN: 9780195168648

Kim, J. (2019). Individual differences in the production of prosodic boundaries in American English. *Proceedings of the International Conference of Phonetic Sciences*. Available at: https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_865.pdf

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43–65. doi: 10.1016/S0167-6393(97)00012-5

Ladd, D. (2008). *Intonational Phonology* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814

Lewandowski, N. (2012). *Talent in Nonnative Phonetic Convergence.* Doctoral Dissertation, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. doi: 10.18419/opus-2858

Lewandowski, N., and Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, *4*, 18. doi: 10.3389/fcomm.2019.00018

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281. doi: 10.1038/36846

Marandin, J.-M., Beyssade, C., Delais-Rousserie, E., Rialland, A. and De Fornel, M. (2004). The meaning of final contour in French. Available at: http://www.llf.cnrs.fr/fr/Marandin/, September 15th.

Meunier, C., Fougeron, C., Fredouille, C., Bigi, B., Crevier-Buchman, L., Delais-Roussarie, E., Georgeton, L., Ghio, A., Laaridh, I., Legou, T., Pillot-Loiseau C., and Pouchoulin, G. (2016). The TYPALOC Corpus: A Collection of Various Dysarthric Speech Recordings in Read and Spontaneous Styles. In *LREC*, 4658–4665. Available at: https://halshs.archives-ouvertes.fr/halshs-01401377/document

Michelas, A. and D'Imperio, M. (2010). Durational cues and prosodic phrasing in French: Evidence for the intermediate phrase. *Proceedings of International Conference on Speech Prosody*, Chicago, United-States, 100881:1-4. Available at: https://www.isca-speech.org/archive/sp2010/papers/sp10_881.pdf

Michelas, A., and Nguyen, N. (2011). Uncovering the effect of imitation on tonal patterns of FrenchAccentual Phrases. *Proceedings of Interspeech*, Florence, Italy, 973–976, 2011. Available at: https://halshs.archives-ouvertes.fr/ hal-01514863/

Michelas, A., and D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French Accentual Phrases. *Journal of Phonetics*, *40*(6), 816–829. doi: 10.1016/j.wocn.2012.08.004

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97. doi: 10.1037/0033-295X.101.2.343

Nakata, S., and Meynadier, Y. (2008). Final accent and lengthening in French. *Proceedings of Speech Prosody*, Campinas, Brazil, 567–570. Available at: https://hal.archives-ouvertes.fr/hal-00285645/

Niebuhr, O., D'Imperio, M., Fivela, B.G., and Cangemi, F. (2011). Are There 'Shapers' and 'Aligners'? Individual Differences in Signalling Pitch Accent Category. *Proceedings of the International Conference on Phonetic Sciences.* Available at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session4/Niebuhr/Niebuhr.pdf

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. doi: 10.1016/j.wocn.2010.12.007

Nguyen, N., and Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics, 53,* 46–54. doi: 10.1016/j.wocn.2015.08.004

Oberauer, K. (2013). The focus of attention in working memory – from metaphors to mechanisms. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00673

Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review. *Journal of Cognition*, 2(1), *36*. doi: 10.5334/joc.58

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. doi: 10.1121/1.2178720

Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole* (Doctoral dissertation, Aix-Marseille 1).

Petrone, C., Fuchs, S. and Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. *Proceedings of the International*

Seminar on Speech Production (ISSP), 393–400. Montréal, Canada. Available at: http://pantheon.yale.edu/~jk736/petrone_fuchs_krivokapic_ISSP_2011.pdf.

Petrone, C., Lancia, L., and Portes, C. (2017). L1 intonational categories as "perceptual attractors during L2 imitation. *7th International Conference on Speech Motor Control*, Groeningen, Netherlands. Poster available at: https://halshs.archives-ouvertes.fr/halshs-01793255/document

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329v347. doi: 10.1017/S0140525X12001495

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club. Available at: http://www.phon.ox.ac.uk/jpierrehumbert/publications/Pierrehumbert_PhD.pdf

Pinheiro, J. C., and D. M. Bates (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag. ISBN: 978-0-387-22747-4

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., and Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language and Communication Disorders*, *48*(3), 329–342. doi: 10.1111/1460-6984.12014

Portes, C., Bertrand, R., and Espesser, R. (2007). Contribution to a Grammar of Intonation in French. Form and Function of Three Rising Patterns. *Nouveaux Cahiers de Linguistique Française*, *28*, 155–162. Available at: http://clf.unige.ch/index.php/download_fi le/view/66/136

Post, B. M. B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Thesus.

Post, B., D'Imperio, M., and Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning. *Proceedings of* the *International Congress of Phonetic Sciences*, Saarbruecken, Germany, 191–196. Available at : https://hal.archives-ouvertes.fr/hal-00380692/document

Postma-Nilsenová, M., and Postma, E. (2013). Auditory perception bias in speech imitation. *Frontiers in Psychology*, *4*, 826. doi: 10.3389/fpsyg.2013.00826

Postman, L., and Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, *17*, 132–138. doi: 10.1080/17470216508416422

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/

Rossi, M. (1985). L'intonation et l'organisation de l'énoncé. *Phonetica*, *42*, 135–153. doi: 10.1159/000261744

Rossi, M. (1999). *L'intonation: Le Système du Français*. Gap: Orpheus. ISBN: 9782708009127

Roy, J., Cole, J. and Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1): 22, pp. 1–36. doi: 10.5334/labphon.108

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J. L., and Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *4*, 422. doi: 10.3389/fpsyg.2013.00422

Selkirk, E. (1977). The French foot: on the status of mute e. *Studies in French Linguistics, 1*(2), 141–150.

Shattuck-Hufnagel, S. and Veilleux, N.M. (2007). The robustness of acoustic landmarks in spontaneous speech. *Proceedings of the International Congress of Phonetic Sciences*, Saarbrueken, Germany, 1–4. Available at: http://www.icphs2007.de/conference/Papers/1584/1584.pdf

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64. doi: 10.1037/0096-3445.136.1.64

Swets, B., Jacovina, M. E., and Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition*, *6*(1), 12–44. doi: 10.1177/0023830911417695

Swets, B., Petrone, C., Fuchs, S., and Krivokapić, J. (2016). Variation in prosodic planning among individuals and across languages. In *Annual CUNY Conference in Sentence Processing,* Gainesville, US. Available at: https://hal.archives-ouvertes.fr/halshs-01459819/document

Turk, A., and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase final lengthening in American English words. *Journal of Phonetics*, *35*, 445–473. doi: 10.1016/j.wocn.2006.12.001

Vaissière, J. (1992). Rhythm, accentuation and final lengthening in French. In: Johan Sundberg, Lennard Nord and RolfCarlson (eds.), *Music, Language, Speech and Brain*. Houndmills:MacMillan Press. Available at: https://halshs.archives-ouvertes.fr/halshs-00363980/en/

Welby, P. (2003). The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation. (Doctoral dissertation, The Ohio State University). Available at: http://www.ling.ohio-state.edu/publications/dissertations

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, *34*(3), 343–371. doi: 10.1016/j.wocn.2005.09.001

Welby, P. and Niebuhr, O. (2019). Segmental intonation information in French fricatives. *International Congress of Phonetic Sciences*, Melbourne, Australia. Available at: https://hal.archives-ouvertes.fr/hal-02106625/file/SegmentalInformationFrenchFricatives_final.pdf

Yu, A. C. L., Abrego-Collier, C., and Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLoS ONE*, *8*(9), e74746. doi: 10.1371/journal.pone.0074746

Appendix

Example of the base code used for categorical (a) and continuous (b) dependent variables :

(a) glmer(dependent_variable ~ WM * CORPUS + (1|LISTENER) + (1|ITEM) + (1|REPETITION) + (0+ WM + CORPUS|LISTENER), family= "binomial").

(b) lmer(dependent_variable ~ WM * CORPUS + (1|LISTENER) + (1|ITEM) + (1|REPETITION) + (0+ WM + CORPUS|LISTENER)).

Legenda

"depedent_variable"stands for whichever categorical or continuous variable chosen for the analysis.
WM : Working Memory as indexed by centered reading span scores.

Note that he base code was further enriched by including boundary_tone (for boundary tone analysis), position (for Hi tones). Sex was added as fixed factor for all f0 variables to account for sex differences among participants. However, effects of sex are not discussed as not interesting for the topic of the paper.

BOUNDARY TONE

	β	SE	z	p	
(Intercept)	2.14898	0.35934	5.980	2.23e-09	***
BOUNDARYTONE	2.17431	0.45450	4.784	1.72e-06	***
WM	0.18654	0.03600	5.182	2.20e-07	***
CORPUS	-0.12944	0.36057	-0.359	0.719605	
BOUNDARYTONE:WM	-0.19576	0.04470	-4.379	1.19e-05	***
BOUNDARYTONE:CORPUS	0.37796	0.63577	0.594	0.552184	
WM: CORPUS	-0.11387	0.03007	-3.787	0.000152	***
BOUNDARYTONE: WM: CORPUS	0.10370	0.06392	1.622	0.104729	

H*TONES

	β	SE	z	p	
(Intercept)	3.655680	0.455876	8.019	1.07e-15	***
WM	-0.040115	0.026704	-1.502	0.133	
CORPUS	-1.873268	0.182967	-10.238	< 2e-16	***
WM:CORPUS	-0.001356	0.021129	-0.064	0.949	

Hi TONES

	β	SE	z	p	
(Intercept)	0.28950	0.52473	0.552	0.58114	
WM	0.06364	0.02091	3.043	0.00234	**
POSITION	2.07746	0.19306	10.761	< 2e-16	***
WM:POSITION	-0.02424	0.02188	-1.108	0.26794	

AMOUNT OF FINAL LENGTHENING

	β	SE	t	p	
(Intercept)	0.456249	0.083945	5.435	0.000152	***
WM	0.003258	0.003766	-0.865	0.392669	
CORPUS	-0.001043	0.023220	-0.045	0.964373	
WM:CORPUS	0.004249	0.003061	1.388	0.173030	

SLOPE OF H%

	β	SE	t	p	
(Intercept)	4.381e+00	4.529e-01	9.673	4.22e-08	***
WM	-5.132e-02	3.518e-02	-1.459	0.152	
CORPUS	3.696e-01	2.395e-01	1.543	0.123	
SEX	5.917e-04	1.020e+00	0.001	1.000	
cWM:CORPUS	2.268e-02	2.753e-02	0.824	0.410	
WM: SEX	1.699e-01	1.342e-01	1.266	0.264	
CORPUS: SEX	-1.429e+00	8.868e-01	-1.611	0.107	

WM: CORPUS: SEX	-6.217e-02	1.220e-01	-0.510	0.611	

SLOPE OF LH* RISE

	β	SE	t	p	
(Intercept)	0.322693	0.033967	9.500	2.59e-09	***
WM	0.005736	0.002883	1.989	0.05722	
CORPUS	-0.033520	0.013048	-2.569	0.01327	*
SEX	-0.059969	0.101465	-0.591	0.58110	
WM: CORPUS	-0.005552	0.001639	-3.388	0.00162	**
WM:SEX	0.008242	0.013785	0.598	0.57238	
CORPUS:SEX	0.018033	0.050803	0.355	0.72440	
WM:CORPUS:SEX	0.003568	0.007108	0.502	0.61858	

SPEECH RATE

	β	SE	t	p	
(Intercept)	6.242937	0.219583	28.431	3.26e-13	***
WM	-0.013551	0.010369	-1.307	0.2008	
CORPUS	-0.945871	0.047154	-20.059	< 2e-16	***
WM: CORPUS	-0.011220	0.006048	-1.855	0.0641	.

F0 MEDIAN

	β	SE	t	p	
(Intercept)	200.86455	2.77709	72.329	<2e-16	***
WM	0.19286	0.36514	0.535	0.599	
CORPUS	-6.93480	0.66795	-10.382	2e-14	***
SEX	-24.51643	31.65407	-0.775	0.487	
WM:CORPUS	0.21475	0.08557	-2.510	0.015	*
WM:SEX	5.89004	4.24723	1.387	0.245	
CORPUS:SEX	2.17028	2.76242	0.786	0.435	
WM:CORPUS:SEX	-0.22476	0.39027	-0.576	0.567	

F0 RANGE

	β	SE	t	p	

(Intercept)	10.424777	0.406291	25.658	< 2e-16	***
WM	0.010241	0.044710	0.229	0.820670	
CORPUS	-1.054419	0.245824	-4.289	0.000116	***
SEX	-0.270475	1.444036	-0.187	0.86236	
WM:CORPUS	-0.072953	0.031708	-2.301	0.027158	
WM:SEX	0.346185	0.236443	1.464	0.20446	
CORPUS:SEX	0.215439	1.014716	0.212	0.83286	
WM:CORPUS:SEX	-0.008388	0.143449	-0.058	0.95365	