THE PHONETICS OF INTONATION IN LEARNER VARIETIES OF FRENCH

by

SCOTT LEE

(Under the Direction of Keith Langston)

This dissertation presents a phonetic analysis of intonation in second language French spoken by native speakers of American English. After reviewing methodological gaps in contemporary intonation research, it presents two methods for analyzing L2 intonation, one based on mean speaker values for pitch range and span, and the other based on the Tilt model of intonation. The second method is based on a hidden Markov model (HMM)-based intonational variety classifier that offers an objective and rigorous way for evaluating the similarity of native and non-native pitch contours. Results from both analyses suggest that while learners are capable of improving certain aspects of their L2 intonation, full acquisition is unlikely, even given the application of non-traditional pedagogical methods, such as, in the case of this study, a study abroad experience. The data also suggest that phonological and phonetic transfer from the L1 is responsible for many of the learners' errors in the L2, contradicting the learner variety hypothesis (LVH), which broadly states that learner varieties of intonation are distinct from their related L1 varieties.

INDEX WORDS: Intonation, Second Language Acquisition, Phonetics, Phonology

THE PHONETICS OF INTONATION IN LEARNER VARIETIES OF FRENCH

by Scott Lee

A.B., The University of Georgia, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

THE PHONETICS OF INTONATION IN LEARNER VARIETIES OF FRENCH

by Scott Lee

Major Professor: Keith Langston

Committee: Diana L. Ranson
Timothy Gupton
Margaret E. L. Renwick

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

DEDICATION

This dissertation is dedicated to my parents and my sisters, who supported me throughout its planning and completion.

ACKNOWLEDGEMENTS

I would like to acknowledge my major professor Keith Langston, without whose patience and guidance I would not have succeeded in completing the coursework and research that were crucial in setting and keeping me on the path to finishing this dissertation. My committee members, Diana Ranson and Timothy Gupton, were a valuable source of intellectual support, and Margaret Renwick was crucial in helping me analyze my experimental results. I would also be remiss not to thank Meghan Armstrong for sharing her time and incredible depth of knowledge with me during the planning of the pilot study for this dissertation, and for helping me turn my enthusiasm for intonation into the seeds of sound scientific research. The dissertation would also have been impossible without the help of my study participants, who were willing to meet at odd times and places to be recorded for my projects, and without the generosity of Debbie Bell, who gave me a place to stay in Montpellier and is truly one of the most wonderfully selfless people I have met. Finally, I would like to thank my friends Andrew Simmonett, Holger Meyer, and Mariette Marano for their emotional support during the writing process, and for helping me keep a lighthearted attitude about what would have otherwise been a daunting, perhaps impossible, task.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This dissertation investigates the acquisition of French intonation by adult native speakers of American English (AE), by looking specifically at the domains of rises in polar questions and continuation contours. The study takes into account the phonological and phonetic factors of their acquisition through a production analysis based on two audio corpora comprising unguided and guided speech. For adult learners to successfully acquire the intonational grammar of a second language (L2), they must not only acquire its syntax, pragmatics, and intonational phonology, but they must also acquire its intonational phonetics, that is the variations in pitch alignment, duration, velocity, and range that define how its intonational categories are realized. While some pitch contours seem to be associated with similar meanings across languages (Ohala 1983, 1984; Gussenhoven 2004; Ladd 2008), many of them do not. To succeed in acquiring these contours, then, learners must not only remember specific syntactic structures to which they may attach (syntax), but they must also understand how to use them in context (pragmatics), and they must be able to discern subtle phonetic contrasts that may be distinctive in one language but phonologically irrelevant in the other. Because of this complexity, and because of the inherent variability in the phonetic realization of pitch contours among both native and non-native speakers, it has been relatively difficult to accurately measure learner progress in acquiring L2 intonation. Stemming from the fact that there is no universally-accepted methodology for comparing intonational structures across languages—even the Tones and Break Indices (ToBI) transcription system (Pitrelli, Beckman, and Hirschberg 1994), which is perhaps the most widely used, has been met with its fair share of criticism (see below for a brief discussion)—the study of intonational acquisition has been caught between something of a rock and a hard place: if pitch range is too variable to make phonetic generalizations about a particular language, but intonational categories are too language-specific to make generalizations across languages, how can researchers measure learner progress in acquiring L2 intonational systems? The goal of this dissertation is to explore ways we might be able to solve this problem, beginning with a careful examination of L2 intonational phenomena at a local level, and then concluding by exploring global models of intonation and testing their applications as measures of similarity between native and non-native varieties of intonation.

While there is a small but growing body of research on the acquisition of L2 intonation, researchers have mainly focused on the acquisition of broad phonological categories within the intonational grammar, such as interrogatives and declaratives. Research on the phonetic implementation of these contours is more scarce (see for example Grabe 1997), and only a handful of recent studies (e.g. Mennen 2004; Mennen, Chen, and Karlsson 2010) have addressed differences in pitch range and duration in learner varieties of language. Moreover, the methods for evaluating learner progress in acquiring L2 intonational phonetics in these studies have been limited to very basic measurements, such as mean fundamental frequency (F0) and duration of pitch accents (Mennen 2010). The present research extends these efforts by focusing on L2 varieties of Standard French (SF) and Meridional French (MF) to investigate how adult learners acquire the underlying tonal specifications and phonetic realizations for pitch contours in two specific domains: continuative rises, and unbiased polar (yes/no) questions. In prosodic terms, MF is perceptually distinct from SF (Coquillon 2003), but the two varieties share the same basic intonational structure, whereby top-level Intonation Phrases (IPs) are built from one or more base-level prosodic units, generally known as either Tonal Units (TUs) (DiCristo 1998) or Accentual Phrases (APs) (Jun and Fougeron 2000; 2002); as these two are theoretically similar, if not equivalent, this dissertation assumes the latter. Because French lacks lexical stress, tonal alignment is determined largely by syntactic structure, with primary stress normally falling on the last full syllable of the phrase. By contrast, English requires pitch accents to align with metrically strong syllables. For native English speakers to acquire French intonational phonology, then, they must be able to perceive and produce both components of its prosodic system, namely its post-lexical stress, fixed intonational structure, and proper tonal alignment. Moreover, because pitch range is the primary phonetic cue that distinguishes AP-final and IP-final rises (DiCristo 1998), learners must also be able to discriminate subtle changes in F0 and be able to reproduce them in the appropriate context. In line with both the Perceptual Assimilation Model (PAM) (Best, McRoberts, and Sithole 1988; Best 1993, 1995) and Speech Learning Model (SLM) (Flege 1995), the suggestion is that the perceived phonetic difference in tonal structure and scaling between the learners' L1 and L2 will bear on their abilities to acquire the L2 contours and produce them in speech. By extension, then, the expectation is that learners will have less success acquiring L2 intonational phonology and phonetics when they are similar to those in the L1, for instance in the case of the AP-final rises in French, which are very phonetically similar to the L*+H rising pitch accent in English. In order to conduct these sorts of perceptual experiments, however, which are based on models of segmental rather than suprasegmental phonology, we would need to understand the phonetic boundaries between intonational categories in each of the target languages, which we currently do not. Therefore, this dissertation will focus on exploring the phonetic aspects of learner intonation rather than the perceptual mechanisms underlying its structure, with the main goal being to develop an empirical foundation for future research seeking to integrate theoretical accounts of prosody and intonation with the prevailing experimental frameworks for second language acquisition.

The dissertation is divided into five chapters. The first two chapters are devoted to introducing the problem of intonational acquisition and reviewing what research has been done in hopes of solving it. The second chapter also outlines the normalization procedures used to process the production data, and it presents a brief argument in favor of using transformed rather than raw pitch data for conducting intonational research. The third chapter describes the production experiments used to gather L2 intonational data, and it outlines the analytical methods used to compare the phonetic realization of native and non-native tonal categories at a local level, i.e. at a level smaller than the utterance. These methods have been used for the same purpose by a number of linguists in recent years (Jun and Oh 2000; Mennen 2006; Chen and Mennen 2008), and so a secondary goal of the chapter is to provide more evidence in support of their use, and by extension the use of the Autosegmental Metrical (AM) framework (Pierrehumbert 1980; Beckman and Pierrehumbert 1986), for work on intonational acquisition. The fourth chapter discusses two novel methods for analyzing L2 intonation: a clustering algorithm that maps the acoustic properties of intonation in three-dimensional space, and a version of the Tilt Model for intonational modeling and synthesis (Taylor 1998, 1999, 2000) adapted for measuring the acoustic distance between the L1 and L2 tonal systems. Neither of these methods has been used for analyzing pitch data in L2 speech, so they are presented here more as experimental approaches to answering our research questions than as definitive answers in and of themselves. In other words, they may describe certain, hopefully important, aspects of learner intonation, but they are not taken to be cognitively realistic in the same way as formal linguistic frameworks like Optimality Theory (OT) or generative syntax. That being said, because they are based on robust statistical models, both methods are able to answer the question of whether learner varieties of intonation are phonetically distinct from native varieties, which is an important and equally relevant goal, considering the rising influence of the learner variety hypothesis on L2 acquisition research (Klein and Perdue 1996; Mennen, Chen, and Karlsson 2010). The fifth and final chapter summarizes the results of the production experiments and phonetic analysis, and it discusses several potential directions for future research on the acquisition of non-native prosody. It also recommends methods for establishing phonetic boundaries between a language's tonal categories, which will be crucial for testing the extension of current models of SLA to the realm of prosody.

1.1 Background

The main focus of this dissertation is intonation in second language acquisition. Intonation stands at the intersection of several subfields in linguistics, including phonology, phonetics, syntax, and pragmatics. It is rooted in the fundamentals of suprasegmental phonology, taking much of its structure from the prosodic structure of language, but it is unique in that it does not have the same clear form-to-meaning relationship with sound as segmental phonology (by way of morphology), where we can examine a series of segments in any given language and accurately describe the bundle of morphological, semantic, and syntactic features they represent. Instead, the rises and falls in pitch that constitute the broadly-defined phenomenon we know as intonation serve many purposes in spoken language, sometimes providing syntactic information, sometimes clarifying the discourse structure, and sometimes reflecting the speaker's own attitudes about an utterance. This ambiguity, along with the fact that it is situated at the interface of so many components of the grammar, has caused intonation to be the subject of both whimsical contemplation and scientific study since the advent of philology and the birth of modern linguistics, and has made it a particularly tough theoretical nut to crack. Even contemporary phonologists disagree about the basic formal characteristics of intonation (Ladd 2008), and there has been no definitive answer to the question of whether it is primarily linguistic or paralinguistic in nature (e.g. Gussenhoven 2002, 2004; Chen, Gussenhoven, and Rietveld 2004).

Before presenting the research questions this dissertation seeks to answer, we will review the basic structure of intonation as it has been characterized in the past thirty or so years. The following section also includes information about the theoretical assumptions underlying contemporary research on intonational phonology, and it presents some basic evidence for why those assumptions do not hold for work on intonational acquisition, which is a relatively new and underdeveloped area of research. A brief introduction to English and French intonational phonology is also included at the end of the section to provide a phonological foundation and theoretical motivation for the phonetic methods described in Chapters 3 and 4, and to illustrate the numerous reasons why language learners often struggle to acquire L2 intonation, which can to some extent be explained by comparing the tonal structures of the L1 and L2.

1.1.1 Prosody, Intonation, and Typology

Language learners face a unique challenge in acquiring the prosody of another language in that it may differ from their own not only in terms of its inventory of phonological categories, but also in terms of the scope of these categories and where they are employed within the prosodic hierarchy. Word-level prosody, for example, specifies whether lexical items have tone, as in Mandarin, pitch accent, as in Japanese, or stress, as in English and German. These prosodic properties often help speakers distinguish lexical items that are otherwise phonologically indistinguishable, like the well-known noun *ma* in Mandarin and many deverbal nouns and their corresponding verbs in English. Higher-level constituents of the Prosodic Hierarchy include the Accentual Phrase (AP), which typically includes a content word and its surrounding function words, the intermediate phrase (ip), and the Intonation Phrase (IP), which is the largest of the constituents and often contains the most prominent pitch accent in the phrase. There are also several units smaller than the word, like the mora, the syllable, and the foot, which together create a language's spoken rhythm. Although the Prosodic Hierarchy is somewhat universal in that most languages employ all of its constituents to some degree, the level of linguistic significance they assume when they are used can and does vary. In Japanese, for instance, both the mora and the word-level accent are important to prosodic structure, but in French, where rhythm and stress are determined by post-lexical processes, they are not. A good deal of work has been done on how these characteristics can be used to classify languages (Jun 2000), i.e. to develop a fully fledged prosodic typology, but the reason for mentioning them here is simply to show that prosodic systems are often structurally quite different, which can make them very difficult to acquire as part of an L2, especially considering that the phonological constituents of the hierarchy are not always marked by the same phonetic cues. Jun and Oh (2000) found evidence for this problem in L2 Korean, where the learners, who were native speakers of American English, were mostly unable to learn the grouping of tones into phrases, presumably because the latter are marked by subtle differences in the scaling of boundary tones rather than by an increase in vowel length or the duration of the interphrasal breaks. Likewise, Mennen (2006) showed that Dutch speakers of Greek struggled to produce native-like tonal alignment in phrase-medial pitch accents, even though the accents have the same underlying phonological structure. Both cases demonstrate the close relationship between phonetics and phonology in L2 intonation, and both suggest that difficulty with L2 intonational phonology may sometimes be traced back to its phonetics.

One of the main benefits of using a phonological model of intonation to study intonational acquisition is the ability to identify which tonal categories or phrase structures speakers might struggle to learn. To use Jun and Oh's study as an example, the authors, knowing that Korean has a default tonal structure but that English does not, focused on whether learners could group words into APs instead of marking them with pitch accents. To establish a similar phonological point of comparison for English and French, the target languages for this study, this dissertation adopts the Autosegmental Metrical (AM) model (Pierrehumbert 1980, Pierrehumbert and Beckman 1988, among others) of intonation, developed in the late 1970s after the advent of autosegmental phonology. A number of other models exist, like the Dutch IPO model ('t Hart, Collier, and Cohen 1990) and Dilley's (2004) syntagmatic-paradigmatic model built around tonal intervals, but the AM model is the most widely used in contemporary phonology and is well-suited for research on acquisition because of its formal simplicity and cross-linguistic adaptability. The AM model assumes that surface realizations of pitch contours are derived from underlying sequences of high (H) and low (L) tones, with mid (M) and downstepped high (!H) tones being available in some languages. The model is particularly useful for studies in acquisition since it assumes a division between two main types of underlying tonal entities: pitch accents, which are associated with prominent or metrically strong syllables, and boundary tones, which are associated with the edges of prosodic domains. Pitch accents may be monotonal, in which case the tone simply aligns with the prominent syllable, or bitonal, in which case the stressed tone (*) is either preceded by a leading tone or followed by a trailing tone. Bitonal accents comprising both an H and an L are the most common, but the other combinations (e.g. H+H and L+L) are theoretically possible and have been attested. The most perceptually salient pitch accent, typically the last one in an Intonational Phrase (IP), is called the nuclear pitch accent (Ladd 2008, Gussenhoven 2004). The combination of a nuclear pitch accent and its subsequent boundary tone is known as a nuclear configuration. These sequences are also known to be important for speakers when processing the meaning of a given utterance, and are thus focal points for cross-language studies on intonation. The notion of the 'nuclear configuration' is a useful unit of comparison when investigating the ways in which native speakers choose to communicate meaning through intonation, and has been used as a unit of comparison in a handful of studies on intonational acquisition, notably those focusing on phrase-final phenomena like the realization of question intonation in English or declarative intonation in German (Grabe 1998).

Like any model, the AM framework has its shortcomings, and they are important to bear in mind when applying it to L2 acquisition. The most noticeable issue is that every time researchers want to formalize the intonational grammar of a language that does not have an existing ToBI system, they usually need to develop a brand new system from the ground up, since the prosodic structure and tonal inventory of each language is more or less unique. At the time of writing, approximately 15 language-specific versions of the framework exist (see Jun 2005 for an overview of the systems), each with their own inventory of pitch accents, boundary tones, and alignment rules that determine the shape and functionality of the underlying intonational system and how it interacts with the other components of the grammar. Although we know that learner varieties are often simplified in terms of their structural complexity (Henriks 2005), a characteristic that applies just as well to prosody as to the other components of the grammar (Mennen 2010), their formal relationship to the target languages is not always clear. In particular, phonological ambiguities in learner intonation are often impossible to resolve on the basis of phonetic information, so that it can be very difficult to determine what portion of a speaker's intonational system belongs to the L1 and what portion to the L2.

We can see a simple example of this problem in L2 French spoken by L1 English speakers: because French makes heavy use of a single rising contour, normally as LH*, we would need to decide in advance whether to use a single label for all rises in the learner variety (which is the norm for L1 French), or to use the multiple labels found in L1 English. Gussenhoven (2005) succinctly summarizes this issue by noting that 'in the large majority of cases, the structural discreteness that is assumed in analyses of intonation systems is rooted in native speaker intuition', and that 'it is only in the more subtle cases, such as when a language appears to have two kinds of rises or two kinds of falls, that the issue becomes problematic.' Speakers at either end of the proficiency spectrum (i.e. true beginners and the very advanced) are likely to use the L1 and L2 categories respectively with enough consistency to merit the choice of one system over the other, but speakers in the middle of the spectrum typically do not, making the choice in these cases more stipulative and less grounded in empirical evidence than we would like. While this ambiguity clearly presents a methodological problem for transcribing learner intonation, it also presents a theoretical problem in that there is no clear way to extrapolate the L1 intonational categories to the L2, a problem that does not exist for segmental categories, which have clearly-defined phonetic identities in each of the L1s and can thus be straightforwardly measured in the L2 (see e.g. Flege 1995 for work on vowels and Guion et al. 2000 for work on consonants). Critics of the ToBI system and the AM framework in general (Taylor 1999; Dilley 2005; Ladd 2008; Martin 2012) have been quick to point out this shortcoming, and its supporters have been just as quick to admit it, so there is something of a consensus in the literature that although the model is not completely satisfying, it is more robust and adaptable than the competing alternatives and is thus worth using.

Another reason that conducting a purely phonological analysis of learner intonation is unsatisfying is that it does not let us measure progress in acquiring the phonetics of non-native contours, since the AM framework comprises discrete tonal categories and was not designed to model gradience in pitch. Indeed, one of the more heated debates in prosodic research is whether pitch gradience is linguistic or paralinguistic in nature (Gussenhoven 2005), and only a handful of serious efforts have been made at developing quantitative methods that could be used to fill in the AM framework's phonetic gaps. We could of course tally how often learners implement these categories in the L2, whether they use them in the appropriate semantic and pragmatic contexts, and whether they group them into phrases of appropriate composition and length, which could all be combined into a rough measurement of spoken proficiency, but the assessment would be incomplete without also examining how the categories and boundaries are realized phonetically, especially since phonetic details are among the most challenging for learners not only to produce (Flege 1995; Flege, Yemi-Komshian, and Yu 1999), but also to perceive (Fox, Flege, and Munro 1995; Iverson and Kuhl 1997).

1.1.2 English Intonation

Compared to French, which has a relatively simple intonational phonology, English has a rich inventory of tonal categories within the AM framework. In addition to the nuclear configuration, which often carries important information relating to discourse structure, information structure, and syntactic structure, pitch accents and phrase accents may be used to achieve a wide variety of syntactic, semantic, and pragmatic effects. Although it can be represented as a straightforward finite-state grammar, developed originally by Pierrehumbert (1980) and pictured below, this

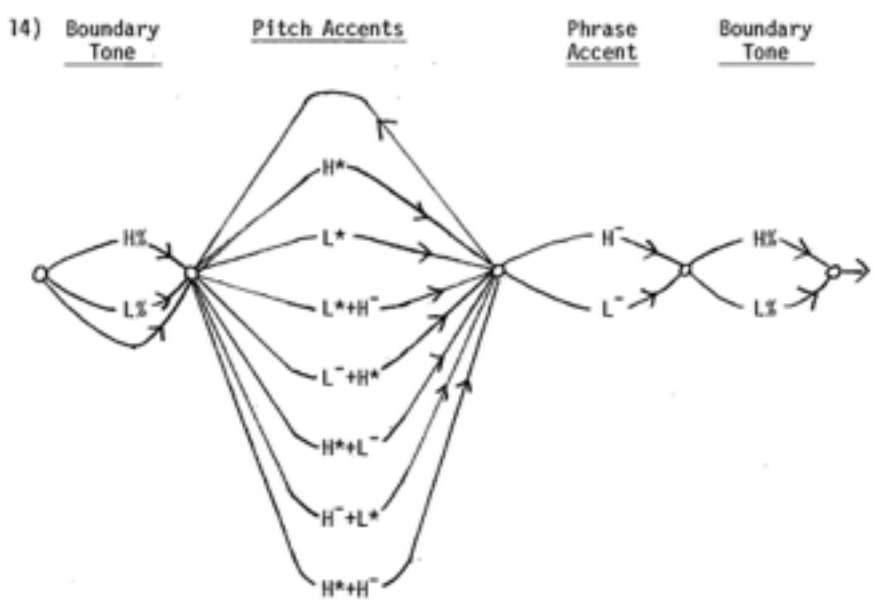


Figure 1.1. *A finite-state transition network for English intonational categories*.

system is subject to a great deal of phonetic variation depending not only on speaker mood and attitude toward the discourse material, but also on factors like speaking level and span that are grounded in physiology and may vary widely among speakers of a single language or dialect (see e.g. Grabe et al. 2000 and Grabe and Post 2002). Separate systems have been developed for American English (Silverman et al. 1992; Pitrelli, Beckman, and Hirschberg 1994) and British English, but they have much in common, including the use of H% boundary tones to signal continuation and interrogativity, and the use of L% boundary tones to signal finality and declaration. Modified versions of the British system have also been proposed for regional varieties (Grabe et al. 2004; Grabe and Post 2002), showing the framework's usefulness for studying sociolinguistic variation in prosody.

In American English, pragmatic information is primarily carried by two tonal entities in the utterance: pitch accents and the nuclear configuration, itself a combination of a pitch accent, a boundary tone, and, depending on the exact system being used, a phrase accent positioned between the two. As shown above, pitch accents come in five varieties: the monotonal accents H* and L*, and the bitonal accents H+L, L+H, and H+H. The level low L+L is omitted from the system, presumably because it would be impossible to differentiate from an L* either preceded or followed by an unaccented syllable, in which case the two syllables would be realized with the same level pitch. The pitch accents align with metrically strong syllables, and the preceding or trailing tone in bitonal accents is marked with a superscript dash ($^{-}$) to show that it is unstressed, although this is commonly left out since there is no need to distinguish a stressed+unstressed accent from a stressed+stressed accent (e.g. H*+H*), which does not exist, likely to avoid violating some basic prosodic markedness constraint like the Obligatory Contour Principle (McCarthy 1986; Yip 1988).

Although there is a great deal of variation in the system, many of the pitch accents serve relatively clear pragmatic purposes. The H* is most commonly used in declarative utterances and indicates that a word represents new information in the discourse, and the L* accent by contrast indicates givenness (Hirschberg 1994). In the L+H*, the high tone aligns with the stressed syllable in the word, and the accent often creates a contrastive effect, as in *They invited Adele,* with focus on the name *Adele,* as a response to the statement *The Rolling Stones didn't invite any Brits to their after-party*. When the L tone instead of the H tone aligns with the strong syllable, the same accent can convey a sense of uncertainty or surprise—in the example above, the speaker would be expressing some sort of doubt that Adele is British, or perhaps that the band had actually invited her to their party. The pragmatic content of other pitch accents, though, especially the extremely common H*, is less clear, and many researchers have suggested that speakers may use gradient phonetic processes, like the scaling of tonal targets, to signal information unavailable in the underlying inventory of tonal categories (Ladd 1994, 2008).

Despite this variation, this framework is useful for analyzing L2 intonation because it gives researchers a rough idea of which elements speakers may struggle to acquire when studying English. For example, the second stage of the transition network above comprises a relatively large selection of pitch accents in English, including both monotonal and bitonal pitch accents, but the same stage in French comprises just one default tonal pattern (Post 2002). Although both are iterative in the sense that speakers may remain in this stage of the network until reaching the end of the phrase, where they would need to select a phrasal accent and boundary tone (for English) or a single boundary tone (for French), the number of tonal choices they have is very different, at least on the phonological level (as mentioned above, the default rising pattern in French has numerous phonetic realizations that are influenced by speaking rate and style, but these are the result of post-lexical processes rather than the underlying tonal specification).

1.1.3 French Intonation

A number of researchers have proposed treatments of French intonational phonology within the AM framework, including Hirst and DiCristo (1984, 1986), Meertens (1987, 1993), Post (1993, 2000, 2002), and most recently Jun and Fougeron (1998, 2000, 2002). Although these treatments disagree on the levels of phrasing and their tonal representations within the IP, they agree on two important points, namely that tones are associated with stressed syllables, and that primary stress, which is purely rhythmic and thus determined post-lexically, falls on the last syllable of the phrase. Jun and Fougeron (2000) proposed that the domain of stress assignment is their Accentual Phrase (AP), which is roughly equivalent to Fonagy's (1979) *arc accentual*, DiCristo and Hirst's (1993) 'rhythmic unit', Delais-Roussarie's *groupe rhythmique* (1995), and DiCristo's (1999) 'prosodic word', among others, and consists of at least one Content Word (CW) and its associated Function Words (FWs) and clitics. In their account, prosodic phrasing in French obeys the Strict Layer Hypothesis (Selkirk 1986; Nespor and Vogel 1988), so that each IP contains one or more APs, and every AP is exhaustively contained by an IP. The default tonal structure for the AP is /LHiLH*/, with H* being an obligatory primary stress and Hi an optional secondary, or initial stress. This basic tonal configuration is variable and may appear as LH*, LLH*, LHiH*, HiLH*, and LHiL*, depending on speech rate, speech style, and the number of syllables in the phrase. Boundary tones in French may be either high (H%), level (%), or low (L%), depending on the semantic and pragmatic function of the utterance. The primary difference between this system and the one proposed for AE within the AM framework (Pierrehumbert 1980) is that the French IP does not have pitch accents; rather, it comprises a sequence of APs built on the default tonal configuration described above.

Rising intonation in French serves a variety of semantic and pragmatic purposes. As with AE, polar yes/no questions in French are often marked by a high rising contour, especially when they are unbiased (DiCristo 1998). Continuation is also marked by a rise, although its phonetic implementation depends on the strength of the following syntactic boundary, with stronger boundaries leading to longer and higher rises. Following Delattre (1966), recent research generally divides these contours into two types: major continuation (MC), and minor continuation (mc). While the mc has been considered a default melodic movement and thus not pragmatically or discursively significant (see for example Marandin et al. 2004 and Delais-Roussarie 2005), the MC is strongly associated with conversational events like turn-taking and backchannel signaling (Portes and Bertrand 2005), or listener responses in conversations that are primarily one-way (e.g. saying *Mm-hmm, Right*, or *Oui* 'yes' to affirm what someone else is saying without becoming an active participant in the dialogue). Syntactically, both contours occur at phrase boundaries, but they apply to different domains, with the mc associating with the AP, and the MC associating with the IP. Examples of the two contours are provided in Figure 1.2, which shows the pitch contour for the sentence *Globalement, il réalise pas trop l'état dans lequel il est* 'Overall, he does not realize the state he's in', taken from Portes and Bertrand (2005, p. 2). At the level of the AP, the two contours are tonally identical and consist of a final LH rise aligned

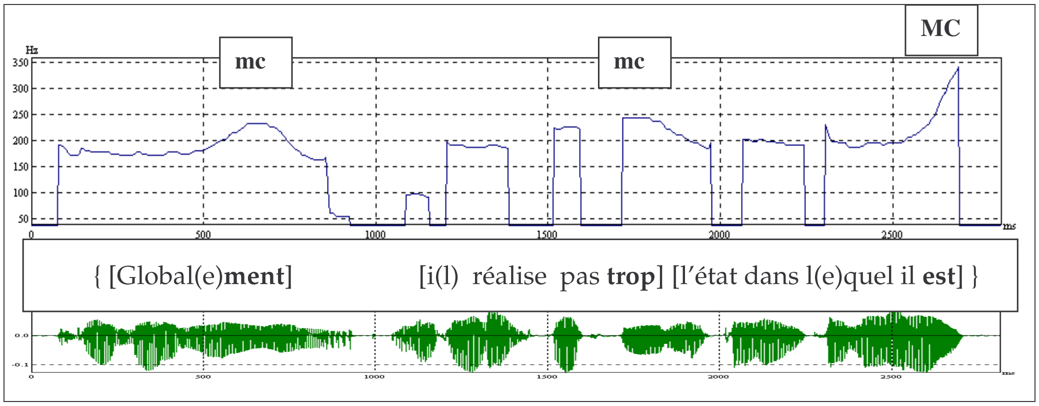


Figure 1.2: *The two types of French continuation (Portes and Bertrand 2005).*

with the last two full syllables of the phrase. Where they differ is at the level of the IP, with the MC being distinguished by a H% boundary tone (Jun and Fougeron 2000). AE native speakers (NSs) learning French must not only acquire this phonological difference between the continuation contours, but they must also acquire the phonetic distinction mentioned above between MCs and unbiased polar questions, which are separated by differences in duration and scaling.

Perhaps the most important aspect of French prosody for this dissertation is the notion of accent and its phonetic realization in the intonational domain. As outlined above, there is a strong syncretism in French between syntactic structure and the realization of pitch accents owing to the absence of lexical stress in the language. Because of this syncretism, the phrase-final *accent primaire* carries a significant amount of both linguistic and paralinguistic information, ranging from the expression of the speaker's belief state (e.g. in cases of uncertainty) to a reflection of the strength of the following syntactic boundary. The accent has a number of potential phonetic realizations, but it typically surfaces as a standard /LH*/ rise during the production of the three rising contours discussed above (mc, MC, and polar questions). Production experiments conducted by Welby (2003; 2004; 2006) indicate that the late L-elbow, or the point where the pitch contour begins its climb toward the phrase-final H*, is most often aligned with the onset of the last full syllable of the phrase, although data from both her studies and the ones presented in Chapter 2 below show that later alignment, while unlikely, is possible. In developing a ToBI system for French, Delais-Roussarie et al. (2014) did posit a second rising pitch accent similar to the scooped L*+H in English, but they suggest that it occurs only with specific types of yes-no questions and is relatively uncommon.

Post (2002, p. 3) proposed a model for French tonal structure that works in much the same way as Pierrehumbert's transition network for English. The model, shown below in Figure 1.3, begins and ends with boundary tones, which may be high, low, or level, the latter appearing only in phrase-final position. Between these bookends is a simple series of tones that are generally realized as a rise from low to high, with the variations attested by Jun and Fougeron (1998; 2000) being possible but not required, including the rare but attested string of only H tones.

The Intonation Phrase:

$$\begin{Bmatrix} \%\mathrm{L} \\ \%\mathrm{H} \end{Bmatrix} \quad (\mathrm{H}^*\ (\mathrm{L}))_0 \quad \begin{Bmatrix} \mathrm{H}^* \\ \mathrm{H}{+}\mathrm{H}^* \end{Bmatrix} \quad \begin{Bmatrix} \mathrm{L}\% \\ \mathrm{H}\% \\ 0\% \end{Bmatrix}$$

Figure 1.3: *Post's tonal structure of French.*

The latter researchers (2002, p. 4) revised this model and proposed their own (Figure 1.4 on the following page), which shares the same core structure as Post's, with the IP consisting of an AP and a following boundary tone, but rests more heavily on the rising movement, which they maintain is the default for French. The second model is the one adopted for the phonological

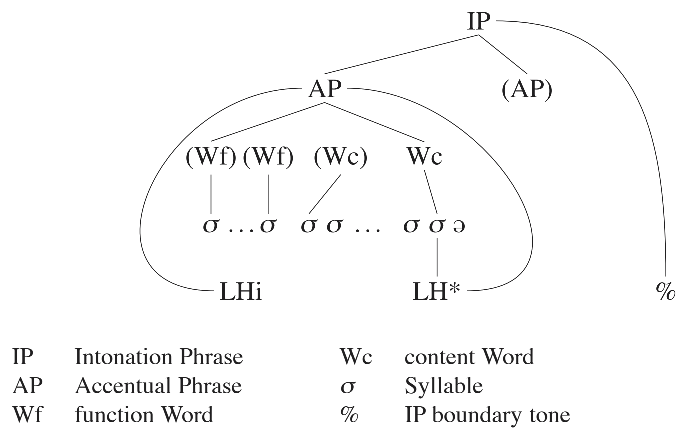


Figure 1.4: *Jun and Fougeron's structure of the French AP*.

comparisons in Chapter 3, but the differences between them are not very important, and the two are in fact combined in the newly-proposed French ToBI system (Delais-Roussarie et al. 2014), along with features from several systems proposed before the advent of AM phonology. The crucial comparison between them and the comparable model proposed for English is simply that they are structurally very different, with English deriving much of its surface variability from its large tonal inventory (i.e. its phonological richness) and French from phonetic factors like speaking rate and speech style. The problem of acquiring French intonation for native speakers of English, then, is as much phonological as it is phonetic, and should yield insight into how the process works more generally.

1.2 Problem Statement

Before addressing the specific research questions this dissertation attempts to answer, it is worth taking a moment to remember why studying intonational acquisition has been problematic, and, perhaps by extension, why it is now so fascinating. When children acquire their first language, they learn to associate these changes in pitch with changes in meaning and store them away for future use, effortlessly weaving them into their knowledge of the language. When adults try to learn a second language, though, acquiring the non-native patterns is much harder, and for a number of good reasons. First, there is strong evidence that for most people, the brain simply becomes worse at processing pitch as we get older—while children can acquire the intonational systems of multiple languages (Snow and Balog 2002) and even develop absolute pitch (Miyazaki and Ozawa 2006) (the ability to name musical notes from memory), adults generally struggle to learn even the most basic intonational patterns, sometimes needing years of exposure to their L2 to succeed at all. Second, researchers have not discovered a good way to teach intonation, so that it is often left out of the foreign language curriculum, even at the college level, where courses mostly focus on grammar, vocabulary, and culture. Third, and perhaps most importantly, the linguistic community itself does not agree on how to measure intonation or compare it across languages, and even the transcription systems we have—somewhat like musical notation, these let researchers write down what they hear—are too underspecified for doing rigorous research on acquisition. For instance, the ToBI system uses only two pitch levels to describe English intonation, which although capable of sketching the basic melodies of the language, like rising contours for questions or falling contours for statements, does not tell us what those melodies actually sound like. Classical models of intonation, like Bolinger's (1988; 1989) intonational morphemes in English or Delattre's (1966) canonical contours in French, have the opposite problem in that they are overspecified, taking whole pitch contours to be the basic building blocks of a language's intonational phonology. A few scholars (e.g. Rose 1991) have tried to solve this problem by studying how speakers scale their pitches—in other words, how high their highs go, and how low their lows—but none of the models is widely accepted (Ladd 2009), and the majority of their studies have focused on scaling in tone languages, which are much less variable in terms of pitch than intonation (i.e. non-tone) languages like English and French. To make matters worse, linguists also do not agree on how changes in pitch should be measured. Because the relationship between frequency and perceived pitch is non-linear (Moore 2012), F0 data from a single speaker should be normalized before being compared to data from others, but this is usually not the case. Many researchers still express observed values for pitch on the Hertz scale, which unlike the semitone, mel, bark, or Equivalent Rectangular Bandwidth (ERB) scales does not account for differences in perceived pitch stemming from variation in raw frequency. The basic problem with this inconsistency is that it makes it very hard for researchers to draw conclusions about variation in pitch range and level by comparing the mean values for groups of speakers, whether they belong to different speech communities (e.g. dialect groups or speakers of a particular language) or not.

These issues are multilayered and complex, but they can be summarized nicely by saying that we do not know why adults have trouble learning the intonation of a second language, and there is no clear methodology for establishing exactly what those limitations are and where they are rooted in the grammar, or indeed whether they are grammatical at all and not simply neurological, as many researchers (e.g. Johnson and Newport 1989) have suggested for the ability to learn other sound-based elements of language, a pattern referred to generally as the Critical Period Hypothesis (CPH). Because linguists do not understand these limitations, the language education community has not been able to develop a pedagogical framework for helping learners overcome them, leaving large gaps in the curriculum that prevent many if not most adult learners from achieving near-native spoken proficiency in their L2s. In the same way, these gaps keep us from being able to design assessments to accurately and objectively measure learner progress in acquiring L2 intonation.

The experiments below do not address this gap directly—rather, they have been designed to tease out the phonetic nature of learner intonation so that researchers will have firmer footing for tackling what is undoubtedly one of the great mysteries remaining in the study of the relationship between sound and language.

1.3 Research Questions

The broad question this dissertation seeks to answer is whether adult language learners can successfully acquire the intonational system of another language. This question, which is admittedly very broad, can be broken down into a small number of more focused questions, some of which we will discuss in the chapters below, but others of which are beyond the scope of this dissertation. Because language acquisition involves both production and perception, the first distinction we can make is between aspects of the acquisition process that we can measure acoustically, like the duration and scaling of tonal targets in L2 speech, and those we can only measure indirectly, like how well learners can distinguish contrastive L2 tonal categories. As mentioned above, a great deal of work has been done on both of these issues for segmental acquisition, but not much has been done for them in terms of prosodic acquisition, intonational or otherwise, so there are many related questions remaining to be answered. Because perceptual data would have been difficult to obtain from the research participants examined in this study, some of whom lived abroad for the duration of the project, the following chapters will focus on L2 production data, and, more specifically, on L2 phonetics. The first research question we will aim to address then, is whether adult learners can successfully acquire the phonetic aspects of L2 intonation.

This question raises a number of other questions, though, each of which must be addressed for our study to be successful. As mentioned above, there are serious problems with the quantitative representation of intonation that make comparing pitch contours between speakers difficult, especially for speakers of different languages. This problem is compounded by the fact that the underlying intonational phonologies of the L1 and L2 may be very different, as is the case for the target languages examined in this study, American English and Standard French. As discussed in Chapter 2, Chen (2008), Mennen (2004; 2006), Grabe (1997), and a handful of other researchers have explored several ways of solving this problem, using basic statistical techniques (e.g. t-tests and analysis of variance) to examine intergroup differences in the alignment, scaling, and duration of pitch accents and nuclear configurations. There are still significant methodological gaps we must fill before we can make reliable empirical judgments about learner progress in acquiring L2 intonation, leading to the second but perhaps most important research question this dissertation will seek to answer: to what extent can languages be distinguished based solely on their intonational phonetics? Provided that we have data from a large enough sample of speakers, and provided that a set of normalization procedures for reducing interspeaker variation in pitch to acceptable levels can be established, this question should be a reasonable one to answer. Research on tone languages like Chinese (Rose 1987; Moore and Jongman 1997; Cheng 2011; Cheng 2012) has shown both percent speaking range and z-score normalization to be promising, although they have not been tested on intonation languages.

Assuming we can distinguish two languages on the basis of their intonational phonetics, the next question we will need to answer is whether adults can acquire the phonological aspects of another language's intonational system and, if so, whether it is easier for them to acquire these intonational categories than their corresponding phonetics. There are very few studies which examine learner intonation on both levels (i.e. phonological and phonetic), but studies by Mennen (2004) and Jun and Oh (2000) suggest that learners may make more progress in learning phrasing and tonal categories than in learning scaling and alignment, and that the phonetic details they master first are those that carry important semantic or syntactic information, like interrogativity and uncertainty. These findings support work on segmental acquisition (Fox, Flege, and Munro 1995; Flege 1999; Flege 2000) that suggests fine-grained phonetic differences between L2 consonants and vowels are hard for learners to acquire, especially if they perceive them as overlapping with phonetically similar L1 categories. Because both the prosodic and the segmental systems develop very early in childhood and tend to crystallize without sufficient exposure to new languages, adult learners are likely to struggle similarly (if not equally) in acquiring L2 intonational phonetics.

The third question we will try to answer, which is more methodological than theoretical in nature, is whether the current low-level phonetic models of intonation are suitable for doing research on comparative intonational phonetics and intonational acquisition. Global phonetic models differ from local phonological models like those developed in the AM framework in that they make no stipulations about the underlying tonal structure of pitch contours. In the Tilt Model (Taylor 1998; 1999), for instance, the entire intonational system is represented by a statistical model based on acoustic features sampled from recorded speech—it makes no assumptions about the underlying tonal structure of the pitch contours, and it places heavy emphasis on the phonetic characteristics of intonation, like the size and duration of pitch excursions, an attribute that makes it particularly well-suited for tasks like speaker identification and intonation synthesis. The system is built from hidden Markov models (HMMs), a parametrical approach that takes a sequence of states or events and assigns each one a probability based solely on the value of the preceding event. So, for any given state in the sequence, the HMM predicts its most likely output, and it also predicts which state is most likely to come next. The full technical details of the Tilt Model are discussed in Chapter 4, but for now it is simply important to know that it treats intonation as a mostly surface-level phenomenon, allowing it to avoid the theoretical problems inherent to transcription systems like ToBI that are designed to provide representations of the F0 contour in terms of its underlying categories. After being trained with L1 speech, the model can also function as a classifier, estimating the likelihood of a particular sequence of intonational events being generated by a particular language.

At least in concept, approaches like the one assumed by the Tilt model should be very helpful for doing work on comparative phonetics, since the way it characterizes or classifies speakers—that is, the way it quantifies the acoustic properties of their speech—is blind to the language's inventory of tonal categories, bringing us very close to an objective evaluation of how the two sound systems relate on the surface. The drawback to this approach is that it limits our ability to analyze how speakers use pitch to express semantics, pragmatics, and other aspects of language that are located at a more abstract level of representation within the grammar, but the benefit is that it allows us to make direct comparisons between how two languages or varieties of a language sound without needing to reconcile their phonological differences, which are significant enough to have merited the development of language- and even variety-specific transcription systems (Hirst and DiCristo 1998).

Because the Tilt Model has some predetermined theoretical limitations—both the number of states and the number of outputs, for instance, must be specified before the model can be run, which lends it a certain degree of stipulation—we could also consider using clustering algorithms to build three-dimensional maps of each speaker's prosodic phonetic space, which may be aggregated and either interpreted as standalone representations of the language's intonational system or used as supplements to guide the development of the Tilt Model's HMMs. The specific algorithm would be technically demanding to implement, but the basic idea would be to record data points for change in pitch, amplitude, and duration—three of the main phonetic elements of prosody—and plot them in three-dimensional space. By applying a statistical technique designed to detect significant clusters to the collection of points, we could assign a rough quantitative identity to each speaker, and by summing the clusters for speakers in a particular group, we would then have a platform for comparing the intonational phonetics of two languages or language varieties. To date, clustering has only been applied to segmental data, and primarily to studies of vowel space (e.g. Miller 1989; Plois, Tromp, and Plomp 2005). Extending the technique to prosodic data will hopefully help us visualize the principal phonetic components (in the informal, non-statistical sense of the term) of each language's intonational system.

Atemporal models like clustering also have the benefit of presenting a single snapshot of each speaker's pitch space, which is something that sequence-based models built on HMMs cannot do. Both the Tilt Model and, to a lesser degree, intonational grammars like Pierrehumbert's finite-state transition network for English rely on sequences of events to describe what pitch contours are possible in a given language. In the Tilt Model, for instance, we might answer the question *What does English intonation sound like?* by explaining which pitch movement would be most likely to follow another (for example, a long high rise is very likely to be followed by a pause). Likewise, grounding our answer in Pierrehumbert's model, we might answer the same question by describing how utterances begin and end with H and L tones that are connected by other sequences of H and L tones, with only a few of the mathematically possible tone sequences being ungrammatical. Under a summative representation like those obtained from algorithms like clustering, however, we could compress minutes' and even hours' worth of pitch data into a single probability distribution, showing how long intonational events are likely to last, how loud they are likely to be, and how much phonetic space they tend to take up for any combination of speakers in the sample population.

This kind of model obviously has a number of drawbacks—it would be impossible to reconstruct a grammatical sequence of tones from a clustering distribution, for instance—but its benefits are significant enough to make it worth testing. A similar approach based on Gaussian mixture models (GMMs) has been frequently used to build speech recognition algorithms aimed at text-independent tasks like speaker identification and dialect identification (Reynolds 1995; Reynolds, Quatieri, and Dunn 2000) (when text is present, a HMM-based classifier is more appropriate, since it is capable of modeling a series of events and not just a summary of their phonetic properties). Clustering is similar in spirit to these approaches, but because it has already been used to study the distribution of vowels in phonetic space, it is likely more suited to examining learner intonation and the kinds of data it generates. Nonetheless, because the Tilt model lends itself well to different structural implementations and is partially linguistic in nature (the intonational events it comprises correspond to pitch accents and boundary tones), it seems a better tool for the task at hand. Clustering will likely play a role in future research on intonational phonetics, but this dissertation leaves it aside in favor of the sequential model presented in Chapter 4.

Related to both these issues is the question of whether learner varieties of intonation are phonetically distinct from those of their L1s. Early work by Jenner (1976), Backman (1979), Adams and Munro (1978), and Willems (1982) shows that learner intonation tends to have the same intonational characteristics regardless of which languages are involved, often having a reduced tonal inventory, simplified phrasing, and compressed pitch space. Mennen, Chen, and Karlsson (2010) support these findings, focusing on the distribution and inventory of pitch accents in L2 English, which they found to be similar for speakers of L1 Punjabi and Italian. All these studies use phonological rather than phonetic models of intonation to characterize the learner varieties, however, meaning the results are perhaps not as robust as we would like. Mennen (2004) is probably the closest to establishing a rigorous method for comparing speaker groups on the basis of intonation, using analysis of variance (ANOVA) to examine the effect of different L1/L2 combinations on phonetic features like pitch accent alignment and scaling, but there is undoubtedly more work to be done. Using the low-level phonetic model outlined in Chapter 4, we will try to answer the question of whether learner varieties really are distinct in terms of their intonation from both L1s involved (i.e. whether the L2 stands on its own), and if so, to what extent. In particular, the chapter presents a modified version of the Tilt model (Taylor 1998; 1999) that uses a robust statistical method to calculate the probability that a speaker's L2 intonation could have been generated by an L1. Because the model considers purely phonetic and acoustic information, it makes no assumptions about the underlying structure of the L2 intonation, and so it should provide a reasonably objective answer to the question regarding the independence of learner varieties.

The last research question this dissertation will answer is whether the quantitative methods outlined in Chapters 3 and 4 are substantial enough to be used in adapting second language acquisition models like the SLM and PAM to prosodic acquisition. These models rely on the establishment of phonological categories from gradient phonetic data, so the only way to apply them to learner varieties of intonation will be if we can demarcate the phonetic boundaries of the tonal categories posited by phonological models of intonation like the AM framework. Establishing these boundaries for the entire inventory of categories belonging to a particular language is likely not possible, since the realization of extremely common tones like the H* pitch accent is more variable in comparison to the realization of bitonal accents and particular nuclear configurations, which are typically more pragmatically and semantically restricted. Information structure (Steedman 1991a, 1991b, 2000, 2007; Büring 2007) and discourse function (Bolinger 1980; Pierrehumbert and Hirschberg 1986; Wennerstrom 1991; Grosz and Hirschberg 1992; Hirschberg, Nakatani, and Grosz 1995; Armstrong 2010) significantly impact the way these tones are realized as well, and there is no current framework for quantitatively evaluating the degree to which they affect pitch height, alignment, duration, or amplitude. Dilley (2007) addressed this shortcoming with her own dissertation research by suggesting that intonation has both paradigmatic and syntagmatic components, with prominence relations between tonemes being governed by different phonetic processes than relations between the same toneme as realized by more than one speaker. Her system is both complex and grounded in the basic principles of music theory and thus has not been widely accepted as a viable alternative to the AM framework or its competitors, but it does lay the foundation for work integrating intonation with pragmatics and discourse analysis. Still, the methodological gaps in this field are currently so large that any results we can produce from the phonetic models below should at least shed light on how much ground we have left to cover before being able to extend the leading SLA models to work on prosody.

1.4 Important Terminology

Although most of the terminology relating to the study of intonation will be familiar to those working in other subfields of linguistics, some of the more technical terms will not. There are also some cases when even intonational phonologists and phoneticians disagree about which terms are the most appropriate, so we will quickly review them here. First, the terms ‘tonal categories’ and ‘intonational categories’ are used interchangeably in the literature to refer to the phoneme-like tones or tone sequences that exist at the phonological (i.e. the underlying) level of description. In the AM framework, categories form the backbone of a language’s intonational grammar and exhaustively describe its possible tunes. Rarely, linguists will distinguish between the two terms, but typically only when comparing the prosodic systems of tone languages and intonation languages; otherwise, the two can be taken as referring to the same thing. In the chapters below, the preference is to use the term ‘tonal categories’, as it is not only shorter, but technically more accurate, since intonation encompasses many different components of the grammar, whereas tone refers specifically to the linguistic use of pitch. The only exception to this rule will be when describing studies or theories that refer to the categories using the broader ‘intonational’ label, in which case the original author’s terminology will be (temporarily) adopted.

This is the same procedure that will be used for discussing the Prosodic Hierarchy (Selkirk 1978, 1986; Nespor and Vogel 1986) and its constituents, whose names depend largely on the formal assumptions underlying the work of the author discussing them. The low-level constituents, i.e. the mora, syllable, foot, and word, are uncontroversial, but those at the higher levels of the hierarchy have been given a number of names. For example, consider the Phonological Phrase (p-phrase), which falls directly below the Intonation Phrase (IP) and is the domain of phrasal stress, and which has been treated both as a single unit and as a composition of smaller units. In syntactic literature, the component units are referred to as the Major phrase (Ma-P) and minor phrase (mi-P) (Kubozono 1987; Selkirk and Tateishi 1991), but in the intonational literature (Hirst 1998), and especially the French intonational literature (Jun 1998; Jun and Fougeron 1998, 2000; Post 2002; Delais-Roussarie et al. 2013), they are called the intermediate phrase (ip) and accentual phrase (AP). The practical difference between the terms is negligible, as they refer to the same general phenomena, but the formal difference can cause problems for doing research on intonation, which is largely concerned with the interfaces between syntax, semantics, pragmatics, and phonology (e.g. Steedman 2000). Because this dissertation deals most explicitly with the prosodic and especially the acoustic properties of intonation, the choice has been made to use the latter pair, i.e. the AP and the ip, especially since, as mentioned above, they feature prominently in the contemporary literature on French intonational phonology.

The terms ‘tonal categories’ and ‘intonational events’ are also a potential source of confusion. The former describes intonational units in the phonological tradition, proposing that the categories are the underlying, contrastive components of the surface F0 contour. The latter, which was pioneered by Taylor (1998; 1999) during his work on intonation synthesis, refers only to significant perturbations in the otherwise approximately linear F0 contour and makes no real reference to the grammar of possible tunes in a language. Tonal categories are essentially realized as intonational events, and intonational events typically correspond in meaningful and predictable ways to tonal categories, but the two are formally distinct and will be treated as such in Chapters 3 and 4.

There are also several terminological ambiguities in the literature on intonational phonetics and speech processing that should be resolved here. First, the size of pitch movements in the speech stream has been referred by both the acoustic term amplitude and the phonological/phonetic term pitch height or scaling. In Chapter 3, which examines the realization of underlying tonal categories in the L2 French, the latter terms will be used interchangeably, and in Chapter 4, which focuses more closely on the acoustic aspects of learner intonation, the former will be used exclusively. Although it would be more streamlined to use one term in both chapters, the division between them in the literature is very clear, so both will be used for the sake of consistency. This decision also applies to terms describing the length of segmental and intonational events, be they vowels, syllables, pitch accents, or larger constituents of the prosodic hierarchy. In traditional segmental phonology, the duration of vowels is commonly referred to as length, but in acoustic phonetics, it is usually just referred to as duration. Since this dissertation focuses almost exclusively on the phonetic aspects of learner intonation and *not* on its underlying phonological structure, only the latter term will be used. Finally, instead of ‘intensity’, which is heavily used in speech processing and the broader acoustic sciences to refer to the amount of pressure per unit area created by a sound wave, the term ‘loudness' is used as a more convenient, though perhaps less scientifically accurate, substitute. It is important to note, however, that amplitude may also be used to describe loudness, but only when referring to the height of the sine wave producing a sound and not its corresponding pitch contour—in the first case, we are measuring the actual loudness of the signal, but in the second, only its change in frequency. In other words, the type of signal being measured is what changes, not the denotation of the term itself. As used in Chapter 4, amplitude refers to the height of intonational events coded in F0, and never to the loudness of the signal.

Another term to clarify is the word ‘acquisition' itself. When used in the binary sense, which seems to be its most common interpretation, the term means that a speaker can consistently and with little effort achieve native-like levels of production and perception; in other words, the speaker is completely fluent in the language. In any real sense, this goal seems to be unattainable for almost all adult speakers who were not exposed to the target language in childhood (for whatever reason). By extension, any study of L2 acquisition wherein the learners do not achieve complete native proficiency in the language must necessarily claim that no acquisition took place. This kind of generalization seems both unfair to the learners and scientifically uninteresting, and so it is abandoned here for a more flexible definition of the term. As it is used in the chapters below, ‘acquisition' may apply to any part of the grammar, and to various degrees. So, learners who consistently realize the AP-final pitch movement in French as a rise may be said to have acquired a small but important portion of the language’s intonational phonology, and those who manage to consistently realize the same rises with native-like alignment may be said to have acquired some of its intonational phonetics. The distinction is also made between partial and full acquisition to reflect the fact that learners may sometimes but not always produce grammatical intonation in the L2. Segmental categories evolve slowly over time (Flege 1995), and if the same is true for tonal categories, then our concept of acquisition must be gradient in order to account for the process and recognize the learners' progress. The reasoning behind these distinctions is simply to make it easier to show which, if any, part of the French intonational system the learners do acquire, and which they do not, with the ultimate goal being to address the question of whether complete acquisition, in the binary sense of the term, might in fact be possible.

Finally, as this dissertation deals almost exclusively with intonational data, it is important to establish a clear conceptual boundary between the terms 'frequency' and 'pitch'. The former refers to the number of cycles a sound wave completes per unit time (typically seconds) and is objective in that it describes an acoustic, rather than a perceptual, phenomenon. F0 is by definition a measure of frequency and is often reported on the Hertz scale for this reason. By contrast, 'pitch' refers to the frequencies of sound waves as they are perceived by the audio circuitry in the brain, and because it is not an objective acoustic measure like frequency, it is typically reported on an adjusted scale. Specific ways of defining the quantitative relationship between frequency and perceived pitch are discussed in Chapter 2, but the important point to make here is that the two terms refer to fundamentally different, albeit related phenomena, and should be treated as such in any study dealing with intonation, which is inherently grounded in not only the acoustics of linguistic pitch, but also in the way speakers perceive them.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter presents a summary and discussion of the most relevant research to date on the acquisition of second language prosody and in particular intonation. The first section outlines the major assumptions and formal principles of the Speech Learning Model (SLM; Flege 1995), a leading model of L2 acquisition that provides a formal link between speech perception and production. The reason for presenting the model here is to provide detailed justification for the phonetic analysis in Chapter 3, which seeks to determine whether this model and others like it may someday be extended to the study of intonational acquisition. Section 2.2 provides an overview of the phonology and phonetics of intonation from a comparative perspective by presenting studies that seek to answer the equally important questions of how languages differ in their tonal structure and how they employ the same tonal structures to different effects. A good deal of research was done to this end in the mid-20th century, with linguists like Delattre (1963; et al. 1962, 1965) and Bolinger (1951) developing theories of intonation and its interface with pragmatics and semantics, which were applied cross-linguistically. The widespread availability of pitch-tracking software beginning in the 1990s saw their research both expanded on and revised by linguists who could now take a more empirical approach to the problem (Beckman and Pierrehumbert 1986; Grover, Jamieson, and Dobrovolsky 1987; Cruttenden 1997; Hirst and DiCristo 1998; Swerts, Krahmer, and Avesani 2002). Recently, these studies have been developed further, and researchers have begun to explore the fine-grained ways that languages and language varieties can be distinguished based on subtle characteristics like tonal alignment and scaling (Grabe et al. 2000; Grabe and Post 2002; Ulbricht 2002).

Section 2.3 deals specifically with research on intonational acquisition. As subfields of linguistics, both intonational acquisition and its parent field prosodic acquisition are relatively new and thus underdeveloped, but work in the two decades or so by Mennen (2004, 2006, 2010), Jun and Oh (2000), Wennerstrom (1994, 1997), Chen (2009), and others has led to some interesting conclusions and important directions for future research. The section is divided into two subsections, the first dealing with the acquisition of L1 intonation during childhood, and the second dealing with the acquisition of L2 intonation in adulthood. The reason for reviewing how we acquire intonation as children is mostly to establish a point of comparison for the particular challenges we face in acquiring it as adults, but the discussion will also serve to highlight some of the basic neurological and psychological bases for the perception and production of pitch in language and in general (e.g in music). The second subsection will provide a brief but detailed review of the most important studies in L2 intonational acquisition, focusing primarily on work by Mennen and her colleagues, who have conducted the most thorough observational and experimental studies to date.

Because these studies bring up important issues in quantitative methods for comparative phonetics, Section 2.4 summarizes methods for normalizing pitch data in speech, which are not only crucial for the phonetic analysis in the following chapters to succeed, but which have also become the subject of a thriving, if small-sided, debate in the international community. There seems to be a rift, although not necessarily intentional, in methodology between linguists working within and outside of the AM framework, with the former leaning toward relatively unaltered measurements to present their data (e.g. use of the Hertz scale and simple summary statistics to compare speakers) and the latter relying heavily on transformed data and advanced statistical methods (e.g. use of logarithmic scales and HMM-based classifiers) to process the data for computational analysis. Because this dissertation explores both methods, arguments for and against each will be discussed, and a suggestion will be made as to which method might be most appropriate for future work on intonational acquisition. The importance of this section cannot be overstated, as any comparison of phonetic data obtained from two or more speakers must be based on normalized acoustic measurements, whether they pertain to fundamental frequency, vowel formants, segmental duration, or other quantitative properties of the speech stream. Humans do not hear sound on a linear scale (see e.g. Moore and Glasberg 1983; Hermes and Van Gestel 1991; Brown, Gaskill, Carlyon, and Williams 1993; Nolan 2003), although a few researchers have suggested otherwise (notably Rietveld and Gussenhoven 1985, whose findings are discussed below in detail), and because perception and production in language are inextricably linked, we cannot form judgments about spoken material without also considering how it is received by the ear. The specific nature of this relationship is beyond the scope of this dissertation, but its general properties are outlined below in our discussion of perception's role in second language production and acquisition.

2.1 Second Language Acquisition Models

While studies like Mennen, Chen, and Karlsson (2010) have addressed the acquisition of L2 intonational phonetics, they are somewhat at odds with other areas of research in L2 phonology in that they have not addressed the main prediction of both the PAM and the SLM, namely that perceived phonetic difference affects discrimination, which in turn affects learning (Flege 1995).

Although the two models are similar, they differ primarily in that the SLM makes predictions about learning, whereas the PAM mostly makes predictions about the relationship between perception and discrimination, although it has been successfully applied to acquisition studies (Guion et al. 2000); both models predict that speakers will have trouble discriminating sounds in the L2 when they perceive them to be phonetically similar to sounds in the L1. For reasons discussed in the following subsection, not much progress has been made in adapting these models to the study of intonational acquisition, but they have been used extensively in the study of segmental acquisition and make fairly strong predictions about L1 interference and interlanguage effects. Recent research on intonational acquisition has addressed this notion, namely that L2 intonation should be considered as stemming from a learner variety and not simply a degenerate variety of the L1 (Mennen, Chen, and Karlsson 2010). Because of this suggestion, and because these models have met with such success in the segmental domain, one of the main goals of this dissertation is to provide quantitative evidence that can be used to test whether these models are also capable of accounting for the phonological and phonetic characteristics of learner intonation.

Generally speaking, complete acquisition of an L2 sound system is likely only in cases where the learners have been raised by first-generation immigrants or in other situations where the target language was spoken extensively if not exclusively in the household. To my knowledge, this generalization is not supported by studies of intonational acquisition in particular, but neither is it refuted. As Flege (1997) notes, the ability for learners to acquire non-native vowels seems to decline with age, so it would seem to be one grounded in some sort of psychological or neurological mechanism, like the language centers in the brain losing their sensitivity to (linguistically meaningful) sound. This idea has been around for some time, with a number of researchers in the second half of the 20th century (e.g. Lenneberg 1967; Scovel 1988; Patkowski 1990) proposing the existence of a critical period for language acquisition, comprising a soft boundary between puberty and adulthood beyond which the capacity to learn language rapidly declines. More recent research on L2 acquisition, however, has shown such a Critical Period Hypothesis (CPH) to be untenable in its strongest form, demonstrating, for example, that adult native speakers of Japanese are able to improve both their perception (Yamada and Tohkura, 1991; Best and Strange, 1992; Scott et al. 1994; Flege, Takagi, and Mann 1996) and their production (Yamada et al. 1994; Flege, Takagi, and Mann 1995) of English /l/ and /r/, both sounds which do not occur in their L1. Although these improvements may take years for learners to make (Flege 1997), they imply that phonological acquisition is indeed possible in adulthood, at least for non-native consonants.

Whether or not adult learners can acquire non-native vowels is less clear. Stops, fricatives, and other sounds normally classified as consonants are made by altering the airstream at a single primary point of articulation, whereas vowels are formed by shaping the tongue to alter the resonance of the sounds produced by the vocal cords; in other words, while consonants are shaped by one continuous variable, the place of articulation (manner is, for the most part, categorical), vowels are shaped by at least two, tongue height and tongue advancement (lip rounding is a third factor, but less critical than the others). Consonants of course have subtle phonetic properties like aspiration that when contrastive in one language but not in another can be very difficult to perceive and thus acquire, but in general, vowels are more subtly differentiated, especially in languages with rich vowel systems. The literature indicates that learners have the highest likelihood of success in overcoming this challenge when they are exposed to the L2 at an early age (Newport and Johnson 1987), a trend that runs against the grain of data gathered from studies on consonantal acquisition like those mentioned above, where learners continue to make progress as long as they consistently use and are exposed to the L2. A long line of research has shown the task to be difficult for adult learners, however. In examining German speakers of English, Bohn and Flege (1992) found that both experienced and inexperienced learners produced open mid-front vowels that English L1 speakers often rated as unintelligible, and McAllister, Flege, and Piske (2002) found that speakers of L2 Swedish could acquire the language's differences in vowel quantity, but only so far as they could perceive the contrasts in terms of the duration features of their L1. Regarding longitudinal acquisition, which has been a beacon of hope for applied linguists working in the past few decades, Flege (1994) maintains that any non-native category can be acquired with enough time, although Munro and Derwing (2002) found that while learners initial progress tends to level off after a few months, after which increased exposure to and training in the L2 does not lead to increased proficiency.

As mentioned above, studies based on Flege's SLM (1995) typically focused on late- rather than early-stage language learners, suggesting that perceptual limitations are responsible for the apparent detrimental effects of maturation on L2 acquisition. The model comprises a number of postulates and hypotheses, but only those which bear directly on the experiments designed for this dissertation will be discussed here. The first and perhaps most controversial postulate with respect to the CPH is that the mechanisms and processes enabling L1 acquisition, be they cognitive, psychological, neurobiological, or otherwise, remain active throughout the lifespan and are therefore also available for L2 acquisition. This idea is supported by the results of Flege and MacKay's (1995) earlier study on the perceived degree of foreign accentedness in the speech of Italian L2 English speakers living in Canada. In their study, the authors found a linear relationship between age of arrival in Canada and the strength of the speakers' foreign accent as judged by English L1 speakers. The result would seem to support the CPH, but their study included speakers who had emigrated from Italy at almost 25 years of age, well past the age when acquisition should rapidly decline. Despite this fact, the increase in accentedness was consistent throughout the range of ages studied, so that full acquisition seemed to be less likely with each passing year, rather than with the expiration of the proposed critical period.

Another important postulate of the SLM is that phonemes are stored in long-term memory as exemplars of phonetic categories, which are established in childhood for the L1 but may continue to evolve under the influence of sufficient exposure to multiple other languages. As it relates to intonational acquisition, this postulate has two important implications: first, that underlying tonal categories exist and that they alone determine the shape of the F0 contour—there has been some debate in the literature about whether this is true (see e.g. Ladd 2008: 209-210)—and second, that learners should be able to acquire new categories in adulthood, which echoes the implications of the first postulate discussed above. Crucially, because the establishment of new categories relies on speakers making generalizations from phonetic input, this postulate implicitly requires that they be able to perceive the phonetic details that differentiate categories in the L2, which may at times seem very subtle, even to the language's native speakers; this is something that limited perception work in L2 intonation does not strongly support (Best PC; So and Best 2008, 2010, 2011, 2014). It also means that while learners may be able to acquire new intonational categories, for example a bitonal pitch accent not present in their

L1, it does not necessarily mean that they can acquire those categories' corresponding phonetics. Research by Jun and Oh (2000), discussed more fully below, supports these two implications, showing that adults generally have more trouble acquiring non-native intonational phonetics than phonology, with even advanced learners failing to consistently produce proper pitch scaling in the L2.

The SLM has two important implications for intonational acquisition. First, if intonational categories exist, then it should be possible for learners to acquire new ones as adults, at least to a certain extent. The caveat to this suggestion is of course that learners can only acquire categories in the L2 that they can distinguish from their L1, so it may not be possible (or at least in any way likely) for them to acquire the full inventory of non-native sounds if there is indeed some phonetic overlap between the two systems. Even in these cases, however, the fact that categories are perceptually indistinguishable means that they are often very nearly, if not completely, identical phonetically, and so learners may not need to acquire them in order to achieve de facto near-native proficiency. Another factor affecting the limit of practical success is the perception of non-native intonation by native speakers of the learners' target language, who may not be able to perceive the statistically significant but acoustically subtle differences in how certain intonational events are realized. If we abandon the underlying assumption that acquisition is all or nothing, however, the basic tenet of the SLM—that there are no psychological, neurological, or developmental factors preventing the development of new phonological categories given sufficient exposure to the L2—then we should expect learners to make measurable, empirically-verifiable gains in acquiring L2 intonation.

Second, because each speaker's phonological space is shaped by each language they learn, then we should be able to find evidence of two-way intonational transfer in advanced learners and bilinguals who have had sufficient exposure to and experience with their L2s for their inventory of phonetic categories to have grown beyond those provided by the L1. As discussed in Section 2.4 below, there is strong evidence that this kind of two-way transfer is in fact possible in the intonational domain (Mennen 2008), indicating that the mechanisms underlying language learning remain active into adulthood, at least in certain groups of speakers (e.g. bilinguals), and that it can be strengthened by increasing exposure to the L2.

As mentioned above, this dissertation does not undertake perceptual research pertaining to intonational acquisition. Yet, models like the SLM are important to its pursuit because they present a clearly defined set of empirical goals that must be met in order to establish a link between the perception and the production of prosody, namely that intonational categories must be clearly defined in phonetic and not simply phonological terms.

2.2 Comparative Studies of Intonation

The earliest work on comparative intonational phonology is likely that of Pierre Delattre (1962; 1963; 1965; 1966), whose systematic comparison of English, German, Spanish, and French (1963) laid the methodological and formal foundation for approaches developed toward the turn of the century, including the AM framework. Delattre examined a number of core intonational features from each language, finding, for example, that both the placement and shape of final accents (i.e. nuclear configurations) signaling finality varied systematically between them, with English and German contours being convex and French contours being concave. He also studied the distribution of similar contour shapes, finding that, although continuation was signaled in all languages with some kind of rise, it was the least prominent in English, with only very good English speakers producing the contour reliably. Although the study is based entirely on observational rather than experimental data, and its methods are largely qualitative rather than quantitative, Delattre's insights into the subtleties of interlanguage prosodic variation still inform contemporary research, especially on intonation in French; his 10 melodies of French intonation (1966) are considered canonical, although they have been adapted to fit with the most recent models of French intonation in the AM framework.

Although intonational data are hard to quantify (Mennen, Schaeffler, and Docherty 2012), contemporary researchers have found evidence that languages may differ not only in their phonological inventory of tonal categories, as Delattre and others have shown, but also in the way speakers implement these categories phonetically. Comparative studies of intonation both between (Grabe 1997) and within languages (Grabe, Nolan, and Farrar 1998; Grabe, Post, Nolan, and Farrar 2000; Grabe and Post 2002) generally focus on the distribution of pitch accents and the phonetic implementation of the nuclear configuration. In her analysis of English and German phrase-final rising and falling pitch accents, Grabe (1997; 1998) found that English speakers were more likely to compress both contours under the reduction of segmental material available for voicing, whereas German speakers were more likely to compress only the rises, truncating the falls and terminating the contour before reaching the bottom of their range. Because the contours she examined have identical tonal specifications in both languages, she concluded that the primary difference between them was the way they were realized in F0, suggesting that languages can be distinguished intonationally not only by phonological, but also by phonetic information.

Work by Mennen, Schaeffler, and Docherty (2007; 2008) also supports the notion of intonational phonetics as important in defining intonational boundaries between languages. Beyond a general interest in comparative phonetics, the motivation for their particular project stemmed from evidence from both experimental studies and anecdotal reports that English speakers not only have a higher level and wider range than speakers of German, but also that (monolingual) German speakers perceive them to sound overexcited (Eckert and Laver 1994) and even aggressive (Gibbon 1998). Using a set of 25 sentences and a text, they collected production data from monolingual native speakers of both English and German and analyzed long-term distributional pitch data for both level and span. Perhaps not surprisingly, what they found was that, on average, female speakers of Standard Southern British English (SSBE) varied more in pitch span than did speakers of Northern Standard German (NSG), with span measured as the distance in semitones between the 90th and 10th percentiles of the speaker's range (i.e. the 80% quantal range). They did not find, however, that the SSBE speakers had higher overall levels than the NSG speakers; in fact, the mean difference between the two groups for the text reading task was just 6.6 Hz, which even toward the lower end of the frequency spectrum occupied by F0 is very small.

This is of course not to say that languages cannot differ in terms of level. Not much research has been done on this topic, but van Bezooijen's (1995) study of Dutch and Japanese women provided solid evidence that the distinction is possible and can indeed be made on the basis of experimental data. Like Mennen, Schaeffler, and Docherty's work (2007; 2008), van Bezooijen's research was prompted by anecdotal evidence that Japanese women intentionally raised their speaking levels 'in order to project a vocal image with feminine attributes of powerlessness' (1995:253). Her study was guided by three research questions: first, whether high pitch was indeed associated with physical and psychological powerlessness in the Dutch and Japanese cultures; second, whether there was a stronger differentiation between masculine and feminine ideals in the two cultures; and third, whether there was a preference for women with high pitch in Japan and medium to low pitch in the Netherlands. Her results confirmed all three hypotheses, and they showed that Japanese women did speak at higher levels on average than English, American, and Swedish women. The only potential methodological weakness in van Bezooijen's study was that pitch values were reported in Hz and not ERBs or some other transformed unit, leaving the data open to confounding via physiological factors like body size. If, for instance, Japanese women tended to be smaller in stature and have proportionately smaller larynxes than the other women in these studies, we would expect them to speak at higher levels, all other things being equal, owing to the basic principles of the Frequency Code (Ohala 1983; 1984). Given the relatively small difference in mean F0 between the three groups of speakers reviewed for this study, with Japanese women recording the highest average level at 232 Hz and Dutch speakers the lowest at 191 Hz (a difference of about 3.4 semitones or 0.9 ERBs), we might expect this kind of variation to be a factor. Formant data are often transformed to the Bark or mel scales to avoid this problem in sociophonetic studies, but mostly to prevent differences in gender from skewing the results, since male larynxes tend to be enough larger than female larynxes to lower the entire range of frequencies in speech (see e.g. Fant 1975; Wakita 1977; Disner 1980; Adank, Smits, and Van Hout 2004), so it is understandable (but still not desirable) that the same procedures would not have been applied to pitch data obtained from only females. Another potential issue is that because the mean F0 values for each group were obtained from separate studies, each with its own set of materials and procedures, it is possible that one particular production task led to more variability and thus higher mean values than others. There are not many precedents in the literature to support this possibility, but the production data from the experiments presented in Chapter 3 do make it seem likely, as the mean and standard deviation for F0 in the guided discourse completion task were higher than in the unguided storytelling task. Despite these two possible confounders in the metadata used for her initial comparison of interlanguage speaking levels, van Bezooijen's perception experiments confirm that Dutch speakers preferred a lower level and Japanese speakers a higher level in females, indicating that there is at least a grain of truth to the idea that level can vary significantly between languages when potentially confounding factors like gender are controlled for.

2.3 Previous Studies of Intonational Acquisition

Early research on L2 acquisition led linguists to form general conclusions about the aspects of intonation that adult leaners are most likely to struggle with. Backman (1977) and Jenner (1976) both found, for example, that learners had problems with the placement of prominence in pitch accents, and Willems (1982) and McGory (1997) found that they placed pitch accents incorrectly on unstressed syllables. All these studies point to the challenge of acquiring the structure of non-native intonation. Adams and Munro (1978) and Lepetit (1989) also found that learners had a tendency to replace rises with falls and vice versa, indicating a similar problem. For the most part, recent research has borne out these findings, although a few studies, most notably Jun and Oh (2003) (to be discussed below), have shown that learners can improve their ability to use the L2 tonal categories in speech. Many of these studies also revealed that learners struggled to acquire the phonetics of L2 intonation as well, producing final rises that were either too high

(Willems 1982) or too low (Backman 1979) and speaking with a narrower pitch range (Jenner 1976) than native speakers. More subtle errors were also possible, with Willems finding that learners of Dutch both had a lower than normal declination rate and did not properly reset their starting pitch from a low to a mid level after phrase boundaries.

Following these early efforts, contemporary experimental research has examined the question of L2 intonational acquisition within the AM framework. To examine whether learners could acquire L2 intonational phonetics, Mennen (1998) examined peak alignment and pitch range in the speech of Dutch speakers of L2 Greek. Although the two languages are prosodically different in many ways, they are similar in that their IPs commonly contain prenuclear rises that have the same underlying tonal specification (Mennen transcribes them both in a later study as L*H). Interested in whether speakers could acquire L2 prosodic phonetics when the L1 and L2 phonology were the same, Mennen focused her investigation on these rises, which despite having the same basic structure in both languages, peak earlier and have a smaller range in Dutch than in Greek. The experiment featured three groups of speakers—one for each of the L1s, and one for the L2 Greek—and was based on production data gathered from a single task requiring participants to read 20 sentences aloud in both languages. The Greek sentences were borrowed from an earlier experiment on peak alignment in that language (Arvaniti, Ladd, and Mennen 1998), and the Dutch sentences were written to match them in structure and content. In both sets of sentences, word stress for words bearing prenuclear accents fell on the antepenultimate syllable. After performing a one-way ANOVA for both peak alignment and pitch scaling with speaker group as the independent variable, Mennen found that the learners' alignment times were similar to those of the monolingual L1 Dutch speakers [F (2.57) = 49.38; $p < 0.0001$] and that their overall range was smaller than that of both them and the monolingual L1 Greek speakers [$F_{(2.57)} = 10.018$; $p < 0.0001$]. Her main conclusion, which would also be borne out by later research on the same general topic, was that learners struggled to acquire the phonetics of the L2 prosodic system, especially its components relating to pitch. One of the speakers in the study did produce alignment and scaling values within the corresponding ranges for Greek L1 speakers, so she also concluded that full acquisition is possible, though certainly not probable.

To the present author's knowledge, there have been no studies on the acquisition of French intonational phonetics by adult native speakers of English. Research has been done, however, on the acquisition of both the phonology and phonetics of similar intonational systems. In their study of native English speakers learning Korean, Jun and Oh (2000) found that advanced learners produced more native-like prosodic structures than intermediate and beginning learners. Like French, Korean lacks lexical stress (Jun 2005), and its IPs are composed of APs built on an underlying /LHLH/ tonal configuration, with the final H being perceptually the most prominent. While questions in both languages are marked by a H% boundary tone, Wh-questions are distinguished from polar questions in Korean by intonational phrasing rather than syntactic or lexical information. To measure learner success, Jun and Oh examined two main variables: phrase boundary placement in questions, and the realization of AP-final H tones. In terms of phrasing differences between question types, which they note was the most difficult contrast for the learners to acquire, the advanced learners succeeded 63% of the time, while the intermediate and beginning learners succeeded on average only 31% of the time. Beginning learners also tended to give each word its own AP regardless of utterance length, showing a difficulty with intonational phrasing. In terms of tonal realization, the learners reliably produced AP-final but not AP-initial H tones, and their pitch range was much narrower than that of the native speakers, comprising on average only 20% of their speaking range. Because phrase boundary tones are distinctive in Korean, but the realization of other tonal patterns is not, Jun and Oh concluded that the phonological aspects of intonation are easier to acquire than its phonetic aspects.

Continuing her work from 1998, Mennen (2004) examined bi-directional intonational transfer in Dutch non-native speakers of Greek. Prenuclear or non-final rises have the same phonological structure in both languages, but they differ in their phonetic realization, with peak alignment not only occurring earlier but also being affected by vowel length in Dutch. The study consisted of two experiments, one looking for transfer from the L1 to the L2, and one looking for transfer from the L2 to the L1. Results for the first experiment showed that only one of the speakers was able to produce native-like L2 rises; the remaining four speakers produced L2 rises with alignment patterns from the L1. Results for the second experiment showed that the same four speakers produced L1 alignment patterns that were significantly different from those produced by monolingual L1 speakers, indicating an effect of exposure to the L2. Together, the experiments support the claim that intonational transfer can go both ways, i.e. from the L1 to the L2 and from the L2 to the L1. Crucially, they also support Mennen's (1998) and Jun and Oh's (2003) findings that the phonetics of L2 intonation are difficult for speakers to acquire, regardless of whether the L1 and the L2 have similar tonal inventories and distribution. However, because prenuclear rises have the same phonological structure in both languages, the study could not test the difficulty of acquiring the phonology of L2 intonation.

Further work by Mennen, Chen, and Karlsson (2010) extended this research by examining the interaction of L1 and L2 intonational systems in learner varieties of Standard

British English spoken by Italian and Punjabi L1 adults. One reason this particular study is not only innovative but is also relevant to our present research questions is that it assumes the L2 intonational systems to be fully-formed learner varieties (Klein and Perdue 1997) rather than malformed varieties of the target language. This approach is based on the observation that learners, regardless of whether they share the same L1, tend to use similar tonal structures in the L2, suggesting that they learn to communicate in the L2 by tapping into some sort of a universal but reduced phonological system, a suggestion only recently considered for L2 intonation. In this particular study, learners were recent immigrants to the United Kingdom, where they lived but importantly did not receive classroom instruction in English, meaning the majority of their exposure to the target language was in non-formal environments, like work, school, or social settings. This distinction is important because we can assume that most progress the learners make in acquiring the foreign intonational system will be the result of their creating or discovering their own learner varieties based on informal L2 input rather than the result of their imitating instructors with their sometimes intentionally artificial pronunciation (see e.g. Rubin 1974; Chaudron 1988; Morley 1991; Wallace 1991; Lockhart 1994) or of completing exercises from a text.

To gather production data for their analysis, the authors used a balanced subset of the European Sciences Foundation (ESF) L2 Database (Perdue 1993), which they then analyzed for a number of intonational features, including tonal inventory, tonal distribution, and pitch scaling. Their results were important in two ways. First, they showed that at the beginning of the 30 month study period, the learner varieties of L2 English for the two groups of speakers were structurally very similar in that they shared the same inventory of pitch accents and featured short IPs that were on average only two words long. The learners also had a reduced pitch span as compared with English L1 speakers. Both facts support the learner variety hypothesis, and we can assume that because Punjabi and Italian are both genetically only distantly related and structurally quite different, the learners formed their L2 intonational systems using a reduced set of underlying tonal categories. Second, their work showed that learners had improved their ability to use the L2 tonal categories by the end of the study period, with the Punjabi speakers producing more falling contours and relying less on their preferred L*H accent, and both groups of speakers producing longer IPs and more native-like distribution of tones. Where the speakers did not improve, however, was in the number of tonal categories they were able to produce, which remained the same (i.e. in its reduced state as compared to L1 English) throughout the study period. Related to this finding was that the speakers did not learn to use complex contours, which Grabe (2004) points out are common in London English (particularly in questions) and thus composed a significant portion of the L2 contours to which they were exposed.

The improvements made by the speakers in this study were largely phonological, though, and do not speak much to the ability of speakers in general to acquire L2 intonational phonetics, which much of Mennen's other work (2004; 2007; with Schaeffler and Docherty 2012) has shown to be difficult, a conclusion supported by Jun and Oh's (2003) study of English L1-Korean L2 learners described above. Some questions we might ask in response to these results is whether the learners perceive the L2 categories as distinct from their own, whether their pitch range expands with longer exposure to the L2, and whether the slope, alignment, and duration of tonal targets in their L2 is similar to either of the L1s. Evidence from studies of segmental acquisition (Flege 1994) suggests that learners will implement L2 phonemes with L1 phonetics in cases of perceived phonetic overlap, so it would be interesting to know whether the speakers in Mennen's study were actually using a reduced universal intonational system or were simply using a compressed or simplified version of the tonal inventory from their L1.

2.4 Modeling Pitch Data in Speech

The strongest reason that perceptual models of L2 acquisition like the PAM and SLM are not widely used in studies of L2 intonation is that researchers have struggled to agree on a method for quantifying and then normalizing variation in pitch range (Ladd 2008; Mennen, Schaeffer, and Docherty 2007), which as mentioned above would be necessary to extrapolate coarse-grained phonological categories from fine-grained phonetic detail. While there are a number of methodologies for handling variation within a single language, including the additive-multiplicative model (Pierrehumbert 1980; Liberman and Pierrehumbert 1984; Pierrehumbert and Beckman 1988; Liberman et al. 1993) and z-score normalization (Rose 1987), the latter being very commonly used to develop language models for computational tasks in speech processing and recognition, there is no consensus in the literature on which method(s) should or even could be used cross-linguistically.

The main issue that these models attempt to address is one described nicely by Gussenhoven (2004), who points out that there are two major sources of variation in intonational phonetic data: pitch range or span, and pitch level. In theory, pitch span refers to the distance in Hertz between a speaker's minimum and maximum pitch, but in practice, it is often reported as the distance between the 10th and 90th percentiles of a speaker's total range (i.e. the 80% quantal range) to prevent F0 perturbations, speech disfluencies, and pitch tracking errors from affecting mean values. Span has been given a few other names in the literature, most notably excursion size ('t Hart, Collier, and Cohen 1990) and key (Cruttenden 1997), but the basic principle is the same. There is some disagreement about the extent to which span varies among speakers, with some researchers maintaining that differences can be largely accounted for by normalizing pitch values within the speech community (Rose 1987), and others maintaining that no amount of mathematical manipulation can completely standardize the data (Ladd 2008). In either case, span is intended to be a simple measure of how widely a speaker's pitch space varies, which studies in L2 acquisition have found to be a rough indicator of spoken fluency, with a smaller span being typical for beginner and intermediate learners (e.g. Mennen 2010; Jun and Oh 2003) and a larger span for advanced learners. Generally, for both learners and native speakers, span will increase when F0 maxima increase, since F0 minima usually hold constant rather than decrease (Gussenhoven 2004). Simultaneously raising the minima and maxima would simply constitute an increase in pitch level rather than range.

Pitch level (Ladd 2008) or register (Cruttenden 1997), by contrast, simply describes the height of a speaker's mean fundamental frequency, which, like span, is normally reported in Hertz. The most common way this measurement is used to distinguish speakers is by sex, with males typically speaking at lower levels than females. The linguistic importance of pitch level varies. In the broad sense, speakers may use pitch level to signal their attitude toward other speakers, with low level signaling confidence, relaxation, or aggression (Ohala 1983, 1984), and high level signaling friendliness, nervousness, and submission (Uldall 1964; Gussenhoven 2002). Pitch level may also reflect differences in the speaker's attitude toward the discourse. Gussenhoven (2002) suggests, for example, that low pitch signals certainty and assertion, while high pitch signals uncertainty and questioning or interrogation. Collectively, these trends are taken to be examples of the Frequency Code, one of the three proposed biological codes relating pitch use in language to basic features of human physiology and evolutionary psychology. There is also some evidence that languages may employ pitch level differently. As described above, recent research has found, for instance, that female speakers of Standard British English (SBE) to have a higher average speaking level than female speakers of German (Mennen, Schaeffler, and Docherty 2007, 2008, 2012; Mennen, Scobbie, de Leeuw, Schaeffler and Schaeffler 2010).

2.4.1 Alternative Transcription Systems

Before discussing methods for dealing with intra- and interspeaker variation in linguistic pitch, we should take a brief look at the available transcription systems aside from ToBI and discuss their drawbacks and benefits. Although this dissertation assumes an AM-centric view of intonation and uses a simplified ToBI system for its transcriptions, the other systems make certain assumptions about the structure of intonation and the shape of pitch contours shared by HMM-based intonational variety classifier outlined in Chapter 4, and so they are worth reviewing here.

Aside from the Tilt Model of intonation (Taylor 1998; 2000), there have not been many attempts to describe intonation from the top down, i.e. from the pitch contours themselves rather than from the proposed underlying tonal sequences that generate them. The main reason for this is probably theoretical in that phonological tones are easier to integrate with other components of the grammar than are chunks of continuous pitch data. Because the tones were designed with the other components in mind, this does make sense. Still, as large digital corpora became available and pitch tracking software improved in the late 1980s and the 1990s, linguists began to experiment with ways of using mathematics to model intonational systems in their entirety. The

Momel algorithm (Hirst 1988, 1997, 2000; Hirst and Espesser 1993), short for 'modeling melody', is a good example of such a system, having been developed for research on French intonation but extended in recent years to research on other languages, including English (Auran 2004), Italian (Giordano 2005), Catalan (Estruch 2004), Brazilian Portuguese (Fernandez-Cruz 2000), Venezuelan Spanish (Gallardo 1996), Arabic (Najim 1995), Russian (Nesterenko 2006), isiZulu (Louw and Barnard 2004), and Korean (Hirst, Cho, Kim, and Yu 2007). The basic idea behind the algorithm is we can model any given pitch contour by stitching together shorter curves that approximate each of its main components. Figure 2.1 shows an example of how the process might work for a single LH* rising contour. The first step is to divide the contour into smaller

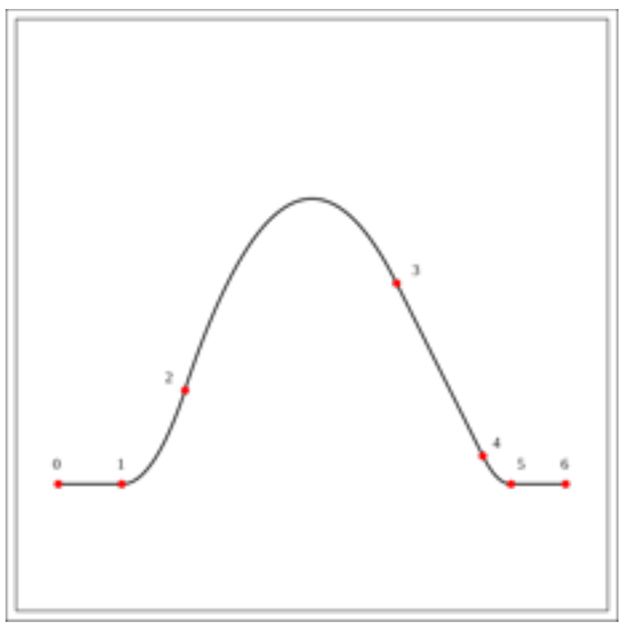


Figure 2.1. *Exemplification of the Momel algorithm (Hirst and Espouser 1993)*.

segments, denoted by the numbers along the curve below, and the second is simply to approximate the shape of each segment using a quadratic function, with the function's second derivative indicating whether its corresponding segment is concave or convex. In the example above, the first section, between points 0 and 1, would be approximated by a straight line with a slope of zero, as would the last section—although they are straightforward to model, both sections are important to include because they are tied to key phonetic parameters like syllable onset and peak alignment. The intermediate sections, however, are trickier, and if we tried to model them all at once, we would have to make some sacrifices to local accuracy in order to capture the shape of the whole curve. Treating these segments separately improves the model's fit, since each segment has its own slope and duration. The end result, then, is a piecewise quadratic function that mathematically summarizes the pitch accent's most important components, making the model clean, accurate, and relatively efficient (doing this kind of analysis by hand would be time-prohibitive, but with microprocessors performing millions of operations per second, it is very feasible). Because the equations describe the shape of the pitch contour itself and not its parameters or underlying components, the model is also very useful for intonational synthesis.

The Momel algorithm is ideal because it produces pitch contours that are almost identical to those found in natural speech. Its downside, however, is that it assumes all of the information in surface-level contours is worth modeling; in other words, it does not discriminate between linguistic, paralinguistic, and completely non-linguistic events that are realized in F0, and it produces quadratic functions that are affected by variables that are essentially unrelated to those we might want to study in examining intonation and, more specifically, learner varieties of intonation. An exaggerated analogy to this kind of catch-all modeling would be measuring the angles and rate of change in speed for each of a runner's legs over a period of time to calculate her distance covered, when it would in fact not only be much more straightforward but also less subject to errors to measure the distance between where she started running and where she stopped. The extra information is interesting and would be relevant in other contexts (say, during a study of human biomechanics), but we do not need it to answer the question at hand. Momel has successfully been used to synthesize intonation in text-to-speech applications (DiCristo and Hirst 1986; Hirst and Espesser 1993; Hirst 2000; Hirst 2012), however, so its inclusiveness as a statistical algorithm does not appear to have hindered its accuracy as a linguistic model.

The other primary drawback to Momel, and the reason it is not ideal for research on comparative intonation or learner intonation, is that it is a purely mathematical rather than a statistical model, meaning it can not straightforwardly be used to draw inferences about the intonational parameters of an entire language or dialect. The Tilt model, the Fujisaki model (Fujisaki and Hirose 1982; Fujisaki 1992; Fujisaki and Ohno 1995), and even tone-based models in the AM framework can all be used to determine the average values for intonational phenomena like pitch accents, boundary tones, downstep, and declination, but Momel can only model individual contours. This limitation does not pose a major problem for studies based on data obtained from a single speaker, but when data have been gathered from a group of speakers or from speakers of different languages, it becomes much more problematic.

Another prominent model used to produce stylized intonation contours is the Prosogram (Mertens 1995; 2004). The model is built around what Mertens defines as the basic objectives for prosodic transcription, which are mainly that the process should be semiautomatic or if possible automatic, that the resulting representation should be objective and robust, and that it should preserve the relationship between intonational events and acoustic speech parameters like time and pitch. Unlike Momel, the Prosogram is designed to produce stylized or simplified transcriptions of intonational data, preserving only the perceptually salient parts of the contour and discarding the rest. The resulting transcription (Figure 2.2) is similar in style to a musical score, with events occurring at discrete levels in the pitch space and lasting for discrete amounts of time (Meertens 2004, p. 2).

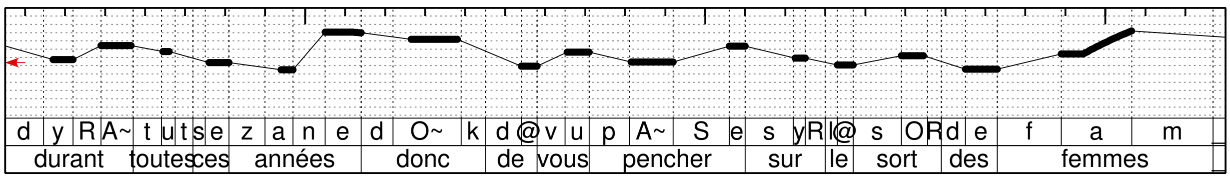


Figure 2.2 *Sample French Prosogram*.

The strength of this model is that it greatly simplifies the raw F0 contour, which, even when smoothed during processing, can be jagged and highly variable. Like Momel, the model also ignores phonological information in the utterance, focusing instead on the duration, amplitude, and intensity of the pitch movements. The question remains, however, of how the Prosogram could be converted into a set of parameters that could be compared across languages. We could calculate the average duration and level of the pitch events, but it would not give us a very detailed picture of the language's phonetics. We could also parameterize the Prosogram values and model them sequentially as a Markov chain, like in the Tilt model, but we would lose important information about the shape of the contours themselves, like their slope and concavity. As a transcription system, the Prosogram is extremely promising, but as a statistical model, it does not have the robustness or resolution we need to compare language varieties on the basis of pitch and is thus abandoned here in favor of the other models.

2.4.2 Psychoacoustic Scales

Intonation, pitch accent, and tone are all linguistic manifestations of pitch, which, along with loudness and duration makes up the acoustic properties of prosody. When linguists refer to pitch, they usually refer to fundamental frequency, or the rate at which the vocal cords vibrate during voiced segments in speech. In acoustics and other branches of physics that deal with sound, frequency is most commonly reported in Hertz, a scale measuring how many times a particular sound wave peaks per second, a tradition that has been carried over into linguistics. Frequency and pitch, however, refer to two separate phenomena, and the difference is crucial for understanding research in intonational phonetics. Measurements of raw frequency, whether they are reported in Hertz or in transformed units like mels or Equivalent Rectangular Bandwidths (ERBs), which are discussed below, refer only to the number of cycles a sound wave completes per second and do not tell us anything about how it is perceived by the ear, since measurements were neither designed nor intended to convey perceptual information. Pitch, on the other hand, refers to the perceived frequency of a sound wave and is of much more importance to intonation, pitch accent, and tone, since speakers learn these systems by listening to others and refine them by listening to themselves. There is of course a significant body of work that addresses the production and perception of prosody by profoundly hearing impaired individuals, but the research is beyond the scope of this dissertation to discuss. The particular way in which humans perceive sound frequency is due to the structure and function of the cochlea, a part of the inner ear that bends around itself in a spiral shape similar to that of a nautilus's shell, with the beginning of the tube vibrating in response to high frequencies and the end in response to low frequencies. Because of this shape, the relationship between frequency and perceived pitch (i.e.

the brain's interpretation of frequency after it has passed through the cochlea and been converted to electrical signals by the small hairs lining its inner wall) is not directly linear, so that a 100-Hz increase in frequency will yield different increases in pitch depending on the frequency of the original tone. We can see a simple example of this transformation in the semitone scale, where moving from one note up to the next note with the same pitch (i.e. up an octave) requires doubling the frequency of the first, regardless of where along the scale the two notes are located. The semitone scale is widely used in Western music, but because it also approximates the basic shape of the human auditory filter, linguists have used it for research on intonation and tone. Contemporary proponents of the scale include Kügler (2007; 2009), Féry (2008), Mennen (2007a; 2007b; 2008; 2010), and many of the researchers associated with the phonetics labs at the University of Cambridge and the University of Oxford, including Kochanski (2009), Grabe (in Eriksson, Grabe, and Traunmüller 2002), and especially Nolan (2003; 2006), who has suggested outright that the scale is more appropriate for representing pitch data than the Hertz. The other so-called pscyhoacoustic scales that have been used for work in phonetics are the mel, Bark, and Equivalent Rectangular Bandwidth (ERB) scales (Glasberg and Moore 1983; 1990). The first two have been in use for about 50 years and are still used today to normalize vowel formants in studies on sociophonetics, but the third scale is relatively new.

The ERB scale is something of a hybrid in that it is approximately linear at low frequencies but logarithmic at high frequencies, a result of its being derived directly from the estimated shape of the auditory filter. Recent work in psychoacoustics points to the scale as being the most accurate, and a number of authors in linguistics have suggested the same (Hermes and van Gestel 1991; Hermes and Rump 1994, 1996; Terken and Hermes 2000), so it will be used as the basis for the phonetic analyses in Chapters 3 and 4 below. Using the ERB scale also sidesteps the issue of transformed values raised by Rietveld and Gussenhoven (1986), who maintain that the Hertz scale more closely matches speakers' perception of pitch in judging the prominence of accented syllables, because both scales are approximately linear in the frequency range characteristic of the F0 data in speech. Thus, we can have the satisfaction of knowing our scale is psychoacoustically realistic and linguistically relevant. The choice to use *some* transformed scale is obvious, though, as there should be no reason to report values for perceived pitch on a perceptually unrealistic scale, regardless of whether linguistic experiments support their use. The one counterargument to this statement, that the ERB and Hertz scales have very similar shapes in the frequency range characteristic of most speakers' F0, is valid, especially since the latter has been widely used for the past two to three decades in phonetics and acoustics research. This similarity notwithstanding, it still seems more justifiable, at least in theory, to use the ERB scale, if not for the reasons given above, then for the sake of ensuring that intonational data are reported in the same way as other phonetic data.

2.4.3 Normalizing Pitch Data

The psychoacoustic scales presented in the previous section describe pitch data in a way that is perceptually accurate, but they do not account for a great deal of interspeaker variation in the realization of intonational events (Ladd 2008). Standard German provides an excellent example of this shortcoming, with Gilles (2005) finding a difference of three semitones in the realization of a declarative falling contour by speakers from Dresden and Duisburg. Other examples of this phenomenon have been found by Earle (1975) for Vietnamese, Rose for Mandarin Chinese (1987), Grabe, Kochanski, and Coleman (2003) for English, and Ladd (2008), among others. The basic problem is that speakers may have similar speaking levels but not ranges or similar ranges but not levels, and that simply reporting pitch data on a psychoacoustic scale does not eliminate these kinds of irregularities.

Broadly speaking, there are two ways to reduce this kind of interspeaker variation in pitch data: z-score normalization and percent speaking range (PSR) normalization (Ladd 2008). To begin with the simpler of the two procedures, PSR normalization represents pitch height for tonal events in terms of how big they are in relation to the speaker's total pitch space. As a simple example, we can consider a female speaker with a total speaking range of about 10 semitones, just over three-quarters of an octave, whose friends consider her to be very expressive. She gets excited when asking questions, so on average she realizes her H*H% nuclear configuration as a 7-semitone rise, equivalent to a perfect fifth on the musical scale, and also a very common height for unbiased polar questions. Her friend, who is much less excitable, realizes the same configuration as a 3-semitone rise, but his total speaking range is only 6 semitones, and the same group of friends consider him to be somewhat monotone and drab. The PSR-normalized pitch ranges for the two speakers and this particular tone sequence, then, would be 0.7 and 0.5, respectively, and the mean height for the two as a group would be 0.6. The normalization procedure nicely captures the generalization that although some speakers may have a wider overall range than others, the relative height of any particular pitch accent or nuclear configuration tends to be the same. Although this concept has only recently been applied to pitch data from intonation languages like English (Mennen 2010), it has been used very successfully to normalize pitch ranges and slopes for lexical tones in tone languages. In particular, Earle (1975) showed the method to be effective in reducing interspeaker variation in

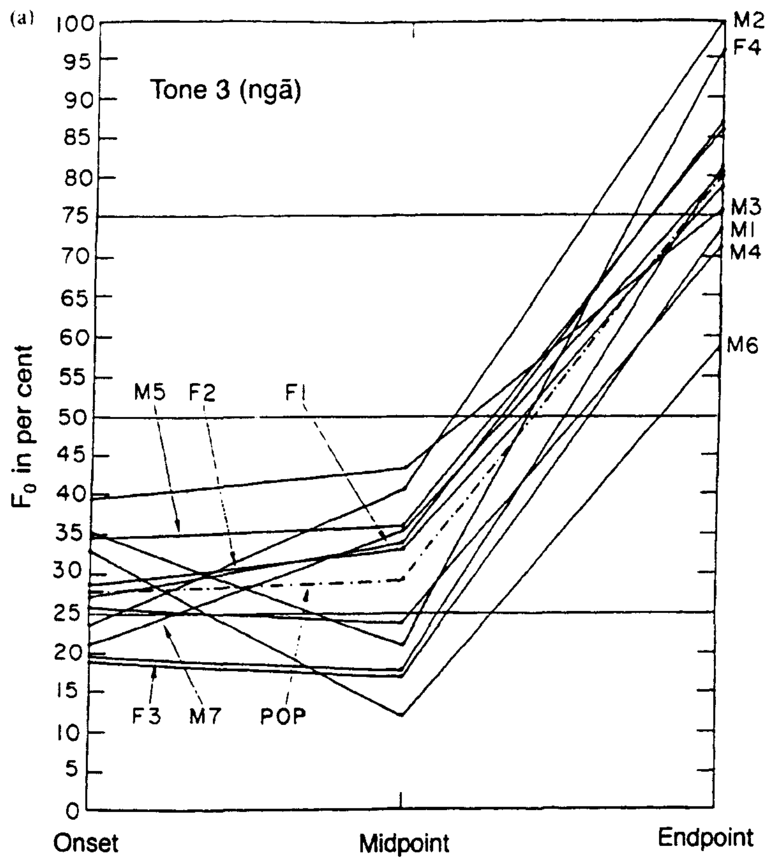


Figure 2.3. *PSR-normalized pitch contours for Vietnamese Tone 3.*

the height of Vietnamese tones, with time- and PSR-normalized pitch contours (Figure 2.3 above) converging much more closely than the raw contours (image from Ladd 2008, p. 194). Ladd (2008) also found the same results for Mambila, a Bantoid language spoken in the border area between Cameroon and Nigeria, with Earle's model reducing all but a small remaining amount of the variation in realizations of the language's four level tone phonemes. Ladd also makes the closely related point that reporting the pitch data on a psychoacoustic scale does not yield much consistency in range, with speakers differing in the realization of the language's High-Mid tone, for example, by over a semitone. In reality, a difference of a semitone is likely not that significant, especially since Ladd's dataset was based on five speakers, which even with hours of recorded speech is going to be too small to allow for inference about the population (in this case native speakers of Mambila) to be drawn. This issue is, in fact, widespread in laboratory phonetics and phonology, and it will be addressed in Chapter 3 in discussing the participants and resulting corpus of speech from which production data where gathered.

The difference of a semitone may not be perceptually relevant, either: ’t Hart (1981) suggests that listeners do not perceive pitch changes of less than three semitones in running speech, and although Rietveld and Gussenhoven (1987) refute this view by pointing out that they appear to hear changes down to a one and a half semitones, the basic idea probably has some merit (especially since the semitone is essentially an invented unit of measurement that is not employed universally even in music—many scales used outside of the Western tradition are built exclusively on intervals larger than the semitone, and some are even built on smaller intervals, like the quarter-tone). A final point worth mentioning here is that although psychoacoustic scales do not transform away interspeaker variation completely, there are likely soft limits on the acceptable size of overall speaking range and individual tone height in each language. For example, there is little doubt that a opera singer who used her full two- to three- octave (24- to 36-semitone) singing range to realize pitch contours in speech would sound strange. The example is somewhat hyperbolic, but it does illustrate the point that just because the parameters of a system are subject to variation does not necessarily mean they have no practical (if not theoretical) limits.

Z-score normalization is based on the mean and standard deviations for each speaker’s F0 and is widely used in speech processing because of its statistical robustness and computational efficiency. Rose (1987) was one of the first to use the approach for research on linguistic pitch in his study of Wu Chinese tones. He showed that reporting pitch values on a standardized scale reduced interspeaker variation in the realization of isolated tones and made it possible to determine the approximate height and slope of their corresponding contours. In his later work on

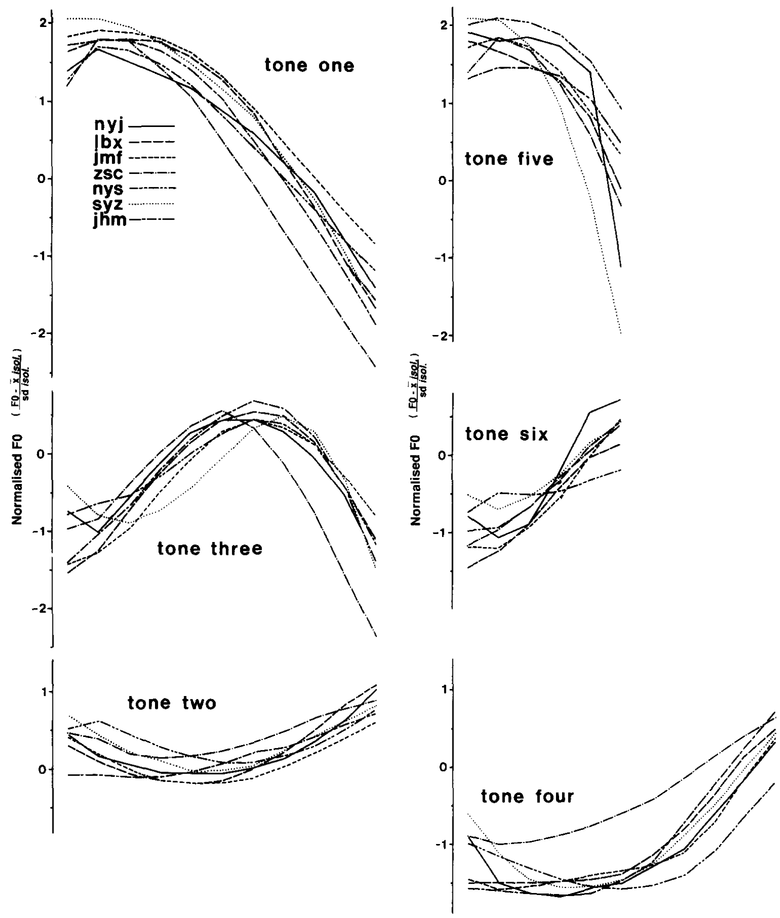


Figure 2.4. *Longterm normalized tones in Zhenhai Chinese.*

Zhenhai Chinese (1991), he revised the method to include long-term normalization, which was not better at reducing interspeaker variation than normalization with isolation tone parameters, but which has subsequently been used to model speaking range in non-tonal (i.e. intonational) contexts. Figure 2.4 (Ladd 2008, p. 195) shows the normalized pitch contours for six tones in this particular variety of Chinese. What is particularly interesting about these plots is not only that they clearly demarcate the phonetic boundaries between the tones, but also that they show significant consistency among speakers within each tone. Also, they show that the procedure can identify tones that are structurally similar. With the most significant change in pitch occurring halfway through the contour, tones 2 and 4 are separated only by the depth of the onset and the height of the eventual rise, both of which differ by only about one standard deviation. This sort of acoustic resolution is ideal for working on intonation, since pitch accents are often only distinguished by relatively subtle changes in scaling and alignment, and separating them on the basis of phonetic data is necessary to understand both the structure and relationship to the L1 of the L2 prosodic system.

Applying this normalization procedure to intonational data presents a methodological challenge, however, and one that is not insignificant. As Rose (1991) rightly points out, F0 data from tone languages tend to have larger means and standard deviations than those from non-tone languages like English, a difference likely due in part to the Effort Code (de Jong 1995), since the greater articulatory effort required to accurately produce lexical tones should lead to an increase in the size of pitch movements. In a basic sense, this means we are at an automatic disadvantage for doing similar research on intonation languages, since the phonetic boundaries between intonational categories will not be as clear as they are between tones (essentially, this is the Effort Code working in reverse) and will thus be more difficult to tease out of intra- and interspeaker pitch data. Z-score normalization has been used to establish acoustic parameters for text-to-speech (TTS) and speech-to-text (STT) applications, but only to establish likelihood estimates for features like energy and frequency and not to construct distinct intonational categories from the data, which the process of which becomes much more difficult when considering the number of linguistic and paralinguistic factors that influence their number, distribution, and shape.

Another problem is one that stems from the comparative nature of assessing L2 intonation. Ideally, we would be able to produce summary statistics for each speaker's particular variety of L1 intonation, including mean level, mean range, standard deviation, and slope for specific categories and contours. Assuming we could produce the same statistics for their variety of L2 intonation, we could then compare the two datasets and come to some conclusion about well the speaker seems to have mastered the target language. The primary issue here, though, is that we do not know what these parameters would be for the speaker's L1 in the target language, i.e. if she had acquired the target language as a native language in childhood, so we cannot say what her variety of L2 intonation really should sound like. Even the most basic components of intonation, like pitch level and speaking rate, can vary by language, so there is no guarantee that they would be the same for two native languages spoken by the same speaker (in fact, they probably would not). Statistically, of course, we could measure the phonetic parameters of her acquired intonation and the probability of their being found in the target language—this is, in fact, one method outlined in Chapter 3, and it is likely the most practical given the expertise required to design and run the more advanced models mentioned in Chapter 4. This is still only an approximation, however, made on the basis of empirical evidence and one that is subject to error and uncertainty. In other words, there is no practical way to calculate the parameters of an individual speaker's native-like L2 intonation, and so any evaluation of her progress toward that goal is at best a highly sophisticated guess.

Both of these normalization methods will reduce interspeaker variation in pitch data to acceptable levels, at least in theory, making the issue of how those data are reported (e.g. on the Hertz vs. a logarithmic pscyhoacoustic scale) less important. Nonetheless, because the ERB scale is the most realistic approximation of how the ear processes pitch, it does not make much sense to use anything else, even if the results are minimally different or even the same in some cases. Therefore, the following chapters will make use of both psychoacoustic and z-score normalization in preparing the pitch data gathered from the production tasks for phonetic analysis in the hope that combining the methods will give us the best chance of finding clear patterns in the learners' L2 intonation. After seeing how well the normalization procedures work, recommendations will be made in Chapter 5 regarding whether they should continue to be used in studies on intonation, and, if so, in what combination.

2.4.4 Pitch Data and Perception Research

Related to the empirical issues of measuring learner progress in acquiring L2 intonation is a formal one in that models of L2 phonology like the SLM, the PAM, and the perceptual magnet effect model (Iverson and Kuhl 1995) have been used almost exclusively to evaluate the perception and production of segmental categories rather than prosodic categories. What makes this shortcoming more bittersweet for researchers interested in prosody is the success of the models in accounting for the subtleties of segmental acquisition. For example, in examining learners' perception of non-native consonants, Best, McRoberts, and Goodell (2001) found that the success rate was near ceiling for discriminating L2 consonant contrasts they perceived as similar to consonant contrasts in L1, but was much worse when the L2 sounds converged perceptually on a single L1 category. Best (2002) found similar results in a study on the perception of non-native vowel contrasts by speakers of AE, and Flege (1999) found that perceived phonetic difference also affected long-term acquisition of English vowels by speakers of Italian. Both models have been tested on many more occasions, and the results have led to some of our clearest insights into learners' perception of non-native phonological contrasts. Following these successes, Best and her colleagues have attempted to extend the PAM to prosody, and in particular to the analysis of learners' perception of non-native tonal contrasts. We see an early example of this work in Best, Levitt, and McRoberts (1991), where the authors examined how well the PAM can account for L1 intonational development (this is relatively unrelated to the issue of L2 acquisition, but it is the first real look at how perceptual models might be integrated with intonational theory, so it is worth mentioning). Their primary finding was that infants between 10 and 12 months old could discriminate native but not non-native prosodic contrasts between statements and questions. The study did not examine intonational data per se, but it did clearly show that the prosodic system is shaped by language exposure, and that it develops very early in childhood, which in one sense might explain at least informally why speakers struggle to learn new prosodic systems as adults.

More relevant perhaps is Best's later work with So on the perception of Mandarin tones by speakers of intonation (i.e. non-tone) languages. The two have produced a string of studies (e.g. 2008; 2010; 2011; 2014) examining the interaction between Mandarin tones and the English intonation system as it is defined by the categories of the AM framework. The experimental approach for each study is unique, but in general their results are the same, with English speakers perceiving tonal contrasts as long as the individual tones were phonetically similar to intonational categories. They also examined the perception of the same contrasts by native speakers of French (2011; 2014), whom they found to be more successful in identifying fine-grained differences between tones than the native English speakers, particularly with Tones 3 (low/falling) and 4 (high-falling). What the studies do not do, however, is determine to what extent the speakers' perceptual assimilation of the tones affects their ability to discriminate them in L1 Mandarin or acquire them in L2 Mandarin, which is of principal interest to our efforts to understand the actual effect of non-native prosody on native prosody and vice-versa. The other methodological issue with these and other studies in their attempts to characterize how learners perceive non-native prosody is that they use very broad intonational categories as the basis for their comparisons. So and Best (2011), for example, used only statement, question, and exclamation intonation when asking listeners to determine whether the Mandarin tones were good examples of English categories. There are many reasons for this kind of generalization, and it does not mean their results are not interesting, but it does point very strongly toward what should be a common goal for research on L2 prosody, which is the development of phonetically-motivated fine-grained categories that can be used for this sort of comparative research. Still, the studies are promising in that they prove that perceptual models of language acquisition like the PAM can be extended to the prosodic domain, and as the methods for quantifying pitch data are refined, their results should become increasingly applicable to the field.

Because of the methodological shortcomings in So and Best's work, a secondary goal of this dissertation is to revisit these attempts and determine whether a more clearly defined picture of the phonetic variation in L2 intonation might increase both the PAM and the SLM's potential to predict whether learners will acquire L2 prosody and, if so, to what extent. Because the SLM addresses language learning more directly than the PAM, it seems like the most likely model to be used for perception experiments on L2 prosody and intonation. Because the SLM claims learners can form new phonological categories at any stage of life but only if they are distinguishable from neighboring categories in their L1, and despite evidence that L2 acquisition becomes difficult after the L1 phonological system is established (Flege 1995; Kuhl 2000; McClelland 2001; Iverson et al. 2003), the model means that adult speakers (as opposed to early-stage bilingual children) can successfully acquire and produce certain L2 phonological categories, provided that certain perceptual criteria are met. Testing this prediction requires the sort of phonetic data presented below, and although describing the characteristics of learner intonation is reason enough to pursue such an analysis, developing a foundation for perceptual research in prosodic acquisition must be (and in fact is) another if not equally as strong motivator for doing so. Current research has not approached the issue of measuring learner intonation from this standpoint, and so it will be an explicit goal of the analysis to follow in Chapters 3 and 4.

2.4.5 A note on ToBI Labeling

Wherever possible, ToBI transcriptions have been provided for example pitch contours taken from the L1 data. For reasons described above, it was not practical to do the same for the L2 contours, and so they are presented in the chapters below without transcriptions. Features like phrasing and tonal alignment are marked to facilitate phonetic comparison of the two varieties.

CHAPTER 3

THE PHONETIC REALIZATION OF L2 INTONATIONAL PHRASING AND TONAL CATEGORIES

The acquisition of second language phonology is influenced by many factors, including the intensity and duration of exposure to the target language, the difference in phonetic and phonological structure between the target language and the first or native languages, and the learner's own aptitude for acquisition (see Oxford 1994 for an overview of the topic). The goal for this chapter is to evaluate all three of these factors as they relate to the ability of adult native speakers of English to acquire the intonational systems of two varieties of French, Standard or Parisian (SF), and Southern or Meridional (MF). Despite the fact that the perception of non-native phonological categories and phonetic gradience plays a crucial role in adult L2 acquisition (Flege 1995; Best 1995), the following discussion will focus on the results of several production experiments that were carried out with the intention of determining the precise structural and acoustic characteristics of the learners' L2 French. The reasoning behind this approach was detailed in Section 1.3 of the Introduction, but the basic idea is that too few studies have been done on L2 intonational phonetics to make perception experiments possible, as these rely on the establishment of phonetic boundaries between phonological categories in each of the languages to be compared. In addition to answering a number of fundamental questions about the nature of L2 intonation, this chapter presents analyses of L1 and L2 French corpora that will hopefully serve as the foundation for this kind of research in the future.

The chapter is organized as follows. Section 3.1 provides information about the speakers who were used to gather production data for the study. Section 3.2 is is divided into two subsections. The first details the production tasks that the speakers completed and explains the rationale behind them, beginning with a quick summary of how they have been used in the past for similar experiments and ending with a discussion of why they were chosen over other methods; the second subsection explains how the tasks were delivered during the experiments. Section 3.3 begins the core of the chapter and presents an analysis of the learners' intonational phrasing in French. Section 3.4 focuses on the alignment of tonal targets in the L2 phrase-final rises, beginning with a short introduction of the rises' main syntactic, semantic, and pragmatic features, and followed by a comparison of their realization in the L1 and the L2 speech. Data from all groups of speakers will be examined in each of these subsections, with the idea being to organize the discussion around French tonal categories rather than the methods of foreign language instruction (these will be discussed in the conclusion to the dissertation because they do not bear directly on the research questions of interest to this particular chapter). Section 3.5 completes the phonetic analysis by examining pitch scaling in the L2 rises, as well as overarching trends in the learners' speaking range and level. Section 3.6 summarizes the results of the chapter and discusses directions for future research into L2 intonational phonetics within the AM framework based on its findings.

3.1 Participants

Production data were gathered from 20 adult native speakers of English studying at the University of Georgia. The participants were divided into three groups: those enrolled in a semester-long (15-week) study abroad program in France (Group A), those enrolled in a summer-long (6-week) study abroad program in France (Group B), and those enrolled in French 3010 (Group C), a course designed for intermediate students beginning a major or a minor in French to improve their conversation, reading, and writing skills. Although their spoken proficiency in French varied, all speakers had completed the equivalent of four semesters' (i.e. two years') worth of undergraduate French, meaning they had basic competence in French grammar, vocabulary, and culture. The four courses were usually completed in sequence during the first two years of their stay at the university, but two of the students received their credit from Advanced Placement (AP) and International Baccalaureate (IB) programs in high school and were thus taking their French 3010 as their first in-class college French course. The distinction likely does not matter, since speakers at the intermediate level usually range widely in their spoken proficiency, but there is an outside chance that the two students' earlier exposure to high-level coursework gave them something of a leg up on their college peers. These differences would be accounted for in a study designed to test the effectiveness of specific foreign language instructional methods (e.g. coursework vs. study abroad), but since the goal of this study is simply to describe the phonetic characteristics of learner intonation, they are treated as stemming from variation in the underlying population of learners and are thus assumed to be acceptable.

Because the participants were close in age and educational achievement, the sociolinguistic axis along which the participants varied the most was language background, described fully in Appendix A. Two speakers had completed college coursework at the same level or higher in Spanish as they had in French; another speaker had working knowledge of Spanish, Italian, and Portuguese; and the remaining participants had only French as their L2. All speakers were raised in mono- or bi-lingual households where English was the primary spoken language and where the second language was only used in specific contexts, for example in speaking with members of the student's extended family who were not proficient in English. As indicated by self-reported language background questionnaires, all the speakers were educated in monolingual English classrooms where foreign language instruction was reserved for elective coursework and was normally not required until the students had reached secondary school. This pattern is typical of American public education, and it provides a nice counterpoint to studies on adult language acquisition based on speakers from Europe, who are often exposed to at least one non-native language (normally English) relatively early on in childhood (Cenoz and Jessner 2000; 2001).

Based on these facts, we can draw a number of conclusions about the participants and their backgrounds in French. First, whether they completed the study abroad program or the advanced conversation course, the participants had their most prolonged and substantial exposure to spoken French during the study period; this is important not only because it puts them on equal footing coming into the study, but also because it means we can make some basic inferences about immersion (either abroad or domestically) as a pedagogical tool that would otherwise be impossible to make. As a direct result of this exposure, we can also assume that the proportion of time the participants spent speaking French increased during the study period. Freed (1995) found that students often use their L1 when studying abroad, but it is hard to imagine them speaking less French during a program with home stays than during a traditional semester-long course; at a minimum, the amount should still be higher, all other factors being equal, since the study abroad participants still take three courses during the program that are given in French.

Finally, although this is somewhat stating the obvious, the quality of the L2 exposure was much higher for the study abroad participants than for the traditional students in Group C, and for a number of reasons. First and foremost, any French the students heard outside of the classroom was most likely native, since the study abroad program was conducted on a decentralized campus where classes were held in multiple locations across the city, meaning that the students did not have the opportunity to, say, leave a lecture in the morning and order breakfast in English at a campus cafeteria or meet with advisors who spoke English, which is often the case for programs conducted on foreign university campuses that are directly affiliated with the students' home institution. Second, most of the students' exposure to French outside of the classroom was also in contexts where it was being spoken by native speakers: all but one of the home stay parents had French as their L1, and it was rare for the students to encounter people outside of the tourist attractions in downtown Montpellier who would voluntarily speak English (P.C. with three speakers from Groups A and B) or who were not native speakers. This fact in particular contrasts with extracurricular L2 exposure at the University of Georgia, where private tutoring and conversation groups are the two main avenues students pursue when trying to gain additional experience with the language.

Because language immersion (Krashen 1982; Krashen 1984; Cummins 1998; Birdsong 1999) and study abroad (Freed 1995; Freed, Segalowitz, and Dewey 2004; Segalowitz and Freed 2004) have been shown to aid second language acquisition, and especially the acquisition of non-native segmental phonology and phonetics, longitudinal production data were gathered from the first two groups to test whether these results may be extended to the intonational domain. Although the length of time abroad in both cases is relatively short in comparison to time abroad in other longitudinal studies of language acquisition (e.g. Mennen 2010)—learners in these studies have normally stayed abroad for years or have permanently relocated to the country where the target language is spoken—many of the students showed significant improvement in terms of their production of the L2 segmental and suprasegmental phonology, which are discussed in detail below. This effect may have been due in part to the structure of this particular study abroad program, which required students to speak and be spoken to only in French during all program events, including classroom instruction, field excursions to neighboring cities, and meetings with local staff who were affiliated with their home institution. Students were also required to speak French with their host families, with whom they stayed for the duration of the program, and who in all but one case were not proficient in English, either in terms of speaking or comprehension. By comparison, students in the third group spoke French only in class, which was held for an average of three hours per week, and otherwise did not have many opportunities to practice their skills. Only one set of data was gathered for students in this group, since classroom instruction has been shown to be less effective in promoting L2 acquisition than intensive coursework or full immersion (Freed, Segalowitz, and Dewey 2004). There is a strong correlation between hours spent speaking and gains in oral proficiency (see work by the above authors for an overview of the topic; also, Heliger 1977), so although study abroad is perhaps not the best method for improving acquisition—the crown in this case would go to full immersion courses, where participants are allowed to speak only the target language and no others—it is decidedly more effective than traditional classroom instruction, so the two are appropriate to compare here.

The corpus is perhaps not as big as one would like, but it has the distinct benefit of controlling for the major sources of phonological input in the learners' L2, which to the author's knowledge has not been done with respect to intonation for the combination of languages being studied. The data gathered from students in Groups A and B were compared directly to recordings of native speakers of Southern French, and the data gathered from students in Group C was compared to recordings of native speakers of Standard or Parisian French, which is the dialect most commonly taught at the university (a handful of instructors come from former French colonies, like Haiti, but for the most part they are either native speakers of SF or highly proficient non-native speakers who learned the dialect in college and graduate school). The Southern French L1 data came from two sources: a corpus developed by Dr. Diana Ranson at the University of Georgia containing conversational speech and interviews with residents of Montpellier and Rognes, a town 90 miles east of Montpellier, and the online corpus of the Projet Phonologie de Français Contemporain (PFC) (Durand, Laks, and Lyche 2009), a large repository of guided interviews, questionnaire responses, and conversational speech organized by geographical region. Both corpora also contained socioeconomic information about the speakers, including age, sex, place of birth, education level, and profession. Although the goal of the project was not to conduct sociolinguistic experiments, it was helpful for building a combined corpus of L1 French speech that was most likely to be heard by the students during their stay, since the same information was available for the host families, instructors, and administrators associated with the program. These corpora were supplemented by a small number of recordings gathered from the students' host families in Montpellier, where they lived and took classes for the duration of the study abroad programs.

The most important criterion for the L1 French corpora to meet was that they offer a representative sample of MF, which was generally true, as most speakers spoke with clear Southern accents and produced many of the dialect's most recognizable phonological features, like the 'singsong' intonational contour (Coquillon 2004) and denasalized nasal vowels in word-final position (Coquillon 2006). Although Coquillon (2004) found that native speakers of French were able to differentiate SF from MF based solely on prosodic information, she was unable to determine exactly which tonal categories and pitch contours were responsible for the difference, pointing instead to faster speaking rate and lower pitch level as indicators of the southern variety. Thus, although it would be possible to compare the two varieties in our corpus along intonational lines, a comparison that would certainly be an interesting and welcome addition to the somewhat sparse but growing body of literature on French intonational phonology, our focus instead will be on using them as anchors for exploring the phonetic characteristics of the learners' L2 speech.

3.2 Materials and Procedures

3.2.1 Materials

All participants were asked to complete two tasks: a storytelling task, and a discourse completion task (DCT) as outlined by Prieto (2000). The task has been used to study intonation in French (Delais-Roussarie et. al 2014) and in other Romance languages (Vanrell 2007; Borras-Comes, Vanrell, and Prieto 2014) and has become something of a methodological fixture in L2 intonation research. As compared with other methods, like the map task, where participants must repeat directions after hearing them, or unguided interviews, the DCT is advantageous in that it allows the researchers to systematically manipulate combinations of pitch accents and boundary tones in concert with certain linguistic and paralinguistic variables, like interrogativity, exclamation, surprise, and bias, giving them the option of documenting the entire inventory of tonal categories when thoroughness is required or capturing a controlled sample of them when it is not. Related to this benefit is that because its questions are deliberately structured so as for each of them to include a specific number of pitch accents, the DCT can be easily translated from one language to another without losing its accuracy or validity, making it an excellent choice for studying L2 acquisition in contexts where comparative work with the two L1s is required. It has drawbacks—namely, that participants must be relatively proficient in the L2 in order to comprehend and produce appropriate responses to the prompts—but for the most part, it is an excellent tool for exploring the phonological and phonetic space between two L1 intonational systems.

The DCT used to gather the production data for this chapter consisted of 29 scripted sentences elicited from the participants, which they read in response to situational prompts described to them by the interviewer. The task was shortened from a longer version to include only items that contained tone sequences we were interested in studying. Figure 3.1 contains a sample item from the task in French (a) and English (b). Appendix A contains the full list of sentences for each language.

1.2 De type non neutre

Focalization contrastive

(1d1). Tu entres dans un magasin où la vendeuse est un peu sourde. Tu lui dis que tu voudrais quelques oranges, mais elle ne comprend pas bien et elle te demande si ce sont des citrons que tu veux. Dis-lui que non, que ce sont des oranges que tu veux.

Non ! Ce sont des oranges que je veux.

1.2 Non-neutral

Contrastive focus

(1d1). You just walked into a store where the salesperson is hard of hearing. You say you'd like to buy some oranges, but she doesn't quite understand and asks if you'd like to buy some lemons instead. You say no, and that you wanted to buy oranges.

No! I want to buy oranges.

Figure 3.1. *Sample item from the DCT.*

The main goal of the DCT is to elicit specific combinations of pitch accents and boundary tones from the speakers, with an emphasis on their ability to correctly interpret the semantic and pragmatic content of the situations presented by the items and respond appropriately, i.e. by producing the scripted response with some form of grammatical intonation. The DCT is not perfect in this sense, however, as for some items, there was more than one possible pitch contour they could produce, and pitch range in the contours varied widely depending on how animated the speakers were in interpreting the scripted responses. The English sample item from above is a good example of this situation: the pragmatic context requires two main H* pitch accents, one on *No!* and one on *oranges*, but the relative height to which these tones are scaled can vary significantly. On one hand, speakers could scale the H* associated with *No!* the highest, in effect contradicting the imaginary shopkeeper's entire statement, but on the other hand, they could choose to scale the H* on *oranges* higher and contradict only her suggestion that they were interested in buying lemons. The difference between the two is somewhat subtle, but both versions would be acceptable as long as there was at least an H* on *oranges*, which in both cases would need to be emphasized in the response. Speakers are faced with similar choices throughout the DCT, but in general, they do not affect target contour's basic structure.

31 sentences were recorded per speaker in the first task and approximately two minutes of extemporaneous speech in the second. Learners completed the tasks in both languages, while the host families completed them only in French. In total, 40 sets of responses were obtained from the learners (20 speakers x 2 languages) and 8 from the families. The literature does not mention an acceptable minimum number of speakers needed for the DCT to produce valid, generalizable results, but the data below indicate that the number used here was enough.

3.2.2 Procedures

For the discourse completion task, the interviewer read the situational prompt for each item, and the participants read the target sentence in response. In order to ensure that the speakers' best efforts were obtained for phonetic analysis, they were allowed to repeat their responses if they were unsatisfied with their pronunciation of the words or interpretation of the syntactic or pragmatic information, like placement of the final accent in relatively long or complex phrases. From a purely scientific point of view, this means that the eventual data present a somewhat skewed picture of the participants' spoken proficiency, but from a practical point of view, it means that the data are more likely to represent the speakers' best efforts, i.e. the phrasing and contours that most closely match their mental representations of the given utterances in French. Many of them were prone to making errors in pronunciation or reading, which greatly increased the number of intonational disfluencies in the original data; allowing them to repeat items helped smooth out the contours that were eventually compared to the L1 data.

For the storytelling task, participants were asked to tell a story about a social event they recently attended, like a family gathering or party. The stories were between two to three minutes long, depending on each participant's speech rate and general comfort with the task. The task served two main purposes. First, it provided examples of how each speaker realized major and minor continuations in French, which were scarce in the first task due to the scripted and relatively short nature of the items in the DCT. Second, like the DCT, it tested their ability to produce complete sentences in French with appropriate phrasing and tonal structure. Unlike the items on the DCT, however, which the participants often had trouble pronouncing on the first reading, the open-ended structure of the storytelling task gave them more control over the syntactic and phonological structure of the utterances, providing a more accurate assessment of their L2 phrasing. The tradeoff of course is that these data could not be used to compare the participants' phrasing to that of native speakers, since it was impossible to control for the number of phrase boundaries in the speech samples.

For participants in Groups A and B, the tasks were completed twice, once in an interview session before they left for France, and once when they returned; participants in Group C completed only one session. The sessions generally lasted around 30 minutes, although participants were encouraged to take as much time as they needed to complete the tasks, the idea being to prioritize quality of their responses over their speed in reading the DCT items and generating ideas for the storytelling task. The two tasks were were separated by approximately five minutes of time during which the participants could review the instructions for the upcoming task. Instructions for both tasks were printed on a prompt sheet and were explained by the interviewer at the beginning of the session.

3.2.3 Preprocessing

Speaker responses to both tasks were recorded using a Shure SM51 condenser microphone and digitized at a sampling rate of 44.1kHz and a depth of 24kbit/s. Responses to the DCT were segmented by hand and saved as separate files, yielding a total of 580 (29 items x 20 learners) phrases of varying length and tonal structure. Responses to the storytelling task were saved as a single .wav file. Sound files for both tasks were then automatically segmented into words, syllables, and phones using the Speech Phonetization, Alignment, and Syllabification (SPPAS) toolkit (Bigi and Hirst 2012), a labeling program trained on native English and French data. Although SPPAS occasionally mislabeled segments in the L2—it had not been trained on learner speech, so it was forced to treat the segments as belonging to one of the L1s—the segmentation itself was very accurate. The segmented files were saved as TextGrid files in Praat (Boersma and Weenink 2014) and were checked by hand to ensure syllable and phone boundaries were correctly aligned with the corresponding spectrographic data.

Pitch curves were generated using Praat's built-in pitch tracker, and files containing F0 perturbations or disfluencies were discarded; this only applied to two of the learners, who recorded their responses to the questionnaire at home on microphones with lower-than-optimal signal-to-noise ratios (i.e. the sound files were too noisy for the pitch tracking algorithm to produce smooth curves). The program's smoothing feature was used to produce clean pitch contours for analysis. Several intonational features were then automatically labeled for each syllable using a script, including pitch minima and maxima, peak alignment, and position of the elbow in non-level contours, which was based on an algorithm designed by Welby (2003; 2004; 2006) to automatically detect the turning point in the contour. Labeling the elbow by hand is possible but subject to significant variation (Welby 2008), so the script was used in order to make the labeling as accurate as possible. Following this initial analysis, AP and IP boundaries were marked in the French data to allow for the comparison of L1 and L2 phrasing. For the L1 French speech, phrase boundaries were marked using two criteria: the length of the pause between segments, as in the ToBI labeling system, and the syntactic well-formedness of the phrase. As a default, an AP was marked for combination of a content word and its surrounding function words and was expanded only if there was clear phonetic evidence that it comprised more words, e.g. in cases of simplified phrasing during rapid speech. For the L2 speech, boundaries were marked using the procedure detailed in Section 3.3 and were based solely on phonetic information rather than the actual syntactic structure of the utterance, with the intention being to yield the most accurate phrase counts for each speaker.

Numerical data were exported to a spreadsheet and then analyzed using the R statistical package (R Core Team 2014), with the most important calculations being as follows. Phrase breaks were marked in the learner data obtained from the DCT, and the late L-elbow in phrase-final rises was identified using a Praat script. Mean pitch level and pitch span for all of each speaker's utterances were calculated in ERBs, as were mean level and span for the three language varieties, L1 English, L1 French, and L2 French. The same statistics were calculated for the tonal categories presented in Sections 3.4, and z-score normalization was used to estimate the mean size of specific pitch excursion relative to each speaker's range. All of the scaling data were obtained using a Praat script that measured time, pitch, and intensity at 10ms intervals and wrote the results to an external file.

3.3 Phrasing

The basic goals of this section are two-fold: to determine whether the learners could produce well-formed intonational phrases in French, and to assess whether the phrases they did produce support the learner variety hypothesis. As mentioned in Chapter 1, English and French are very different in terms of their tonal structure, although they do share a number of important pitch accents, which are the subject of Sections 3.4 and 3.5. With regards to phrasing, the most important difference between the two languages is the structure of the lowest-level intonational unit, the pitch accent in English, and the AP in French. Although the AP is composed of individual tones, it is generally realized with only one major pitch movement, which is normally either a rise (ip-final) or a fall (IP-final). As a result, two or more words may be strung together with no recognizable changes in F0, and in rapid speech, APs themselves may become linked to

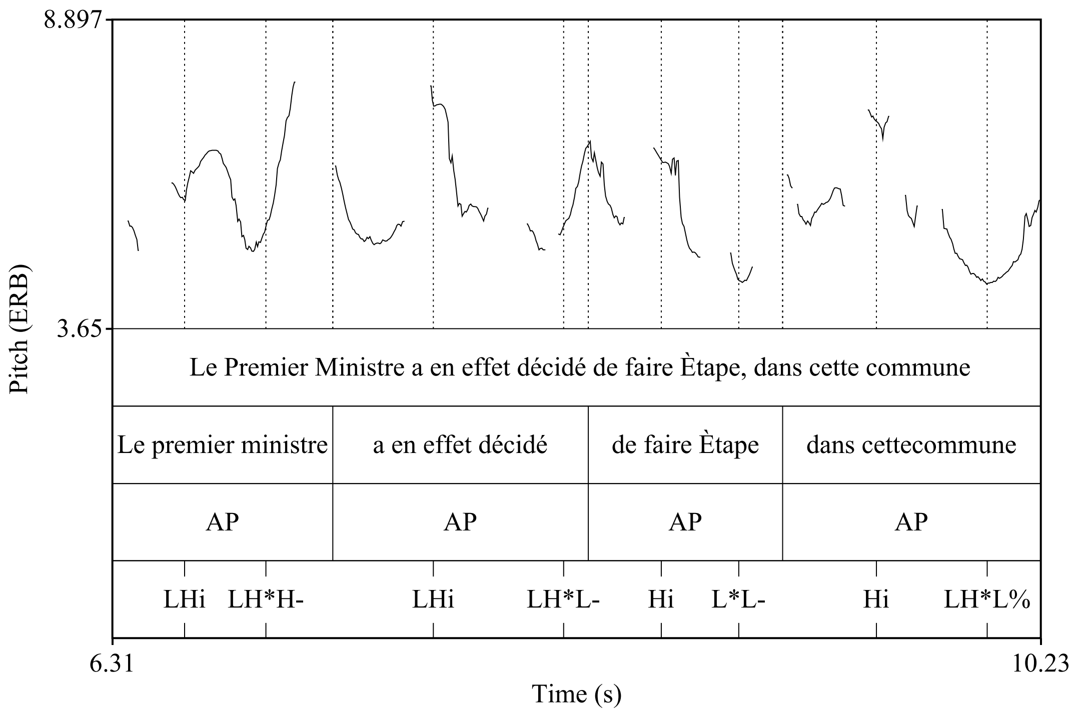


Figure 3.2. *Sample phrasing in L1 French.*

such an extent that the final rise is reduced not only in duration, but also in amplitude. Figure 3.2 above shows an example of this phenomenon, which although a well-established feature of French intonation, is worth mentioning here to establish a point of comparison with English. The phrase *Le Premier Ministre a en effet décidé de faire Étape* 'The Prime Minister has indeed decided to make a stop' is divided into three APs, each with a noticeable rising movement at the right edge. In slow or very careful speech, which native speakers commonly produced during the DCT used for this study, the phrasing would change in two subtle but very important ways: first, there would very likely be AP boundaries after each of the content words rather than after only a few of them; and second, all of the pitch movements would have higher amplitudes, and the boundary tones would have longer durations. Both changes would stem from the speaker's more deliberate pronunciation, which as suggested by the Effort Code (Gussenhoven 2002) tends to increase the prominence of prosodic events. Under these conditions, the intonational phrasing of French and English is similar, at least on the surface, with the IP being broken into smaller units comprising roughly one content word each. One of our experimental hypotheses stems from this fact, then, and is that we expect the learners to have some success in matching the average number of boundaries produced by the French L1 speakers in the DCT. Unfortunately, this was the only hypothesis that was practical to test, since there is no established procedure for measuring the number of phrase breaks in faster, unscripted speech. The results described below show that the hypothesis is not necessarily true, even considering the relatively rigid structure of the DCT, giving some insight into difficulty in acquiring L2 intonation and supporting the suggestion that learner varieties are indeed unique.

3.3.1 Methods

There were several possible methods for comparing native and non-native French phrasing, but the most practical seemed to be counting the number of APs produced during the DCT, which because of its being scripted contained a predetermined number of possible phrases. Another option would have been to derive a statistic for describing the average length in words of each speaker's AP, or perhaps the average number APs per utterance, but neither has any real precedent in the literature, and it is not clear whether the results would be consistent enough to be useable, especially for spontaneous (i.e. unscripted) speech, where utterance length and speaking rate are largely determined by individual factors. Jun and Oh (2000) avoided this issue in much the same way by having speakers complete a 40-item task similar to the DCT that was designed to elicit different combinations of pitch accents and boundary tones, so the method seemed appropriate to adopt here. The only remaining difficulty was that of determining exactly what should constitute a phrase boundary in the learner speech, an issue that Jun and Oh do not address in their article but likely encountered at some point during their study. In the French L1 data, only two criteria were needed to determine phrase boundaries: a rising pitch movement, and some degree of syllabic lengthening. As long as these were met, there was a good chance that the pitch accent was followed by a break and could be considered a boundary tone. In English L1, the criteria are more or less the same, with the exception that the pitch movements are more often falls or complex movements than in French; phrase-final lengthening is still common, though, and rises were very often used to mark continuation and interrogativity, just as in French. However, phrase boundaries occur much less frequently in English than in French, mostly because the building blocks of English intonation, the pitch accents, are relatively small (i.e. they have a more limited prosodic domain) in comparison with the French APs.

Based on these facts, phrase boundaries were marked in the L2 speech when at least one of the following criteria were met:

1. Rising, falling, or complex boundary tone used in either French or English
2. Pitch accents not used in French, e.g. bitonal or scooped accents

3. Segmental lengthening as indicated by normalized vowel duration

The first criterion is somewhat obvious but necessary to state nonetheless: whenever speakers produced a recognizable boundary tone from either of the L1s, a phrase break was marked. Boundary tones in both languages are phonetically distinct from pitch accents in terms of their amplitude, rising to either the top or falling to the bottom of the speaker's range, and they are always paired with a pitch accent (since they associate with the boundary itself and not the preceding syllable), so they are relatively easy to identify. The second criterion stems from what distinguishes the two languages rather from what they have in common, namely that English has bitonal pitch accents and French does not. If a speaker produces one of these pitch accents in L2 French, then it can have only one of two possible functions: marking the end of an AP, as was often the case for the scooped L*+H accent, which the learners struggled to differentiate from the French LH*, or simply performing one of the syntactic, semantic, or pragmatic functions of the same pitch accent in English. In both cases, we can mark a phrase boundary, and the only real difference between the two is that the former is phonologically grammatical while the latter is not. The last criterion is based on the simple fact that both languages have phrase-final lengthening, so whenever a word-final syllable lasted for longer than two standard deviations above the speaker's mean duration, a boundary was marked. The choice of the two-sigma level was arbitrary, but it lowered the chance of marking a boundary where one did not exist, and it was easy to implement after exporting the syllable durations from each speaker's DCT and normalizing them in R.

By comparison, phrase boundaries were much easier to detect in the French L1 data, with speakers tending either to mark most APs with a phrase-final rise and/or lengthening in slow speech, or to string multiple APs together and mark only ip and IP boundaries in rapid speech. In both cases, the first and third criteria were used above to determine whether or not a boundary tone was produced, but not conjointly, since speakers did not always produce phrase-final pitch movements large enough to be distinguished from phrase-medial movements based on amplitude alone, in which case final lengthening was taken as evidence of the underlying boundary. In general, though, speakers tended to produce recognizable boundary tones after most of the APs, since the DCT is structured in a way that encourages careful reading and pronunciation.

3.3.2 Results

Data from the DCT show that learners produced significantly more sentence-medial AP boundaries than native speakers, with boundaries being marked for actual phrases and for lone pitch accents, as described above. Mean counts for both boundary types are presented in Table 3.1 below and organized by speaker group, with the left column showing averages for the French L1 speakers who completed the task (only 10 were available to interview in person), and the right column showing averages for all three groups of the learners. Values for the French L1 speakers were tightly clustered ($\sigma=3.7$) and are thus reported as an aggregate mean, but values for the L2 speakers were more widely dispersed ($\sigma=11.38$). A Welch's *t*-test indicated that the difference between the means was statistically significant ($p<0.0002$), supporting Jun and Oh's findings for Korean that non-native speakers of French are generally less successful than native speakers at grouping words into phrases. The *t*-test was used here instead of ANOVA since data were only being compared between two of the three possible language varieties (the third being L1 English). The Welch's test also assumes unequal variances between the two groups, so the issue of the learners' phrasing being more variable did not impact the similarity score.

Table 3.1. *Number of sentence-medial APs produced by French L1 and L2 speakers in the DCT.*

Speaker	French L1	Speaker	French L2
1	78	1	85
2	77	2	117
3	80	3	108
4	73	4	115
5	74	5	96
6	82	6	94
7	75	7	100
8	71	8	112
		9	140
		10	93
		11	88
		12	112
		13	99
		14	101
		15	90
		16	85
		17	135
		18	112
		19	106
		20	88
μ	76		96

The majority of the errors met the second criterion above, as the learners often produced pitch accents on syllables that would otherwise be either completely unaccented or only marked a very small pitch movement leading to the larger phrase-final rise. In one way, we can think of this as an issue of pitch scaling and alignment, since if the final H* is aligned far too early (like on the penultimate full syllable), or the phrase-initial LH is scaled too high, then the resulting pitch accent would effectively be ungrammatical in French. Figure 3.3 (a) on the following page illustrates the point nicely. The phrase *Peut-être que ça ne lui plaira pas* 'Maybe he won't like it' is categorized as a doubtful declarative in the DCT and can be divided into two main APs: one ending after *être*, and the other ending after *pas*. The initial rising pitch movement in this example is appropriately scaled given the focus on the modifier *peut-être,* but the learner produced phrase boundaries after *ça* and *lui,* which being centered around function and not content words are both ungrammatical. Although pitch appears to decline evenly in the last two phrases, which might make us consider grouping them together, the length and level pitch in *plaira* are more characteristic of a downstepped H* in English than of the AP-initial /LH…/ in French, meeting the second criterion described above. We can also see evidence of the same pattern on *ça*, where despite the initial fall from the H* associated with the end of the first phrase, pitch remains level. It is worth mentioning, however, that the rise associated with *lui* is well-formed, with its alignment and scaling close matching what we would expect from a native speaker. For the sake of comparison, we can consider the second pitch contour in Figure 3.3 (following), which shows the same sentence produced by a French L1 speaker. This time, the utterance is organized into three phrases, with the first two being marked by high or rising pitch movements, and the last being marked by a level low tone, presumably the combined result of downstep and a L% at the IP boundary. The pitch tracking algorithm had some trouble with the

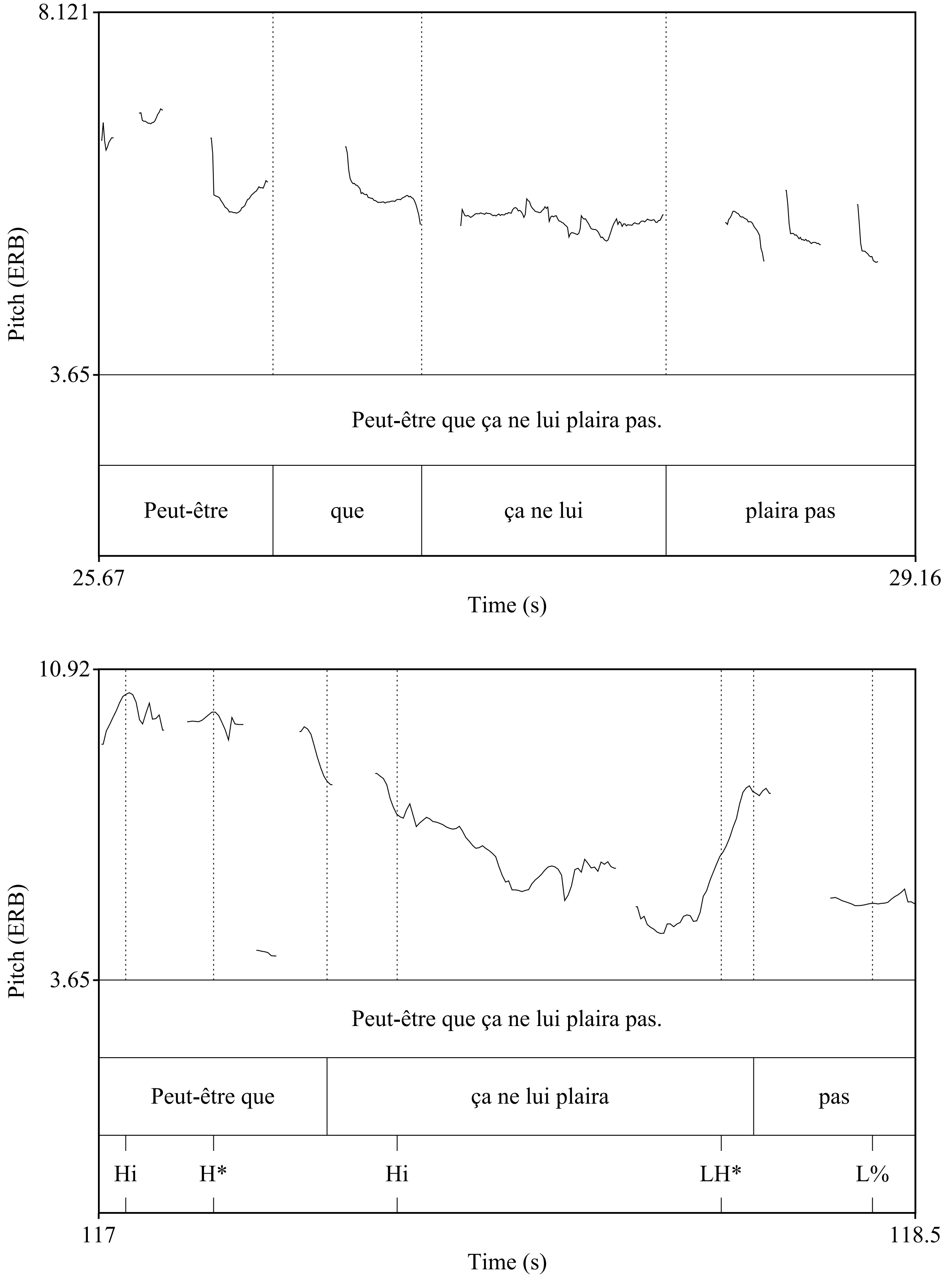
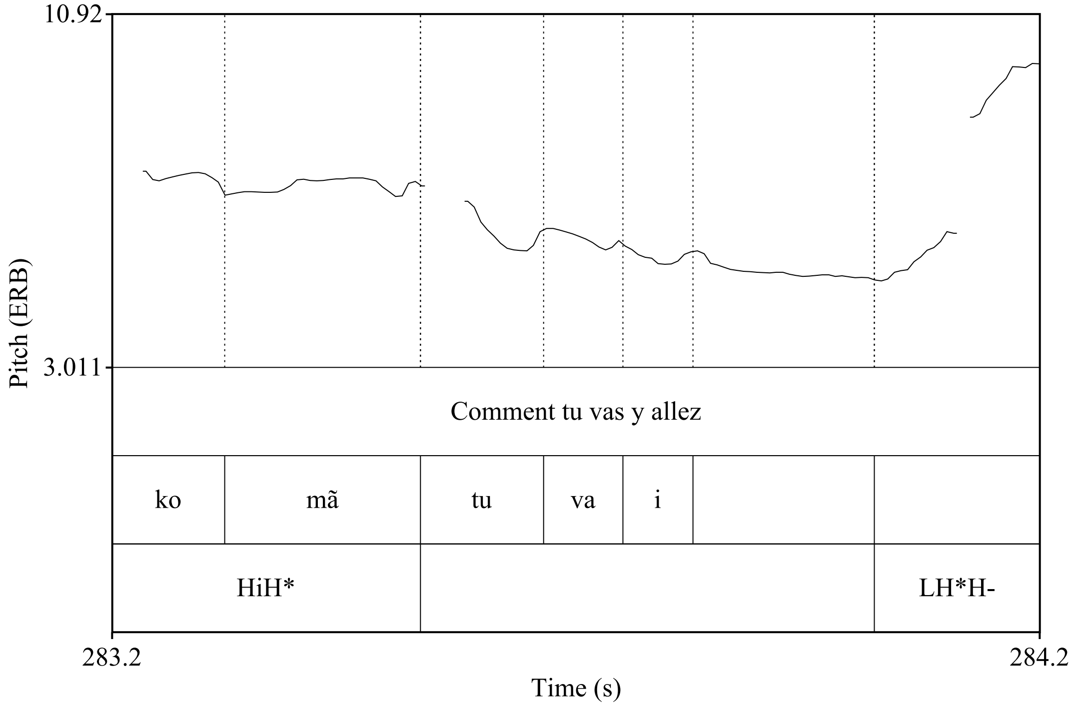


Figure 3.3 *Phrasing in L2 (a) and L1 (b) French*.

first AP, but perceptually, the apparent H* accents associated with *Peut-être* sounded like truncated rises rather than the level peaks produced by the learner. The rise after *plaira* is very clear, however, and meets both the criteria for a phrase boundary, given the subtle but significant lengthening on the final syllable *-ra*. The slope is perhaps steeper than what we would find in casual speech, as the speaker in this example was speaking very deliberately, but the rise provides a beautiful example of how clearly phrase boundaries can be marked in L1 speech.

The two contours in 3.3 also provide supporting evidence for Jun and Oh's (2000) idea that marked tonal categories are easier for learners to acquire than unmarked categories. The H* accent produced by both speakers at the beginning of the utterance places broad focus on the phrase *Peut-être* 'maybe', indicating doubt about the following proposition. French and English both code this information with a high tone, but it surfaces in French as a modified realization of the default LH* rather than a level H* or an H*+L fall, which are both common in English.

The final pitch movement in both of these examples comes at the end of an IP, but learners also struggled to produce native-like phrasing in smaller units, like the ip. The pair of utterances on the following page shows a native (3.4a) and a non-native (3.4b) pronunciation of the phrase *comment vas y aller* from the larger utterance *Ou est-ce que tu vas, comment vas y aller, et à quelle heure tu vas rentrer?* 'Where are you going, how are you getting there, and when will you get back?'. The utterance contains three ips, each being separated by clear syntactic boundaries, and provides examples of the major continuation contour, which is larger in duration and amplitude than the minor continuation contour. The basic structure of the contours is the same in that they start high on *Comment* 'how', the result of its receiving broad focus, and then decline throughout the phrase until rising in response to the H% boundary tone at the final



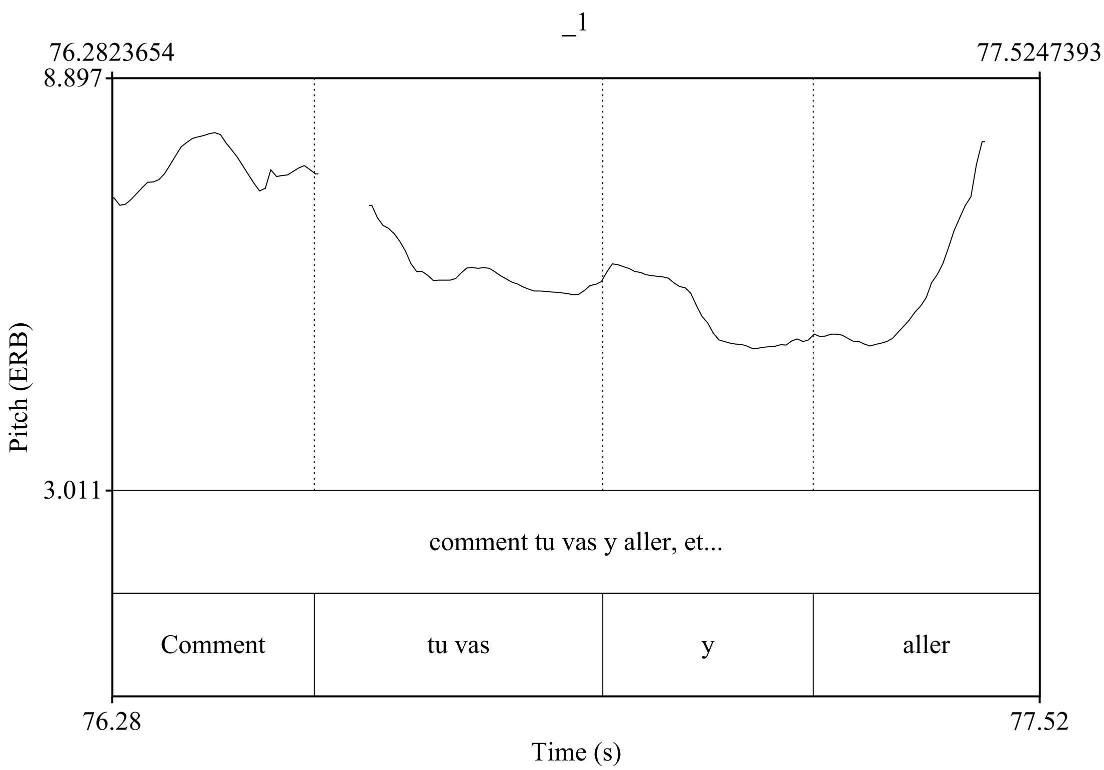


Figure 3.4. *Phrasing in L1 (a) and L2 (b) French*.

boundary. Another point, which we will discuss in Section 3.4 below, is that the final rise in (b) is delayed, beginning well after the onset of the vowel, a very common characteristic of the learners' L2 intonation. With regards to phrasing, however, there are places where pitch increases in (b) but either stays level or falls in (a). Despite the anchoring effect of the movements at the ip's edges, the learner still produced English-style H* pitch accents on *tu* and *y*, effectively grouping the words into separate APs. The accents are downstepped, of course, and there is the suggestion that they have been pulled away from their full potential height by the flatter French contour, but these are both issues of scaling rather than of phrasing. The main point here is that there are clear differences in phrasing between the L1 and L2 speech that very likely stem from the two languages' differing intonational phonologies.

Learners in the two study abroad groups improved their phrasing over time, although none of them achieved native-like proficiency by the end of their programs. Some of the improvements were modest and not statistically significant—two speakers produced fewer phrase boundaries after the trip, but still too many for their counts to fall in line with those expected for L1 French—but a handful of learners made genuine progress. Figure 3.5 (following) shows the pitch curves for the sentence *Peut-être que ça ne lui plaira pas* 'Maybe he won't like it' produced by one of the speakers in Group A, which spent the longest amount of time in Montpellier, both before (a) and after (b) the trip. The contours have some features in common with the L1 French contour shown in Figure 3.4 above, most notably the initial H* on *Peut-être* and the decline in pitch throughout the phrase, but (a) shows evidence of intonational transfer from English. Based on the three criteria given above, we can mark boundaries after all but two of the words in the sentence. In the second contour, however, the rises are confined to

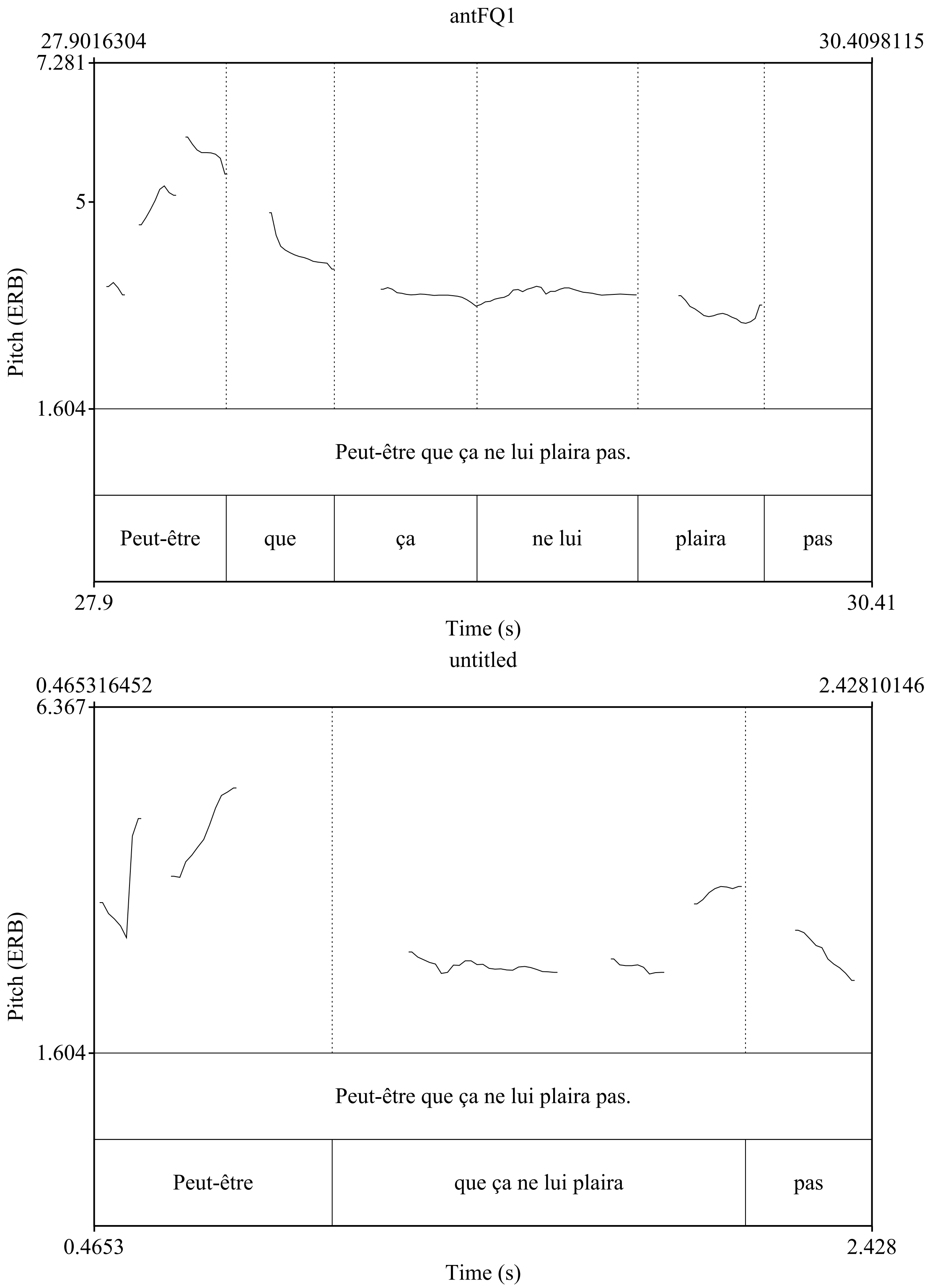


Figure 3.5. *Pre- and post-program phrasing in L2 French*

their two appropriate environments, e.g. the initial movement after *Peut-être* and the final one after *plaira*, which is nearly identical to the phrasing produced by the native speaker above.

3.3.3 Discussion

Following Lee (2014) and Jun and Oh's (2000) work on L1 English and L2 Korean, the results above suggest that learners are able to acquire French intonational phrasing, but only with varying degrees of success. Although each speaker's phrasing was subtly different, two primary trends emerged from the L2 data:

1. IP- and ip-final phrase breaks were marked more frequently than ip-medial AP-final breaks; and
2. Sentence-medial AP-final phrase breaks were marked more often with either final lengthening or an English pitch accent than with the appropriate tonal category in French.

The first trend falls in line with Jun and Oh's (2000) findings that learners had more success grouping words into phrases when the phrase boundaries were somehow marked, for example phonetically with a H*, or semantically by the addition of interrogativity. This seems to have been the case for learners in all three of the study groups as well, as they produced phrase-final rises (instead of English-style pitch accents) more often in questions (IP-final) and in continuation (ip-final) than in sentence-medial APs. There were clear differences between learners and the native French speakers in the phonetic realization of these phrase boundaries—these are discussed below in Section 3.4—so we cannot say that they acquired the contour in full, but we can at least say that they succeeded in parsing the sentences appropriately under these particular conditions. By contrast, the unmarked or default phrase structure was very difficult for the learners to produce. Many of them also realized constituents receiving focus, both broad and narrow, as isolated words with H* pitch accents instead of APs with a lengthened or heightened final rise, which was typical in the L1 French data.

Regarding the learner variety hypothesis, the data neither support nor refute the claim that the phrasing the learners produced belonged to a common intonational system distinct from the two L1s, mostly because it is hard to make a real comparison using a metric as simple as mean phrase count. They do seem to offer some support in favor of the learners transferring parts of their L1 phrase structure to the L2, however, especially considering examples like the contour presented in Figure 3.4b above, which shows rather clear evidence of English-style pitch accents forming the phonological basis of the contour in French. One reason the data may not be as convincing as they could be is that no study has tested the learner variety hypothesis with L1 speakers of English—rather, they have focused mostly on L1 speakers of other languages learning English as an L2. Using Mennen, Chen, and Karlsson's (2010) study as an example, it is possible that the learner variety they posited for L2 English is based not on a universal underlying tonal inventory but rather on tonal categories in the L2 that perceptually overlap with those in the L1. Although the corpus used here is smaller and was gathered over a smaller period of time than the one they used, our data seem to support this analysis, since the learners used complex pitch accents in their L2 French, which was rare for the varieties of L2 English in their study. This possibility is discussed in detail in Chapter 5, but it is worth mentioning here because it will bear on the results of the statistical analysis in Chapter 4, the main point of which is to test the learner variety hypothesis on strictly phonetic data.

3.4 Tonal Alignment

This section examines tonal alignment in phrase-final L2 French rises. As above, the main goal of this section is to determine the extent to which English intonational phonology influences the learners' realization of French tonal categories, only the focus now is on a few specific syntactic domains: ip and IP boundaries.

3.4.1 Background

The two most prominent phrase-final rises in French are the continuation contour and the question contour. The two rises are phonologically identical, comprising an AP, again with the default rising tonal configuration proposed by Jun and Fougeron (2000), and a single H% boundary tone. The two rises differ in excursion size and duration, however, with the question contour exceeding the continuation contour in both dimensions, and the two presumably being perceptually distinct (DiCristo 1979). As in English, the question contour is one of the most extreme pitch movements in French, rising from near the bottom of a speaker's range and potentially climbing all the way to the top. Variation in the prominence of the rise depends primarily on pragmatic and paralinguistic factors. From a phonetic point of view, the continuation contour is somewhat more interesting, however, in that its form and function are an extension of the unmarked AP-final rise, which grows in duration and prominence to signal the speaker's intent to hold the floor. The actual size of the rise relative to the speaker's range depends primarily on speaking rate and emotional state (see data below), but it may take on a relatively large range of values, making it a challenge for learners to produce consistently and accurately. Because of these two features—the rise's importance to French intonation and its predictable but wide-ranging variability—the continuation contour is an excellent litmus test for learner success in acquiring the language and will be the focus of the discussion below, although the question contour will be briefly discussed in Section 3.4.

From Delattre's (1963; 1966) ten fundamental melodies of French to the more recent accounts of French intonation within the AM framework (Post 2000; Jun and Fougeron 2002; Delais-Roussarie et al. 2013), the continuation contour has been consistently described as a moderate rise, beginning near the bottom of the speaker's range and climbing close, but not all the way, to the top (the latter is characteristic of yes/no questions and is discussed in Section 3.5). The pattern is generally the same for English (Beckman and PIerrehumbert 1986; Grover, Jamieson, and Dobrovolsky 1987; Chen 2003), although Delattre (1963; with Poenack and Olson 1965) does suggest that continuation may also be marked by a fall. Delattre's claim is not completely unreasonable—speakers often produce a H*+L as part of the contour, which creates a pronounced dip in F0—but the pitch movement near the boundary is almost always a rise and usually transcribed as an H% preceded by a L- phrase accent. Regardless of the pitch accent leading the nuclear configuration, then, the final movement is realized as a low- to mid-rise, with height depending on speaking rate and paralinguistic factors.

In both languages, the amplitude and duration of the continuation contour correspond to the strength of the following syntactic and/or prosodic boundaries. In French, continuation appears mostly before ip boundaries, where speakers pause before beginning the following phrase in all but very rapid speech, but it may also surface before AP boundaries in particularly careful or hesitant speech. In general, smaller rises are common before ip-medial AP boundaries, and larger rises are common before IP boundaries in questions. The ip-final LH* falls somewhere in between these two extremes and can be seen as the canonical realization of the underlying /

LHLH*/ tonal configuration of the AP, as it is neither small enough to be overlooked nor large enough to be mistaken with the LH*H% rise used in questions. Stemming from this structure, the most important contrast between continuation in the two languages it that the rise itself begins quickly in French but may be delayed in English, especially in cases where the nuclear configuration includes an H* on the final syllable, squeezing the other two tones (L-H%) toward the phrase boundary. In terms of alignment, the elbow between the L and H tones in the English contour is delayed in much the same way as it is in the L*+H pitch accent, which is importantly absent from French. These differences in tonal inventory create a significant problem for English L1 speakers in that they must learn to initiate the rises immediately in French rather than allowing them to gravitate toward the right edge of the phrase. As we will see below, this was indeed an issue for the learners in all three of our groups, and statistically speaking it was the most consistent error they made.

3.4.2 Preliminary Results

Following Welby (2004; 2005; 2006), whose research shows phrase-final rises in French consistently align closer to the onset rather than the middle of their syllables, the goal of this section was to explore whether the learners could produce continuation contours with appropriate alignment in their L2. Welby's work focuses on two main anchoring points in the default / LHLH*/ contour: the elbow of the late L, and the peak of the late H. Continuation contours in both French and English end in an H% boundary tone, so the alignment of the final H in both cases should be the same, i.e. it should coincide with the right edge of the phrase, with the potential to bleed over into beginning of the following phrase in IP-medial positions. This trend was borne out by the data gathered from the learners, and so the following discussion focuses instead on the alignment of the late L elbow, which shows more variation in the L1 French data (Welby 2005), and which poses more of a phonological problem for the L2 speakers.

Results from both tasks show noticeable differences in the phonetic realization of rising tone sequences between the native and non-native speakers. In general, the L2 speakers realized the elbow in the late rise later than the L1 speakers. The placement of this elbow was also more variable. There were, of course, many times when the learners aligned the phrase-final H* with the wrong syllable entirely—these cases are worth discussing briefly, if only to show that the learners made phonological errors aside from those relating to phrasing discussed above. Figure 3.6 on the following page shows the pitch curves for a sentence-initial AP produced by a native (a) and a non-native(b) speaker reciting the days of the week, with time shown in seconds and pitch in ERBs. The pitch curve shows a slow rise in the first syllable followed by a level H in the second. In contrast with the curve in (a), there is no sharp phrase-final rise marking the ends of the AP and the ip, and there is no evidence of final lengthening. Although the rise itself is aligned appropriately, beginning immediately after the vowel onset, it is associated with the wrong syllable and is thus ungrammatical. Under certain pragmatic conditions, and initial LH rise would be possible on the first syllable, but either an H* or a full rise would also need to appear on the second syllable to signal continuation (Delais-Roussarie et al. 2014). We could potentially analyze the tone on the second syllable as a downstepped H*, but we would need a L% boundary tone for that to work, which is also ungrammatical given the continuation. There is no clear method for quantifying these kinds of phonological errors, so the analysis below will focus on phonetic errors we can measure—specifically, the delay in the low-L elbow from vowel onset

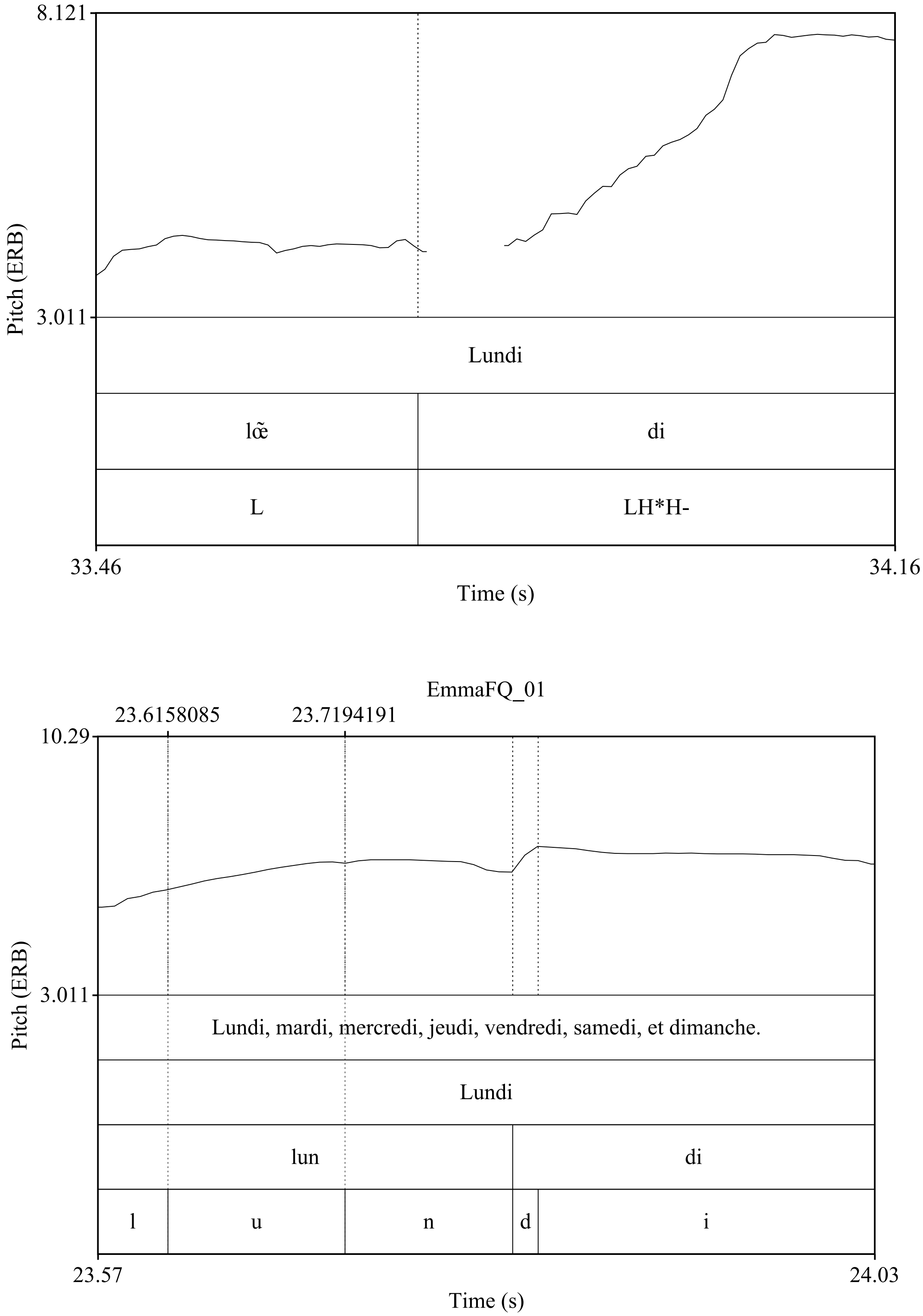


Figure 3.6. *Tonal alignment in ip-final rises in L1 (a) and L2 (b) French*.

that characterizes the learner intonation. Any complete account of L2 intonation must include these characteristics, however, so Chapter 5 presents a number of potential solutions to the problem of measuring and interpreting them with respect to the current theories of L2 acquisition and intonational phonology.

Returning to the issue of tonal alignment, the pitch curve in Figure 3.7 is representative of the learners' typical alignment pattern for late rises, with the L elbow being located near the middle of the final syllable and the H target being reached near its end. The target phrase is taken from the sentence *Qu'est-ce que tu lui offriras*? 'What will you offer him?' and provides a very clear example of the characteristic delayed rise, with the slope of the curve increasing sharply about halfway through the syllable. The contour also has a slight dip in pitch following the final

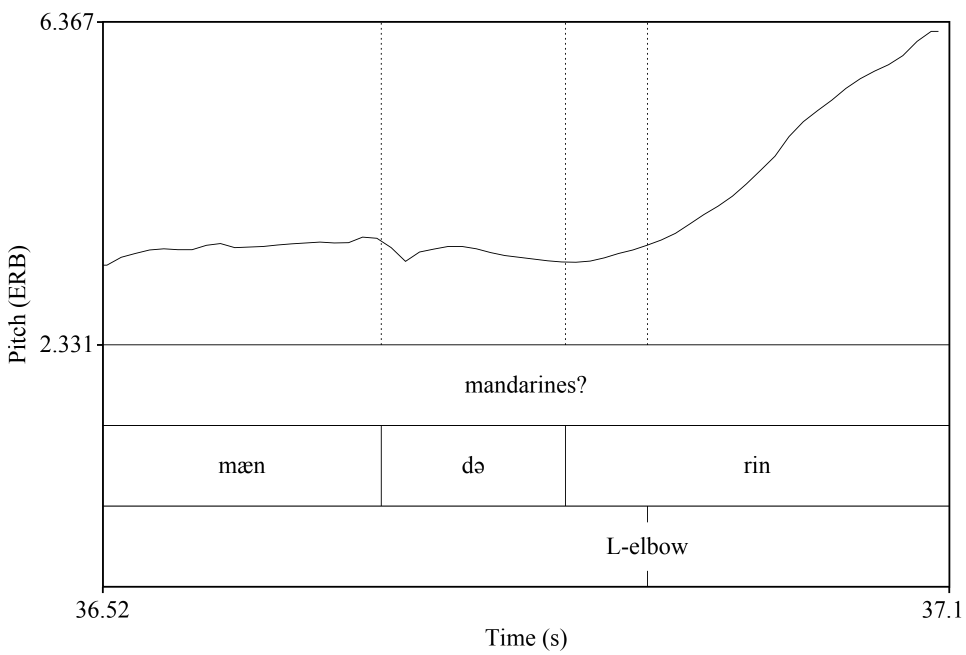


Figure 3.7. *Typical tonal alignment in L2 French*.

H*, which was uncommon in the L1 French data, where most rises continued throughout the final syllable. This pattern was very common in the L2 speech, accounting for 45% of the total rises, and was in many cases produced with an even more pronounced delay than the one above. For the sake of comparison, we can consider the contour in Figure 3.6 (a), which shows earlier alignment of the L-elbow and a more even slope to the rise. There are likely several issues going on here—one of which is phonological and discussed below—but their effect is clear in that the contour has a pronounced L segment before climbing to the H*. In English, the contour would be transcribed as a L*+H pitch accent followed by a H% boundary tone, and using an ad-hoc ToBI system for the L2, the result would likely be the same.

Although French does not have lexical stress, we see further evidence of this phonological transfer from English in the relationship between intensity (i.e. loudness) and pitch in the L2 rises. Taken from the question *Vous avez des mandarines?* 'Do you have any mandarins?', the contour in Figure 3.8 shows pitch and intensity for the phrase-final *mandarines*. At the onset of the vowel, pitch is low, and intensity is high, both phonetic indications of an underlying stressed syllable and corresponding L* pitch accent (both parameters are phonetic correlates of stress in English). As pitch rises, however, intensity drops, and it almost bottoms out before the end of the syllable. In L1 French, the H* and not the L* would be the tonal target for the rise, so intensity would remain fairly level until the peak was reached. The difference between the expected and the actual contours is striking, and it provides strong evidence in favor of analyzing the delayed rise as an instance of phonological and not just phonetic transfer from the L1 to the L2.

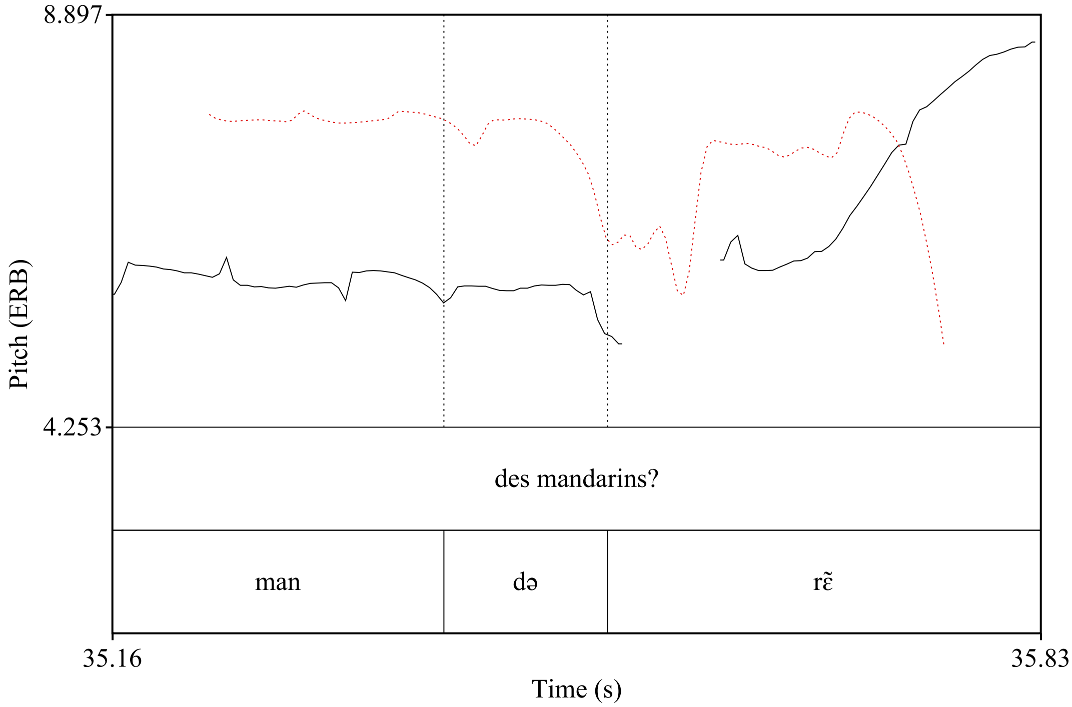


Figure 3.8. *Pitch (solid) and intensity (dotted) in an L2 IP-final rise.*

Table 3.2 (following) presents the mean alignment in ms of the late elbow relative to the onset of the vowel in the accented syllable for both varieties of French (i.e. L1 and L2). Data for the three L2 groups are grouped in a single column, as are the data for the two L1 corpora. As mentioned earlier, vowel onset was used instead of syllable onset to account for within- and between-language variation in the length of onset consonant clusters, which would very likely have skewed the alignment data. In calculating the means, positive values were entered when the elbow was positioned after the vowel onset, and negative values were entered when it was positioned before the syllable onset, again with onset consonants being removed from consideration. Following the pattern described above, the French L2 values show late mean alignment relative to the French L1 values, which are closer to the syllable onset and similar to those reported by Welby (2005; 2006).

Table 3.2. *Latency in* ms *of the L-elbow in AP-final rises for French L2 and L1 speakers*

Speaker	French L2 μ | σ		French L1 μ | σ	
1	32.1	39.3	10.5	8.3
2	51.6	58.6	15.6	10.3
3	45.3	33.2	8.5	9.6
4	39.8	37.4	21.3	11.4
5	52.5	49.3	12.3	9.3
6	61.8	56.7	11.2	14.2
7	33.1	36.5	9.3	16.3
8	71.0	58.8	18.1	15.8
9	45.2	43.1	2.3	10.2
10	83.1	76.3	7.8	8.2
11	27.5	23.1	3.1	5.9
12	28.0	30.2	5.2	10.3
13	35.6	30.0	10.1	7.3
14	45.4	37.7	8.0	9.8
15	23.7	31.4	11.3	5.5
16	75.1	60.3	15.9	6.1
17	33.8	30.2	20.2	15.4
18	49.6	50.0	2.8	9.0
19	45.2	43.9	12.9	5.6
20	56.0	56.0	3.3	9.8
All	45.7	42.7	8.6	8.6

It is important to remember that although a late rise has been proposed for French (Delais-Roussarie et al. 2013), the vast majority of rises begin immediately after vowel onset—even in cases where the syllable onset comprises one or more voiced consonants, as in the words *moi* [mwa] ‘me' and *vouer* [vwe] ‘to vow’, the rise is delayed until the beginning of the vowel (Welby 2005). Statistically speaking, one trend that emerges from these data is the close relationship between the mean and standard deviation for the L2 speakers’ alignment patterns. In effect, the distribution was approximately bimodal, suggesting the speakers implemented one of two options when realizing the rise: early alignment of the L elbow, or late alignment of the L elbow. In cases of early alignment, the elbow most often fell on and occasionally even before the syllable boundary. The alignment of the late H* was relatively invariable, almost always occurring near the syllable boundary (mean latency from the right edge of the phrase was 1.2ms for all speakers, with σ =.53), with the rare but notable exception of contours like the one in Figure 3.8 that begin their descent to the following L before reaching the right edge of the phrase. To visualize the distribution of the L2 alignment values, a kernel estimation was used to calculate the probability distribution function for the alignment of the late L, shown in Figure 3.9 on the following page. In addition to the bimodality, the distribution highlights the fact that the speakers tended to produce more early- than late-aligned L targets, on the whole, which may be

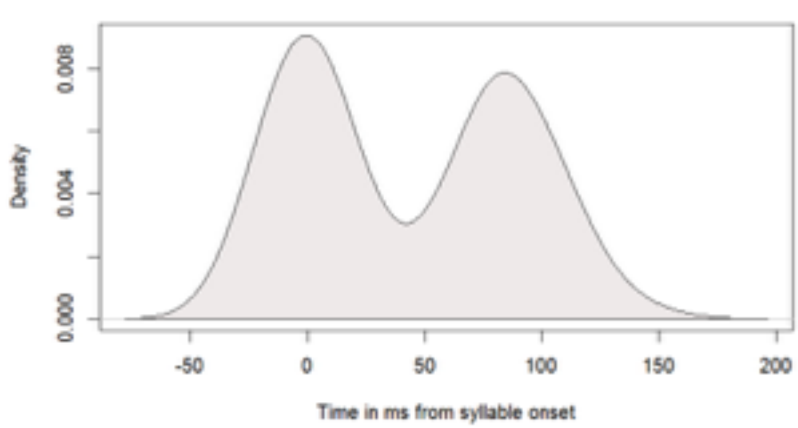


Figure 3.9. *Probability density function for the alignment of the late L elbow in L2 speech*.

attributed to their lack of confidence in completing the production tasks in French—self-doubt triggers high level and more rising intonation than confidence—or perhaps simply chance, given the relatively small sample size of speakers. These data clearly show that the learners struggled to produce correct alignment patterns in their L2, specifically with the late-L elbow, which was often delayed. Before discussing the implications of these results, it is worth remembering that French L1 speakers really only showed one alignment pattern for the LH* rising pitch accent, with the rise beginning almost immediately after the syllable onset in a vast majority of cases. Therefore, any statistically significant bimodality in the alignment patterns for the L2 speakers can be considered a realization of non-native phonetics.

The vast majority of the rises comprised by the probability distribution were either AP-final or ip-final continuation rises, the latter essentially being a pragmatically stronger implementation of the former. The most likely reason for the even bimodal distribution is that of the two possible large rises in French, i.e. continuation rises and question rises, these are the two less marked in the sense that they are tied directly to the default /LHLH*/ tonal pattern. Polar or yes/no questions, however, are syntactically marked, and they tend to carry more pragmatic weight, such that they are realized with larger and longer pitch movements than continuations, a difference that, as mentioned in Chapter 2, appears to be categorical in L1 French (DiCristo 1979), although there is a fairly large lack of perceptual research in recent decades that could either support or falsify such a distinction. An issue worth investigating, then, is whether the questions were implemented with more of a delay than the continuations, and if so, how much of an effect their collective delay had on the overall distribution of alignment values. As a first step, we can compare two continuation and question contours produced by the same speaker, which

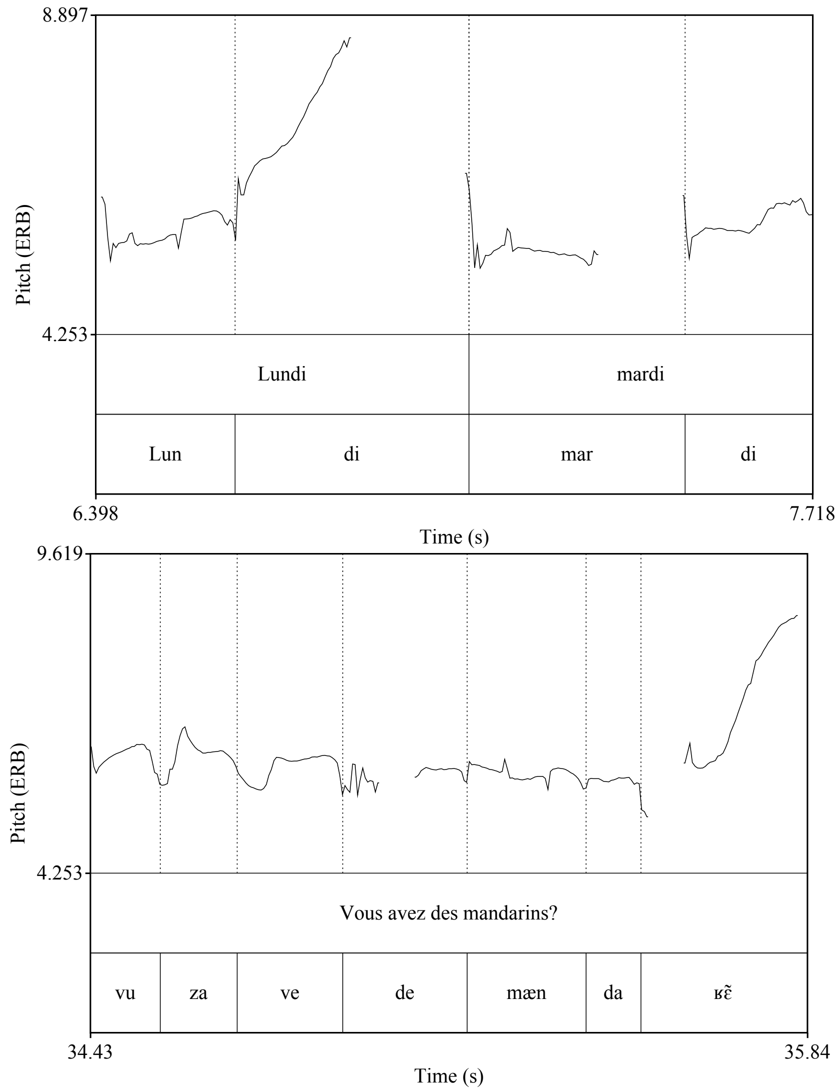


Figure 3.10. *Tonal alignment in L2 continuation and question rises*.

are presented in Figure 3.10 above. The first contour is taken from a sentence where the learner recited the days of the week and contains two continuations, one on *Lundi* ‘monday’, and one on *mardi* ‘Tuesday’. There is a visible difference in size between the two pitch excursions, with the first one being much larger, and in terms of alignment, the late L-elbow appears much sooner after the vowel onset in the first than in the second, where it is delayed until nearly halfway through the syllable. By comparison, the question rise in the second pitch contour shows the same large excursion size as the initial rise in the first contour, but it also shows delayed alignment, although the elbow is not positioned quite as far into the syllable as in the first example.

The first question these data raise is whether the rises in question share the same phonological structure in L1 French and the L2 variety exemplified here. In simple terms, the answer to this question is negative, since the native speakers in the corpus consistently produced early rises (LH*) under the same circumstances. In line with this fact, the second question the data raise is about the source of the phonological deviation in the learner's rises: is it the result of her tapping into the reduced tonal inventory assumed by the LVH, or is it simply a case of tonal transfer from L1 English? Two pieces of information are critical in answering this question, the first being the pragmatic content of each phrase, and the second being its underlying tonal structure. In both English and French, the three contexts (initial focus, continuation, and interrogation) are distinct, and they are all signaled by some kind of rising pitch movement, typically and H* in the former and an LH* in the latter. In this example, however, the phonological and phonetic content of the contours is more or less identical to their typical content in L1 English, especially considering the late alignment in the second and third contours, which would likely be considered instances of a L*+H rising bitonal accent. There is, in effect, no trace of the early-timed French rises in any of the contours other than the first, and even there, we could simply posit an H* on the initial syllable that triggers a rising movement toward the ip-final H%, as it begins too early to be considered a phrase-final LH*. There is no justification, therefore, for considering the rises as anything but examples of L1 transfer, which we can straightforwardly account for using the basic principles of the SLM by proposing that the learner did not (and probably could not) perceive the subtle difference in timing between the English and French rises, and thus substituted the one for the other in all three contexts. Clearly, this analysis, if true, would pose a serious problem for the LVH, which does not assume transfer to be the primary source of structure in L2 intonation. This issue (among others) is discussed more fully in Chapter 5.

3.4.3 Post-Program Data and Improvements in Alignment

Despite the fact that most of the learners in Groups B and C did not produce properly-aligned phrase-final rises in their pre-program interviews, some of them managed to produce them in their post-program interviews gathered after they returned from France. Before considering examples of this improvement, it is important to remember that the learners still struggled with not only phrasing but also the phonological structure of most of the unmarked tone sequences in French, both pre- and post-program. Their phonetic realization of certain contours improved, but they were restricted to very specific syntactic and pragmatic contexts, e.g. ip-final continuations and IP-final questions. That being said, some of the learners did make very significant gains both impressionistically and statistically in the nativeness of their alignment, going somewhat against the grain of previous research suggesting that learners were unlikely to acquire L2 intonational phonetics. Figure 3.11 shows a post-program pitch contour for an ip-final rise taken from the sentence *Vous voulez de la glace à la vanille ou à la noisette?* 'I've got some ice-cream. Do you want vanilla, or hazelnut?' (The English translation was rendered as two sentences to reflect the corresponding tonal structure in French). In terms of both alignment and pitch scaling, which is something examined more closely in Section 3.5, the final rise in this example falls within acceptable ranges for native French. Although the rise increases in slope rather drastically after

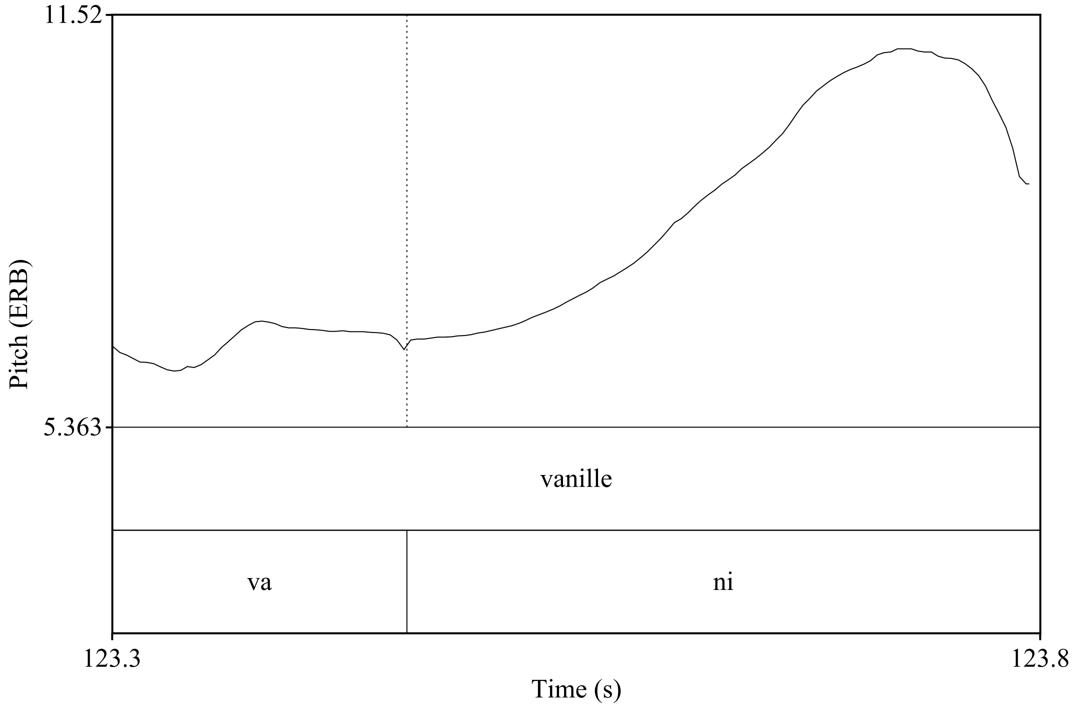


Figure 3.11. *L2 continuation rise with native-like alignment of the late-L elbow.*

the first few milliseconds of the vowel, it does not show the pronounced delay characteristic of L1 English and the pre-program L2 French data. Rather, the rise begins immediately after the vowel onset, and it climbs consistently until reaching the final H* near the right edge of the syllable. The final H* is still prolonged, and perhaps more than we would expect for a native speaker, but the contour on the whole is much closer to the L1 standard than before. This particular speaker produced this kind of alignment on about 75% of his ip-final continuation contours, which was a marked improvement over his original proportion of 32%. One other speaker from Group C (40% to 65%) and two speakers from Group B (51% to 73% and 44% to 69%) made similar gains. No speaker regressed significantly during the program, although some did record slightly lower percentages at the end than at the beginning, a difference likely due to non-linguistic factors influencing their performance on the production tasks, like nervousness, the expectation of improvement, and fatigue. These issues are briefly discussed in Chapter 5, along with more general methodological issues in conducting pitch production research.

3.5 Pitch Range and Tonal Scaling

Overall pitch range, measured as 80% quantal range in Equivalent Rectangular Bandwidths (ERB) was lower for the learners (μ=3.1ERB) than for the native speakers (μ=4.3ERB); the difference was statistically significant ($p<0.002$). This result echoes the findings in the literature, which point to learners having a reduced pitch range in the L2 relative to either of the L1s (Backman 1977, 1979; Mennen 1998, 2007; Mennen, Chen, and Karlsson 2010). Jun and Oh (2000) suggest reduced range may be the result of general processing constraints that limit the amount of information learners can prosodically code, and the suggestion seems to hold here, with the learners showing the most substantial reductions in range as compared to their L1 English also producing the highest number of phrase breaks during the DCT. A point discussed more thoroughly below, however, is that for a handful of the learners, range was more or less identical in the L1 and the L2, with level being the only feature that significantly changed. After z-score normalization, average values for the scaling of the late rises were shown to be significantly less variable for the learners as well, perhaps indicating a restriction in expressivity caused by a general processing constraint limiting the complexity of syntactic and pragmatic information that they can prosodically code. Reporting aggregate data can be misleading, however, especially for parameters as variable as overall pitch range. For some learners, the difference between their range and the native speakers' was statistically significant, but acoustically very small, sometimes reaching only .5 ERBs. To clarify these results, the data are presented in two subsections below, with each subsection corresponding to the relationship between L2 French and one of the L1s.

3.5.2 Differences in Scaling between L1 English and L2 French

Another useful perspective on pitch scaling in the L2 can be had by comparing it to scaling in the learners' L1. Mennen (2004) examined this relationship closely for speakers of L1 Dutch and L2 Greek, finding that phonetic transfer could go both ways (i.e. from the L1 to L2 and from the L2 to the L1), supporting Flege's SLM—speakers can always acquire new categories, so long as they are perceptually distinguishable—and giving further credence to the LVH. Regarding the learners in this study, two trends emerged, the first being that pitch range was either narrower or the same in L2 French as in L1 English, but never higher. Looking at the aggregate data for all of the speakers can be a little tricky, since it conflates the issues of inter- and intra-speaker variation in range and level, but looking at data from a single speaker illustrates the points nicely. Figure 3.12 shows the probability density functions for pitch in ERBs of one learner's L1 English (red) and L2 French (blue).

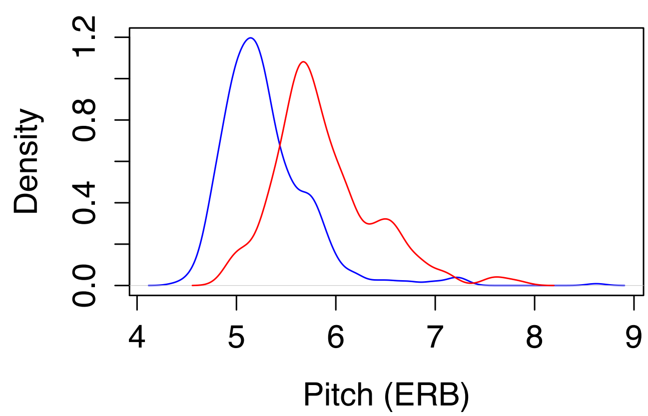


Figure 3.12. *Pitch range in L1 English (blue) and L2 French (red).*

In this case, pitch level is higher in the L2 (μ=5.9 ERBs) than in the L1 (μ=5.3 ERBs), a difference that was statistically significant (p<.0002), but range, as measured by standard deviation, is about the same (σ=0.45 vs. σ=0.5), and an F-test showed they are not significantly different (p=.153). The distributions are also close in terms of interquartile (IQR) range, with the first spanning 0.62 ERBs between the 20th and 80th percentiles, and the second 0.78. The probability distributions have very similar shapes, although the French curve does have a slightly larger hump to the right of the mean. These data contradict those in the literature that suggesting a narrower range in the L2, but they have a potentially simple explanation in that most aspects of this speaker's intonation in French closely followed her intonation in English—she produced relatively few AP-final phrase breaks, produced many English-style pitch accents, and consistently used the late rise instead of the early rise in ip- and IP-final continuations and questions. In other words, she transferred intonational phonology and phonetics from her L1 to the L2, which is reflected in the close similarity between the the two distributions of pitch. The similarity is perhaps more marked when the two distributions are normalized and plotted on the same scale, as shown in Figure 3.13 (following).

Between reporting the pitch data in ERBs and transforming them into z-scores, nearly all of the variation in range is accounted for, despite the two distributions having different actual means (as mentioned above, the speaker's L2 level was higher than her L1). What makes the similarity between these plots even more striking from a statistical point of view is that they were produced from two one-minute samples of speech, which are relatively short. Given more data, the curves would likely converge even more toward a single distribution, making the case even stronger for this particular speaker's L1 transfer. From a phonetic standpoint, it is also important

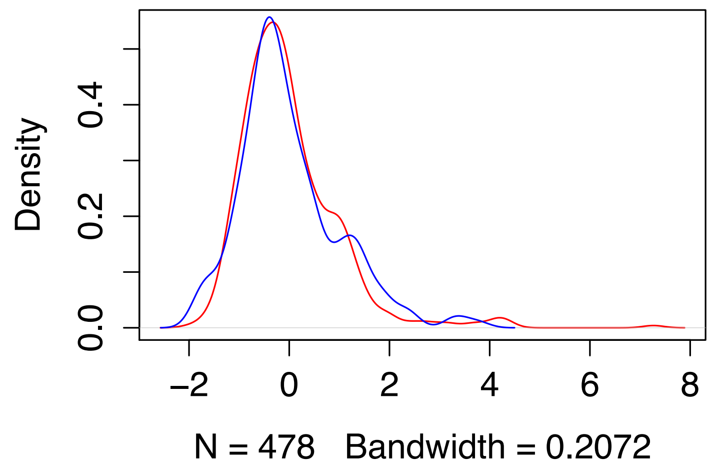


Figure 3.13. *Normalized pitch range for one speaker's L1 English (red) and L2 French (blue)*.

to note that these data provide an interesting look at the LVH, as there is no clear difference on average between the native and the learner variety of this particular speaker. If we were to extrapolate the distributions above to all speakers in the study, then if L2 French indeed had a narrower range than L1 French, it would simply be because L1 English had narrower range than L1 French and not because the learners were somehow limited in speaking the L2. This possibility is discussed more thoroughly in Chapter 5, but it is interesting to mention here, especially because Mennen, Chen, and Karlsson (2010) do not provide L1 pitch data for the learners in their study (in fact, it is not clear from the description of their methods whether they were collected at all).

Two of the learners also had *wider* ranges in L2 French than they did in L1 English, despite them both having relatively low scores for phrasing and alignment. Figure 3.14 shows the pitch distributions for one of them (English is in blue, and French is in red), and it illustrates a core issue with use overall range as a metric for learner progress, which is that there is no way to

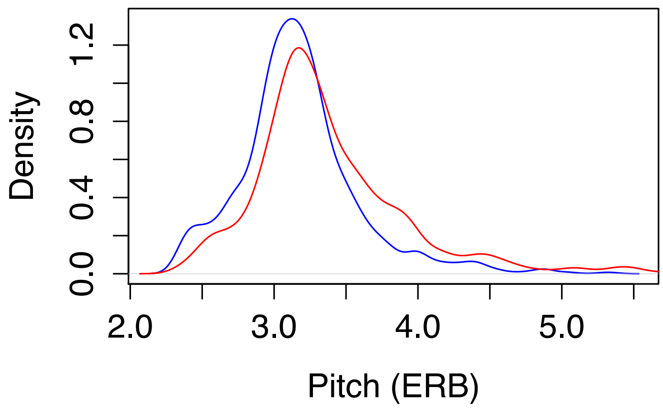


Figure 3.14. *Extended pitch range in L2 French (red) as compared with L1 English (blue).*

know what the source of variation in the pitch data is, or why they are being pulled in a particular direction relative to the two L1s. In this learner's case, his 80% IQR was .57 ERBs in English and .70 ERBs in French, both of which seem small when compared with the L1 French (mean IQR = 1.73 ERBs), but which are within the average ranges for the L1 English data. If we based our judgments about his level of acquisition on range alone, we would end up deciding that he was better at speaking French than English, which of course is not true. More likely is that he uses a relatively small range when compared with the other English speakers, especially when compared to the female speakers, and that he was actually making French-like adjustments to the scaling of particular tonal targets in his L2, widening the range and lowering the amount of time he spent speaking at or around his mean pitch (represented in the distribution plot by the main peak). Whatever the source of the discrepancy, though, the data still contradict the notion that L2 intonation is on the whole narrower in range than L1 intonation—this is the strongest version of the claim made by the LVH—which means we should probably question its place as a metric in the theory of intonational acquisition.

This is also a good spot to mention that while z-score normalization is helpful for comparing the basic shape of pitch distributions, it should probably not be used to compare pitch ranges from two separate speakers, since it transforms most of the variation out of the data, producing distributions that are statistically homogenous (this is, after all, the point of normalization). Because the F-test (i.e. one-sided ANOVA) strongly assumes the data are normally distributed, it is probably not the best choice either, so the IQR method proposed by Mennen is settled on here as the metric for comparing ranges. However, for comparing the scaling of individual tonal categories or pitch contour in specific pragmatic contexts, it is quite good, since we can accurately determine how big the pitch movement is in relation to the speaker's overall pitch space. Of particular interest is whether the learners scaled the two target contours for this study—continuations and questions—the same in French as they did in English, which has implications not only for the possibility of L1 intonational transfer, but also for our goal of evaluating the learners' progress in acquiring French.

On the whole, the learners tended to scale continuation higher in French than in English, but only at strong syntactic boundaries, i.e. ip-final breaks; elsewhere, they scaled the contours much smaller, which was likely a contributing factor to their overall range being narrower than that of the native speakers. Figure 3.15 shows an example of the latter, where the speaker produced a larger rise on the French *Lundi* 'Monday' than on English *Monday* when enumerating the days of the week. The rises differ significantly not only in terms of raw height (8.5 ERBs vs. 6.7 ERBs), but also in terms of z-scores (3.1σ vs. 5.1σ), or standard deviations from the mean, despite their underlying phonological similarity (the syllable break is not clear in the figures, but it occurs halfway through each contour). Across learners, this difference was relatively

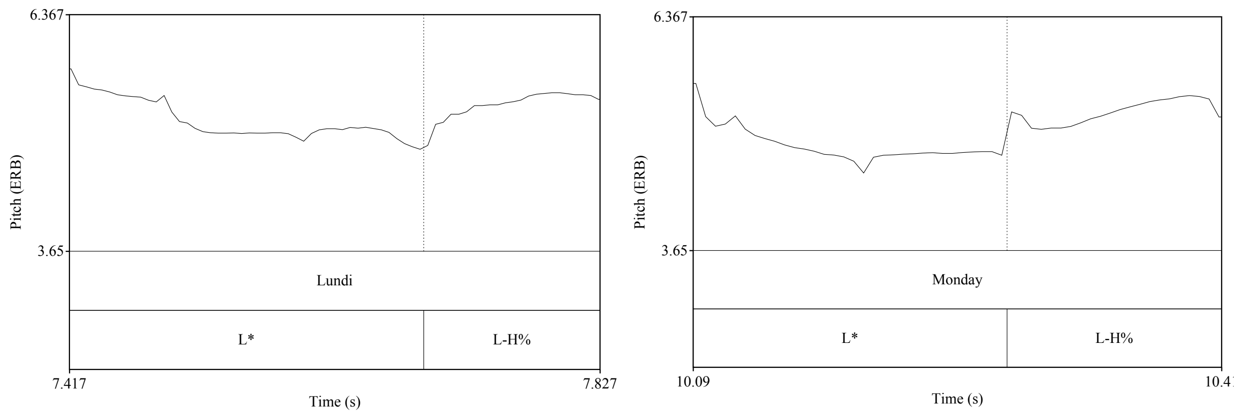


Figure 3.15. *Scaling of ip-final rises in L2 French (left) and L1 English (right)*

consistent, with the L2 continuations being scaled on average 1.3σ higher than those in English, relative to each speaker's individual range (i.e. the results were pooled after and not before . calculating the difference). The main reason for this is likely a small difference in tonal specification between the two contours, since the French rise should technically be a LH*H%, while the English rise is an L*L-H%, with the low phrase accent shortening the height of the final movement. Phrase accents are not usually posited for French (Jun and Fougeron 2000; Post 2002; Delais-Roussarie et al. 2014)

The scaling of polar questions was much more similar in the learner's two varieties than that of continuations. Figure 3.16 shows pitch contours produced by a speaker in Group B for the final rise in the question *Vous-avez des mandarines?* 'Do you sell mandarins?' in French (a) and in English (B). The tonal structure of the utterances is different, but the scaling of the final rises is very similar (7.5 ERBs vs. 8 ERBs), and z-score normalization also shows they are almost identical in terms of the proportion of the speaker's range they comprise (z=4.12 vs. z=4.72). Although the English rise is a bit larger in terms of its z-score, both are very large relative to the other rises in the data, as values four standard deviations away from the mean are rarer than

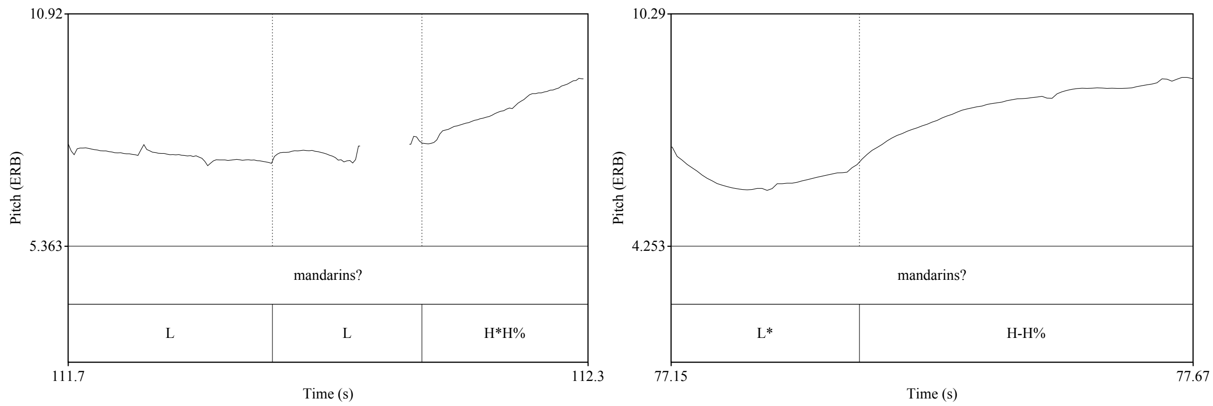


Figure 3.16. *Scaling in L2 French and L1 English*

99.999% of the data. It is important to remember that although there are clear phonological dissimilarities between the two phrases, with the first rise showing near-native alignment of the L-elbow but the second being aligned much earlier, they begin and end at approximately the same height in the speaker's range, which is of primary interest for understanding the relationship between L1 and L2 French.

3.5.1 Differences in Scaling between L2 French and L1 French

Much of the data supported the idea that learner varieties of intonation are phonetically distinct from native varieties. Figure 3.17 shows the aggregate pitch distribution for the female learners

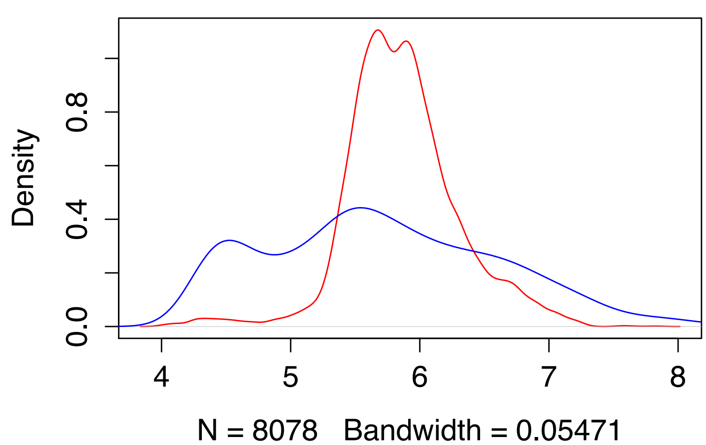


Figure 3.17. *Mean pitch range in L1 French (blue) and L2 French (red).*

plotted alongside the aggregate distribution for the female speakers of L1 French. In general, these results do support the claim that learner intonation is narrower in range than native intonation, as the mean pitch peak is much higher and narrower for the L2 distribution than for the L1. As mentioned above, there is a good chance this difference stems at least in part from the learners' narrower range in their own L1 English, which in most cases followed the same basic shape as the aggregate distribution above, showing a strong central peak instead of the soft hump in the French distribution. Nonetheless, the two distributions clearly have different shapes, and both the IQR comparison (Δ=1.11 ERBs) and F-test ($F=0.189$, $p<2.2e-16$) support the conclusion that the differences in spread are statistically significant. Again, the F-test is probably not to be trusted in this particular case, since the L1 distribution especially does not appear to be normal, and Shapiro-Wilk ($p<0.0002$) tests for both distributions rejected the null hypothesis that the data are normally distributed. Other tests could be used for evaluating the difference in variation, but eyeballing the distributions for these data is probably enough, since they clearly have different fundamental shapes.

3.5.3 Pitch scaling in post-program L2 French

Along with improved phrasing and alignment, several of the learners also produced improved pitch scaling in the post-program L2 French. In general, these results went hand in hand, and only one speaker improved the former two and not the latter. In his case, alignment and phrasing became more native-like, but scaling did not improve, perhaps because the first two can be adjusted phonologically (alignment by using fewer L*+H English pitch accents in phrase-final rises), while scaling cannot. When speakers did make progress in scaling, it surfaced in two main ways: expansion of their overall speaking range, and adjustment of the height of phrase-final rises to match the height of those in French. There is clearly a relationship between these two parameters, since the expansion of range would also require the expansion of the height in at least some of the individual pitch accents or contours, but it is beyond the statistical scope of this dissertation to determine its specific quantitative nature. This limitation notwithstanding, these two features provide the best evidence of intonational acquisition in this data set, so they are worth examining more closely.

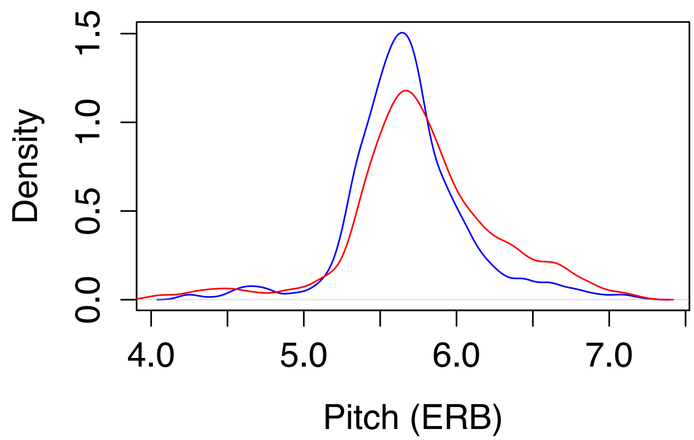


Figure 3.18. *Pre- and post-program pitch range for L2 French.*

Figure 3.18 above shows the pre- and post-program pitch distributions for one of the speakers in Group B. The basic shape of the distributions is the same, with one strong peak and several smaller humps in the lower frequencies, but the pre-program distribution (blue) is significantly narrower than the post-program distribution (red) ($F=0.6376$, $p<0.0001$), which slopes softly to the right after the main peak instead of dropping off sharply; the difference in IQRs (.49 vs. .60 ERBs) reflects the trend as well. On average, the speakers in Group C were more likely to show an expanded range in the post-program French, but both groups made statistically significant, although perhaps not linguistically meaningful, progress. Table 3.3(following) shows the pre- and post-program IQRs for the speakers in Group C who participated in both interview sessions, and the aggregate means for each gender and group are also included.

Table 3.3. *Pre- and post-program pitch ranges for female and male learners in Group C.*

Speaker	Pre IQR	Post IQR	Difference (Δ)
1f	0.49	0.68	0.19
2f	0.89	1.26	0.37
3f	0.75	0.9	0.15
Mean (μ)	0.71	0.95	0.24
Speaker	**Pre IQR**	**Post IQR**	**Difference (Δ)**
1m	0.53	0.67	0.14
2m	0.38	0.73	0.31
3m	0.62	0.83	0.21
Mean (μ)	0.51	0.74	0.22

Although the goals of this study were not primarily sociolinguistic, it should be also noted that the male learners tended to have a much narrower range (even after z-score normalization) than the female learners. Each of these speakers made gains in range, but even their post-program ranges fell well short of the aggregate IQRs for the L1 French speakers, which was 1.73 ERBs for females and 1.2 ERBs for males. Based on this information alone, we can say that the study abroad program did have a positive impact on the speakers' L2 intonation, but we cannot say whether the relative narrowness of their range is due to factors influencing their L2 proficiency or simply to L1 transfer. As noted above, the aggregate IQRs for L1 English were 0.85 ERBs and 0.58 for the females and males respectively, which suggests that the expansion of range during the program is more likely due to the speakers moving closer to their individual L1 ranges (e.g. as they become more comfortable with French) than to the L2 range, although the only way of

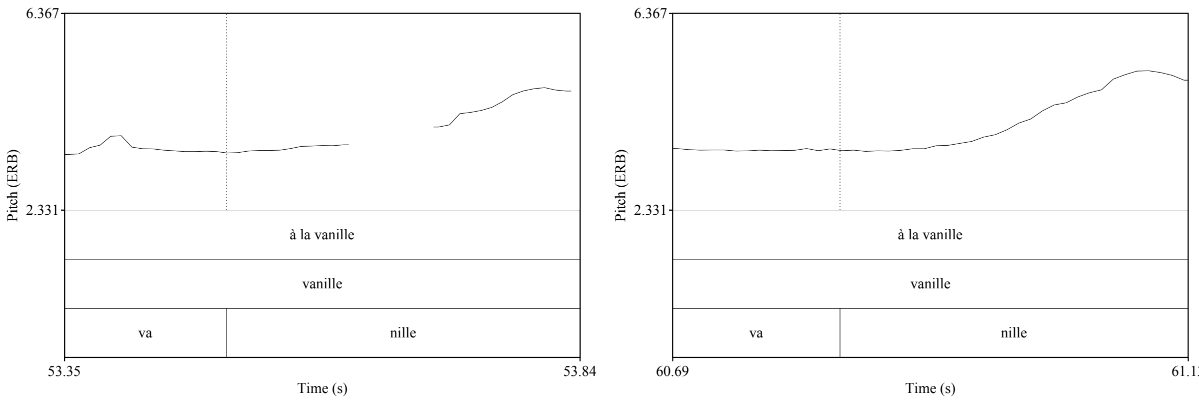


Figure 3.19. *Pre and post-program scaling of ip-final continuation*.

testing this hypothesis would be to extend the length of the program and see whether the ranges level off near the L1 limit or continue expanding. Given that the learners are completely proficient and thus fully expressive in English, the former seems the more likely scenario, but both would certainly be possible.

The other question we should answer is whether the learners improved in the scaling of particular pitch contours during the program, or if the values stayed closer to those in their L1. Overall, the answer to this question appears to be no—for most of the speakers, the height of specific pitch movements remained constant with respect to range in the post-program data, implying that improvements in the latter are due primarily to them expanding all of the pitch movements (perhaps by becoming more comfortable with the L2). This finding is in line with Mennen, Chen, and Karlsson (2010), but it has a few caveats, the first being that as with range, a few of the students managed to make significant improvements. We can see an example of this in Figure 3.19 above, which shows pre- and post-program pitch contours for an ip-final continuation in the sentence *Vous voulez la glace à la vanille ou à la noisette?* 'I've got some ice cream—do you want chocolate, or vanilla?'. The rise on the left is scaled to about 4.5 ERBs

(1.1σ), while the one on the right reaches about 5.2 ERBs (1.9σ), showing a significant (although perhaps not substantial) movement toward what would be the average height of the rise in L1 French. On average, this speaker's AP-final rises went from 3.8 (0.3σ) to 3.6 ERBs (0.03σ) in height, and his ip-final rises from 3.9 (0.83σ) to 4.5 ERBs (2.4σ), showing progress toward but not attainment of the French norms, which were about 1.0σ and 2.0σ respectively. The second rise is also a good example of the over-heightened rises that some of the speakers produced in their post-program data, where ip-final LH* was scaled to the same height (or higher) as the IP-final LH*H% used in questions. There is too little research on L2 intonation to say whether this is a case of phonetic hypercorrection, but on the surface, it seems to be a reasonable explanation of the contours' unexpected height.

3.6 Summary of Results

The main results of this chapter can be summarized as follows. First, although pitch level remained the same in the speakers' L1 and L2, pitch range was significantly (though again not substantially) reduced, indicating a reduction in expressivity in line with the results of previous studies. The scaling of specific contours, however, was much more variable, with some being scaled differently but others the same in the two varieties. Second, the learners produced phrase boundaries after IPs and ips more often than they did after APs, and even post-program production data show they struggled to move from the pitch-accent-oriented English system to the phrase-oriented French. Finally, alignment values for continuation and question rises suggest that the learners transferred tonal categories from the L1 to the L2, supporting perceptual models of L2 acquisition like the SLM, but bringing the learner variety hypothesis (LVH) into question.

With respect to the existing literature on intonational acquisition, the most interesting result of these production experiments is that learners tended to scale phonologically-identical rises the same in both their L1 and their L2, so long as the contours appear in similar pragmatic contexts (e.g. questions). This tendency makes a strong case for L1 transfer being the dominant force in shaping the learner variety, and it falls in line very nicely with the SLM, which broadly predicts that speakers will only acquire non-native phonetics when they are associated with perceptually distinct phonological units. In terms of the SLM, the issue with the phrase-final rise for the learners is that their two English categories (L*+H and L+H*) converge on only one French category (LH*), and that they seem to perceive the former as acceptable instances of the latter. The consistency with which the L-elbow is aligned late in the L2 French contours further suggests that the learners responded to this perceptual equivalence by using the rises in specific semantic and pragmatic contexts exactly as they would use them in English, further supporting the idea that L1 transfer and not a universal learner variety provides the structural backbone for their intonation in the L2. This hypothesis should be easy enough to test, since tonal alignment is easy to manipulate in perception experiments, unlike pitch scaling, which is more difficult to quantify. Suggestions regarding how these results could be implemented with perceptual models of SLA are made in Chapter 5.

The final topic worth mentioning in more detail is the apparent cross-linguistic equivalence of pitch height in certain pragmatic contexts. After converting each speaker's pitch data to z-scores and then aggregating them by language variety, it appears that excursions of certain sizes, especially the 1σ and 2σ levels, play important roles in both L1 and L2 intonational phonetics. Although much more research is needed to flesh out the idea, it is possible, and perhaps even likely, that by parameterizing pitch movements along the axes of height, direction, and alignment, we could develop a grammar of tunes that could be used to make connections between the phonetic, and thus the underlying phonological, intonational components of different languages. Chapter 5 addresses this possibility in slightly more detail, and it presents additional evidence from this chapter's production experiments to give support to the suggestion.

CHAPTER 4

TILT ANALYSIS OF LEARNER INTONATION

This chapter presents an automatic intonational variety classifier based on the Tilt model, originally created by Taylor (1998; 2000) to be used for intonational synthesis in text to speech (TTS) applications. The model is flexible, robust, and trainable on acoustic data, and it increases the linguistic relevance of Taylor's original algorithm. Before describing the model's structure in detail, Section 4.1 provides theoretical and methodological justification for pursuing a low-level or purely phonetic approach to studying L2 intonation, and it provides some background information on the original uses of the Tilt model and its more generic predecessors used in speech processing. The section also describes clustering, an alternative method for processing large amounts of acoustic data, and makes a brief argument in favor of using sequential algorithms instead. Section 4.2 begins the presentation of the Tilt analysis itself by describing how the audio corpora were processed and labeled for phrase boundaries and intonational event types. Section 4.3 continues the presentation by introducing the statistical basis for the Tilt Model, including its methods for parameter estimation and hidden Markov model development, and then explaining the architecture of the models developed for the L1 and L2 varieties, providing a detailed look at their shared characteristics and dissimilarities. Importantly, it also describes how the HMM used for this research differs structurally from those used by Taylor (1999) and others researchers for more general tasks in speech processing. Section 4.4 presents the results of the Tilt analysis, with a special focus on the likelihood estimates produced by the classification task. Finally, Section 4.5 concludes by showing how the model can be used to evaluate the quantitative dissimilarity between two varieties, by summarizing findings from the previous sections, and by proposing directions for future research into L2 acquisition using similar methods.

4.1 Introduction to Global Models

4.1.1 Justifying the Global Approach, or Why not Use ToBI?

Although neither clustering algorithms nor the Tilt model have been applied to L2 speech, both have the potential to provide insight into how learner varieties of intonation are shaped, not only phonologically, but also (and perhaps most especially) phonetically. The reason they are classified here as global rather than local models of intonation is that they are designed to give us insight into how the intonational system works as a whole, rather than as a collection of individual tonal categories. Theories of intonation within the AM framework are intended to be global in this way as well, but because they define the system's components from the bottom up rather than from the top down, they are rigidly language-specific and difficult to adapt for work on L2 intonation. The difference between global and local models might seem largely theoretical and indeed superficial—neither is perfect, and both are helpful—but it is worth suggesting for a number of reasons. To begin with, we know learner varieties of language are distinct from L1 varieties (Klein and Perdue 1997), and that defining them in terms of L1 varieties is theoretically unsatisfying and practically troublesome. Linguists have tried to avoid this problem by simply defining the L2 variety on its own as if it were structurally independent and self-contained—indeed, this is the approach taken by most contemporary studies (Mennen, Chen, and Karlsson 2010)—but we encounter it again when trying to determine how it relates to other L1s, especially phonetically. This limitation stems from the basic fact that the AM framework was not designed for cross-linguistic research, although it has clearly been used for it with some success, and as a result, it is poorly equipped to compare languages in a way that is objective and robust (see Ladd 2008 for a general discussion of this point). Thus, although it can show us that two languages are structurally related, an AM-centered approach cannot tell us to what extent, at least in any meaningful way. Global models do not suffer this problem, which in many ways is their main strength.

Another advantage of using low-level phonetic models to study L2 intonation is that they provide an objective way to test whether learner varieties really are distinct from their related L1s in terms of prosody. Following the boom in L2 acquisition research in the 1980s and 1990s, Klein and Perdue (1997) argued (convincingly) that learner varieties are morphologically, semantically, and phonologically distinct from native varieties, an idea that Mennen and her colleagues (e.g. with Chen and Karlsson 2010) have recently extended to their research on intonation. Her findings suggest that while learner varieties do have some phonetic characteristics in common, they are defined mostly by their structural simplicity, having in most cases a reduced inventory of tonal categories. Although there is nothing wrong with the methods she used to reach this conclusion, the basic fact that they are based on the AM framework is problematic for two reasons: there is no ToBI standard for transcribing learner varieties of intonation, so pitch contours must be labeled subjectively and on an individual basis; and phonetic transfer can go from the L1 to the L2 and vice versa (Mennen 2004), so we are likely to overlook surface-level features that the learner varieties may have in common with their native varieties. By using purely phonetic models, we can test the learner variety hypothesis without labeling and without referring to intonational categories, allowing us to make a purely quantitative judgement of whether L2s are phonetically different from L1s and, if so, to what degree.

Another major strength of using a low-level model to compare varieties of intonation is that we can avoid having to deal directly with issues like the pragmatic and paralinguistic uses of pitch. The first reason for this is purely formal: because the models do not assume that F0 is influenced by underlying tones, they do not need to account for how the tones may be influenced by variables like speaker mood or belief state. If they are designed well and trained on high-quality data, the models can inherently handle this kind of variation, which, as we will see in Section 4.3 below, becomes a statistical problem along the same lines as determining the most likely sequence of pitch accents or the average height of boundary tones. The other main strength of using low-level models is that being purely phonetic, they can process large amounts of speech, meaning their statistical power can also be very high. Intonation models relying on phonological transcriptions, like ToBI and the IPO model, require a great deal of time and effort to process large audio corpora, so they do not usually include the same number of data points per speaker as models constructed for studying segmental phenomena at the same (i.e. the variety- and language-wide) level. Many of the studies using these models only report data from 10 or so speakers (Mennen, Chen, and Karlsson 2010), a number that tends to decrease when the focus is switched to intonational acquisition, which increases not only the difficulty of finding speakers (reasonably proficient L2 speakers are clearly harder to come by than proficient L1 speakers), but also the difficulty of transcribing their intonational data objectively and rigorously. Unfortunately, for the former reason, this study does not exploit the low level model's potential to process large corpora, but the benefit exists, at least in theory, making it a very attractive alternative to the phonological approach.

4.1.2 Parameter-Based Clustering

Unlike finite-state transition networks, which model some number of states and the transitions between them and are the formal basis for most models of intonation, including Pierrehumbert's AM treatment of English and the Tilt model outlined below, clustering algorithms do not represent linguistic data as a series of events; instead, they show how the average values for some group of related events are distributed in acoustic space. Because clustering is not a specific algorithm per se but rather a collection of related statistical techniques, it can be adapted to fit the formal assumptions underlying the linguistic phenomenon being studied. Classic examples of how clustering has been used in phonetics come from sociolinguistic studies of vowel space. Klein, Plomp, and Pols (1970) tested a clustering-based approach for identifying 12 Dutch vowels, finding that maximum likelihood areas based on the logarithm of the first two formants (F1 and F2) correctly classified new vowels with about 70% accuracy. The clustering algorithm they used was itself somewhat rudimentary, being based on what they already knew about the vowel space (i.e. clusters they expected to find) rather than statistical inference. Similar procedures like multidimensional scaling (MDS) and principal component analysis (PCA) have also been used by phoneticians looking to visualize trends in acoustic data. Although clustering algorithms are powerful, they cannot be used to model time series or sequential data, which limits their applications in intonational phonetics (and, for that matter, phonology) for a number of reasons. First and foremost, every contemporary theory of intonation treats pitch contours as corresponding to underlying sequences of tones, which a clustering algorithm would not be able to model. Second, although pitch accents are often realized by changes to both F0 and intensity or loudness, not enough experimental research has been done to determine whether a relationship between the two parameters would be cross-linguistically consistent. Finally, despite the inherent interspeaker variability of pitch range and level, pitch data obtained from a single speaker tend to cluster very tightly around the mean, which would make it hard for the algorithm to detect meaningful clusters (in general, they work best when the data are clearly distributed in groups, like those in the vowel space). Because of these limitations, and because clustering has really not been used at all to study prosody, the analysis below will focus on sequential models of intonation.

4.1.3 The Tilt Model

According to Taylor (1998; 1999), traditional models of intonation like Pierrehumbert's AM model and Hirst's INTSINT model are most appropriately defined as classifiers because they assign linguistic events in the pitch curve identities based on a predetermined set of categories; in other words, given a pitch contour, they determine what sequence of tones is most likely to have produced its shape. The Tilt model, however, is more of an intonational event detector in that its only goals are to locate linguistically meaningful parts of the pitch contour and describe their shape using a basic set of acoustic parameters. Like Hirst's (1993) Momel and Fujisaki's (1995) model for English, the Tilt model makes no assumptions about phonological system underlying the sequence of events, nor does it attempt to determine whether the sequence violates some sort of rule structure within the language's prosodic system (e.g. it cannot be used to make grammaticality judgements). Along with its predecessor, the rise-fall-connection (RFC) model, the Tilt model was developed and has been used primarily for intonational synthesis (Taylor and

Black 1994; Dusterhoff and Black 1997; Dusterhoff, Black, and Taylor 1999), but its statistical underpinnings make it useful for analysis as well (Black 1997).

Whereas the basic unit for models in the AM framework is the phonological tone, the basic unit for the Tilt model is simply the intonational event. Unlike tones, which because they are associated with underlying segments must be contiguous, events are allowed to be separate, occurring between stretches of the pitch contour that are not linguistically relevant, avoiding the issue of over-inclusion posed by Momel and other models driven by curve fitting. Like models in the AM framework, however, the Tilt model assumes two primary classes of events: pitch accents, which speakers use to emphasize particular words or syllables, and boundary tones, which speakers use to encode phrase-level syntactic and pragmatic information like interrogativity and continuation. The events are then assigned values for amplitude, duration, and tilt, which are the model's three parameters. Amplitude and duration measure the same acoustic properties here as elsewhere, i.e. the magnitude of the event's change in pitch and its length in milliseconds, respectively. The tilt parameter, on the other hand, is a dimensionless number describing the overall shape of the event and is based on Taylor's (1993; 1994) earlier RFC model, which described the events using the duration and amplitudes of the rises and falls, as shown by Figure 4.1 (following) (Taylor 1995). The RFC model also takes into account the position of the transition between the rise and fall portions (i.e. the peak alignment), yielding a total of five parameters.

The RFC model's key innovation was treating pitch accents and boundary tones as having more than one part: whereas Hirst's (1993) and Fujisaki's (1992; 1993; 1995)

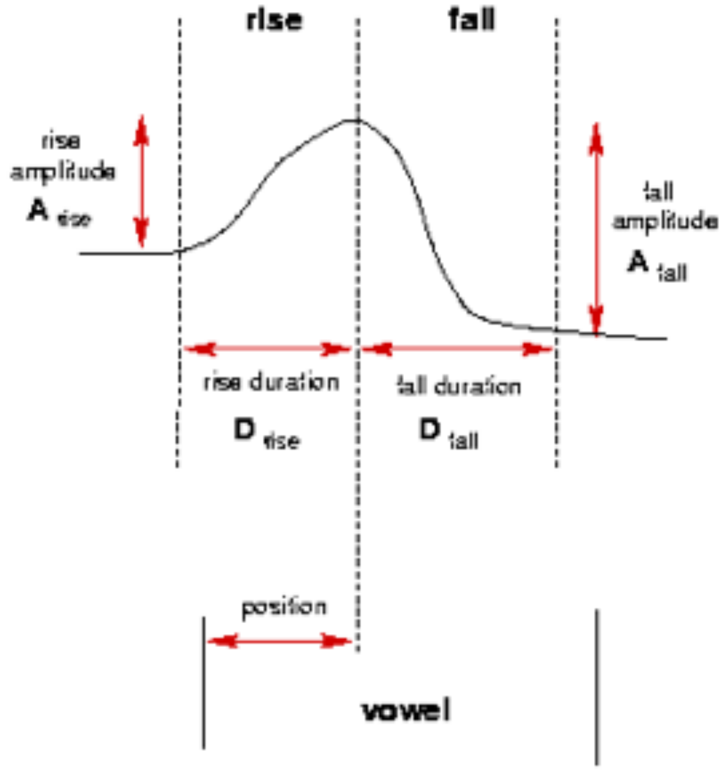


Figure 4.1. *The RFC model (Taylor 1995).*

algorithms seek to model the entire intonational event with a single equation, the RFC's algorithm uses different monomial equations to model its rising and falling portions separately, increasing the model's accuracy while also decreasing its computational complexity (there are twice as many equations as in the other models, but being monomial rather than polynomial, the equations are much simpler to compute). Linguistically, this treatment makes sense, too, because there is no clear evidence in the literature to suggest that speakers store pitch contours as whole curves; in fact, it is more likely that the mental representations are either level tones or tonal targets (Ladd 2008), and so treating them as single movements (e.g. rises or falls) seems not only more efficient, but also more direct. The RFC model's other advantage is that it does not count the spaces between intonational events as being linguistically relevant and thus necessary to model using curved functions, an assumption maintained by the Tilt model. Instead, the spaces, called *connections*, are treated mathematically as straight lines with variable duration and amplitude parameters, allowing them to angle upward or downward when linking events.

While its simplicity relative to the other low-level phonetic intonational models is clear, the RFC model still has room for improvement in that its treatment of single events like a pitch accent or boundary tone as compositional is neither linguistically justified nor maximally computationally efficient. High-level phonological models like Pierrehumbert's for English treat these events as unified categories, and so a better (or perhaps the best) compromise would be to combine the rise-fall measurements into a single parameter, thereby maintaining the formal assumptions of the RFC model but improving its performance. In fact, the Tilt model achieves this goal by reducing the total number of parameters to three: amplitude, duration, and tilt, with the tilt parameter itself being a combination of the ratios of the rise and fall amplitudes and durations. Although this sounds somewhat complicated, Figure 4.2 (Taylor 1999, p. 1704) shows that the calculation is relatively straightforward.

$$tilt = \frac{|A_{rise}| - |A_{fall}|}{2(|A_{rise}| + |A_{fall}|)} + \frac{D_{rise} - D_{fall}}{2(D_{rise} + D_{fall})}$$

Figure 4.2: *Equation for the calculation of tilt.*

This simplicity raises the question, however, of why amplitude and duration, which have traditionally been viewed as separate characteristics of F0, are combined into one parameter. Taylor (1998) explains this reduction by saying that because empirical evidence has shown the two values to be highly correlated, we can present them as a single value without suffering a significant loss of information in the model. It is also important to note that intonational events in the Tilt system are still assigned values for these parameters separately; only in calculating the tilt parameter itself are they combined, which gives the representation of the contour an overall shape in addition to its size in the frequency and time domains.

Figure 4.3 below (Taylor 1999, p. 1699) shows how a pitch contour would be divided into segments according to the Tilt model, with pitch accents being labeled with an 'a', boundary tones a 'b', and syllables an 's'. The diagram makes clear a number of important features of the model that reflect its theoretical assumptions and empirical power. Although the model does not assume an AM treatment of intonation in that syllables are not divided into segments and they are only loosely attached to the events, it does share the basic feature of having multiple levels of representation, with both models linking the surface F0 contour to the

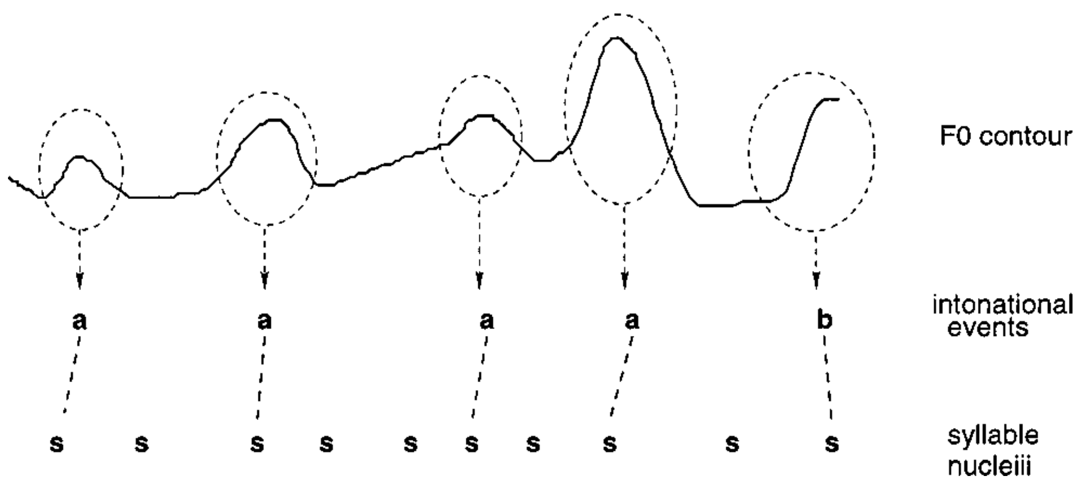


Figure 4.3: *Hypothetical pitch contour and intonational events*.

abstract linguistic representation (the syllables) by means of an intervening level (the events). The diagram also shows that the model leaves out large portions of the pitch contour when calculating its parameters, focusing instead only on the parts where there are noticeable, sometimes drastic, changes in F0. Taylor (2000) notes that this selectivity is indeed a strength of the model, as not all of the information contained by the raw contour is linguistically relevant. The changes in pitch between the events, for example, can be calculated largely by means of straight-line interpolation (Taylor 1998) as they were in the RFC model, so there is no reason to include them in a model intended to capture information that is determined solely by the speaker and thus unpredictable. These connections are distinct from pitch accents and boundary tones, and they may attach to more than one syllable, which makes them unique in the model. Finally, although the Tilt system divides events into two classes, pitch accents and boundary tones, it assumes they both have the same set of parameters, e.g. amplitude, duration, and tilt, in addition to height and alignment relative to the vowel onset. Formally, this division is satisfying in that it corresponds nicely to the tonal categories proposed in AM theories of intonation, and practically, it makes the model itself easy to implement, since the parameters can be calculated event-by-event without needing to switch between types of feature sets. As we will see in Section 4.3, by slightly modifying Taylor's (1999) original model, we can design an HMM-based classifier to reflect the phonological-phonetic dichotomy in the AM framework without losing its ability to robustly and accurately determine the probability of a series of pitch movements being generated by a particular language variety.

4.2 Corpus Preprocessing and Event Labeling

This and the following sections present an analysis of L2 French intonation using Taylor's (1998; 2000) Tilt intonational model. Like the production experiments in Chapter 3, the goal of the analysis is to test the hypothesis that learner varieties of intonation are phonetically distinct from the native varieties to which they are related, only using a primarily statistical rather than phonological approach. The analysis below incorporates temporal information into the model of

L2 intonation, treating the F0 contour as a sequence of intonational events rather than as a single phenomenon occurring within pitch space. Perhaps more importantly, it diverges from Taylor's original model in that its structure reflects the tonal inventory of the language being studied: traditional speech processing algorithms use HMMs to model a sequence of sounds within a single word, but here they are used to model the sequence of tonal events itself. The primary advantage to this approach is that it allows us to incorporate linguistic information into the acoustic model without needing to make (what are often subjective) judgments about tonal sequences underlying the pitch contour. The result is a powerful computational tool for comparing language varieties on the basis of intonation alone, which fills a large gap in the literature on intonational phonetics and, more generally, the quantitative evaluation of L2 acquisition.

The same audio corpus was used for this section as for the AM analysis in Chapter 3. Owing to the particular requirements of the software used to implement the Tilt analysis, however, several changes were made to the sound files before running the model. The Tilt model itself was implemented using the intonation package in the Edinburgh Speech Tools (EST) library, a family of programs designed for general speech processing and annotation. The package takes two files as its input: an F0 pitch file, and a text file with information about the intonational events, like start time and event type (a, b, or silence). After producing tilt parameters for each event in the file, the results can be exported to an output file for statistical analysis, which in this case was performed using an HMM package in R (Team 2014). The package contains programs for designing, training, and testing HMMs and has been used in a variety of fields, including bioinformatics, financial statistics, and speech processing. Its benefit here is that it can take the Tilt parameters produced by EST and build a general HMM that predicts their values, making it possible to build a classifier that automatically determines the probability that a new sequence of events could have been produced by the established model. In order to test the hypothesis that learner varieties are phonetically distinct from native varieties, two base HMMs were developed, one for L1 French, and one for L1 English. The models were trained using extemporaneous speech from the UGA, MR, and PFC corpora, and the parameters were estimated using the Baum-Welch algorithm. The recordings from each L2 speaker were then run through the classifier to produce likelihood estimates for their speech having come from either of the L1s; this was implemented using a forward algorithm. After obtaining likelihood estimates for the individual speakers, the results were pooled as a single group to determine on average whether the classifier could distinguish the L2 speech from the L1s, which was used as the primary metric for determining whether the learner varieties are distinct.

Before loading the files into R for statistical analysis, they were segmented into interpausal units (IPUs) in SPPAS and then labeled by hand for intonational events in Praat. Several of the PFC *enquêtes* were already labeled, in which case the word, syllable, and phone boundaries were simply checked for accuracy and adjusted when needed, which was rare. Event boundaries do not always correspond with syllable onsets and other phonological or syntactic boundaries, so there was no way to automatically label them, although it would have been possible to design a custom HMM for doing so in HTK (Young et al. 1996). The labeling procedure closely followed Taylor's (2000), with events classified as either pitch accents (a), boundary tones (b), or connections (c), and non-event portions of the F0 curve marked as silence. The procedure is exemplified in Figure 4.4 (following), which shows a labeled pitch curve for

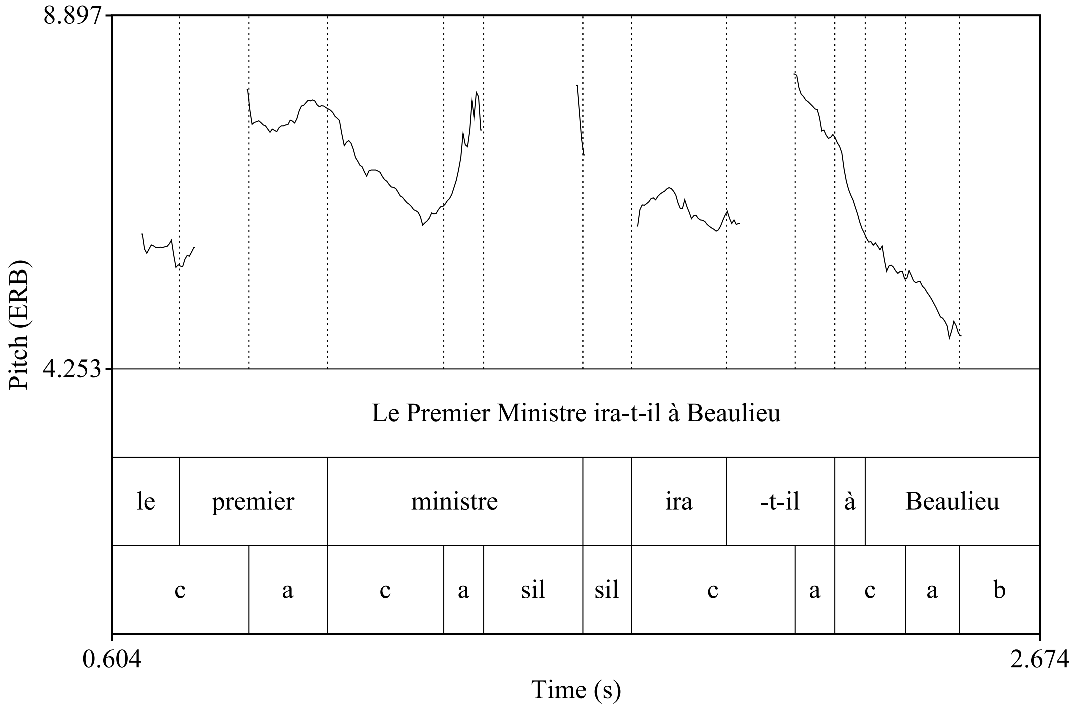


Figure 4.4. *An utterance divided into Tilt intonational event types*.

the French sentence *Le Premier Minstre, ira-t-il à Beaulieu* 'The Prime Minister, he will be in Beaulieu'. The sentence is a good example of the overall labeling procedure used for the analysis, as it shows how boundary tones may appear in non-final or phrase-medial position; in this case, the events associated with the final syllables of *Ministre* and *Beaulieu* qualify as boundary tones, since they are followed by both syntactic and prosodic breaks (the speaker paused after each one for what would be a level 3 break in ToBI). In longer sentences with fewer strong syntactic boundaries between APs, boundary tones were only marked at the end of what Delais-Roussarie et al. (2013) term intermediate phrases (ips) and intonational phrases (IPs), not at the end of regular (i.e. non-phrase-medial) APs. These boundaries generally corresponded with orthographic breaks marked by commas, semicolons, and periods, but not always, as speakers would occasionally parse the sentence differently. In practice, this means that there was a subtle yet noticeable positive correlation between utterance length and the amplitude of the **a** events, which tended to shrink in size in response to the potentially higher number of rising **b** events. This trend was especially noticeable in utterances beginning with topicalized constituents, like left-dislocated subjects, which were very often marked with high-amplitude rises. **C** events were on average much longer than either the **a** or the **b** events; this was expected, considering the relatively simple phonological structure of French intonation. As shown by the figure above, **sil** events were marked anywhere there were disfluencies in the F0 contour, but crucially not during voiceless consonants and devoiced vowels, as this would have distorted the duration of the **c** events. In the example above, the **sil** falls between the coda cluster of *ministre*, which causes the dip in F0 immediately following the **b**, and the onset of the initial vowel in *ira*. In this way, every portion of the sound file was labeled, and the event sequence was fully continuous.

Labeling procedures for L2 French followed the phrase boundary procedures established in Chapter 3. An **a** event was marked each time a syllable was realized with an English-style pitch accent, like H*L, or any of the AP-final pitch movements from French. Boundary tones were marked at syntactically-determined ip and IP boundaries, as well in places where the syllable showed phrase-level final lengthening. These same procedures were also used to label events in the L1 English, with boundary tones only falling on continuations and IP-final pitch movements, and pitch accents falling everywhere else. In general, event labeling is much less prone to misclassification errors than phrase marking, since the Tilt model only interests itself in the significant changes to the F0 contour and not to the underlying structure of the utterance, so the process was relatively straightforward.

Regarding the labeling procedures used here, it is also worth noting that event boundaries were placed in line with vowel onsets rather than syllable onsets, which Taylor (1999) suggests increases the model's linguistic meaningfulness or relevance. More straightforwardly, perhaps, we could simply say that leaving consonants outside of the event boundaries reduces variation in the data, since the consonants themselves add time to each syllable's duration, and somewhat unpredictably. The downside to this approach is of course that it shortens events beginning before the onset of the vowel, for example in syllables beginning with nasal consonants and other voiced continuants, which shortens not only the average duration for each type of event, but also its standard deviation, making the distribution of values seem more uniform than they actually are. Even in these cases, however, pitch often remains constant throughout the syllable onset, changing only at the beginning of the following vowel, meaning that the F0 data would probably be best classified as belonging to the previous continuation rather than to the following accent or boundary tone.

Once the events were labeled, several Praat scripts were written to extract the data needed for Tilt analysis, which included amplitude, height, duration, alignment, and tilt. Duration was normalized for each speaker to reduce variation in the values for parameters relating to time, which were all but the two relating to pitch (amplitude and height). As with the pitch values, the normalization was done using the z-score procedure, which calculated the mean and standard deviation of the duration for the voiced events, i.e. those classified as either **a** or **b** pitch accents; **c** events, or continuations, varied widely in length and were thus not included. The main reason for doing this was to account for the fact that average vowel length can vary not only by speaker, but also by language (Ramus, Nespor, and Mehler 1999; Ling, Grabe, and Nolan 2000; Grabe and Low 2002), meaning there was potential for the learners to be at an automatic disadvantage in producing French pitch accents with native-like duration. Following this normalization, alignment parameters were calculated as the percentage of the normalized syllable duration elapsed at the time of the event peak, with the main goal being simply to ensure that individual differences in speaking rate did not distort the duration data, much in the same way as normalizing pitch values prevents the distortion of the values for amplitude and tilt. Because alignment for pure rises will always equal duration—the peak is not reached until the end of the accent—the method was less important for the French L1 data than for the English L1 and French L2 data, which both had a high number of complex (i.e. either humped or scooped) pitch accents.

4.3 Methods

4.3.1 Basic Properties of HMMs

Before explaining how the Tilt model can be used to answer the learner variety hypothesis question, it is worth discussing the basic properties of HMMs, which are one of the most popular models used for describing the statistical properties of sequential data. As their name implies, the models are defined by two main features: that they satisfy the Markov property, and that they are in some way hidden. The Markov property, which appears in a handful of other statistical models (e.g. Markov chains and Markov fields), simply says that the probability of encountering one item in a sequence only depends on the item that came immediately before it—in simpler terms, the sequence has a very short-term memory. By extension, the property means that for any given sequence of items, we can predict the next item in the series by knowing something about the current item and nothing else, since it is not important to know what item appeared, say, three steps back in the sequence. When the Markov property applies to a series of events, the series is called a Markov chain. Many different kinds of data can be modeled with a Markov chain, from meteorological data, like the likelihood of a sunny day following a rainy day, to genomic data, like the likelihood of finding a G after an ACGGT sequence in DNA. In language studies, the model has been employed in corpus linguistics to construct language models that describe word order (Vogel, Ney, and Tillmann 1996), predict phonotactics (Shen and Reynolds 2007), and disambiguate ambiguous morpheme sequences (Adler and Elhadad 2006).

What separates HMMs from these simpler Markov chains is the idea of their being *hidden*. Although this sounds mysterious, the basic idea is straightforward. In a Markov chain, the only components of the system are the outputs, so we can see every part of the chain, and the probability of encountering one item only depends on the previous one. In an HMM, however, the outputs are in fact being generated by underlying *states* and not by the previous outputs themselves. These hidden states—they are hidden because we cannot directly see them—not only have probabilities for producing each of the outputs, but also have probabilities for transition to each of the other states. The diagram in Figure 4.5 (following) shows the basic structure of the system, with the hidden states X1, X2, and X3 being tied to the outputs y1, y2, y3, and y4. The transition probabilities between states are represented by a12, a23, and a21, and the output probabilities are represented by b11, b12, etc. This state-output structure has been widely applied in the speech sciences—practically every recognizer used for TTS or speaker identification is based on an HMM—but it has also found applications in finance (inferring market growth or decline based on volatility data) and public health (inferring disease status based on a sequence of presented symptoms). What makes the model so useful for studying

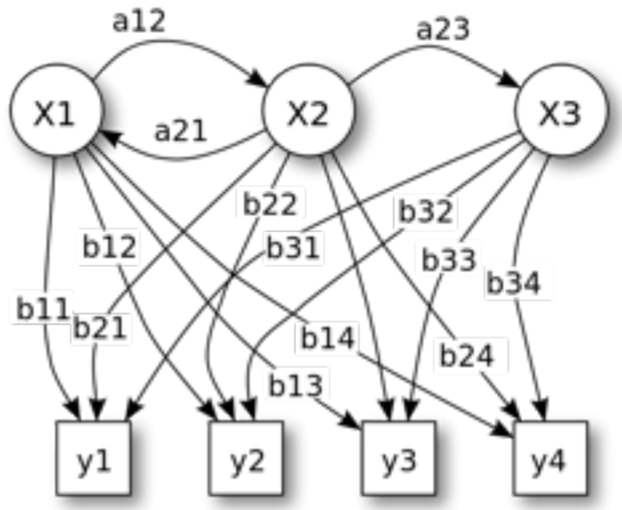


Figure 4.5. *Structure of a simple HMM*.

linguistic data, however, and especially for studying intonation, is that it can capture not only the likelihood of particular sequences of events, like morphemes or pitch accents, but it can also capture their average characteristics, since the outputs may either be discrete (letters) or continuous (parameters like amplitude or duration). The output probabilities may also have any kind of distribution, including mixtures, so a small number of hidden states may be able to produce a large number of distinct outputs. Another way of thinking about this is to imagine that each of the states is an urn containing balls of different colors, and that each urn has a different proportion of one color to the others. The likelihood of seeing a red ball on the surface, then, depends not only on the state sequence (if one urn had no red balls and a 90% chance of transitioning to itself, not many of them would be selected), but also on the mixture of balls in the urns themselves.

The model in Figure 4.5 may look complex, but the basic idea is that the while the outputs are visible, the states are not, and so we can only infer their sequence by calculating its likelihood based on the sequence of outputs. The formal assumptions of interpreting intonational systems using this framework are discussed below. The key point, however, is simply that the model gives us a robust statistical method for estimating the parameters of the hidden states (e.g. the underlying representation or sequence of intonational events) based on information on the surface (e.g. in the F0 contour). Because the data were hand-labeled, the HMM did not need to perform a detection function like the one described by Taylor (1999), meaning that it could also function in a more abstract and less purely acoustic way, which makes it especially attractive for studying L2 intonation.

Before continuing, it might be helpful to review a simple example of an HMM to more concretely conceptualize how the models work. Imagine that, as part of the guided tour to his chocolate factory, Willy Wonka decides to offer children a taste test of his new Wonka's Wondrous Jelly Bean Blends. Each blend has its own distribution of flavors, so he decides to let them try beans from each blend (for the sake of simplicity, assume he has one box of beans per blend). Instead of handing them the candy directly, however, Wonka wants to make things more interesting by blinding the tasting; so, he takes the boxes into a room where the children cannot see him, chooses a box according to some random process, and then sends the beans out one by one on a conveyor belt. As the beans come out, one of the children, who is not only very curious but has a keen interest in statistics, decides to figure out the order of blends from which Wonka is choosing the beans. Wonka's selection process can be modeled with an HMM, and the child could satisfy his curiosity by using the Baum-Welch (explained below) algorithm to establish the parameters for the model, and then find the most likely sequence of blend boxes given a sequence of beans using the Viterbi algorithm. The problem seems hard, if not impossible, to solve at first, but the combination of the model and the algorithms make it not only feasible, but also very computationally efficient.

4.3.3 Architecture of the Intonational Variety Classifier

The HMM-based classifier used to test the learner variety hypothesis was relatively simple. Separate HMMs were established for the two L1s, with each model being composed of three states, i.e. one for the three intonational events: **a**(ccents), **b**(oundary tones), and **c**(ontinuations). Each model was initialized using parameters estimated from the data analysis and then trained on native speech from the three corpora using the Baum-Welch algorithm, which uses the Expectation Maximization (EM) algorithm to estimate model parameters based on a sequence of observed states. In simpler terms, the algorithm takes guessed values for event parameters like amplitude and duration and refines them by processing the events one at a time and updating the estimated values until they match those in the actual data as closely as possible. Although it may seem inefficient—essentially, it is a very sophisticated version of trial and error—the algorithm is computationally feasible and widely used in speech processing (see Rabiner 1989 for an overview) and other sciences to model sequential data.

Perhaps the most important feature of the classifier is its state structure, which makes a number of changes to the original model described by Taylor (1999). His model, along with most HMMs designed for speech processing, uses a relatively sophisticated algorithm to detect intonational events in an unlabeled F0 contour, but it uses a relatively simple algorithm—the n-gram language model—to make predictions about how they are sequentially ordered. In this system, the HMMs are used to model the intonational events themselves rather than their order, and they are defined using an arbitrary number of hidden states; Taylor uses three, but there is no clear motivation for the choice other than that it is what most people use to model other kinds of linguistic information, like feature values for vowels and consonants. The n-gram model is computationally efficient, but it is not as linguistically intuitive as a full HMM, since it does not propose hidden states governing the output of the events. Instead, the probability of the next event occurring is based solely on the previous event, rather than the *state* that generated the previous event. This distinction is subtle, but it has important ramifications for the linguistic realness of the intonational model. First, there is strong evidence that intonation interacts with and/or is governed by other components of the grammar (Gussenhoven 2002; Ladd 2008), particularly syntax and pragmatics. Discourse structure and information structure (Steedman 1999, 2000) play a large role in shaping intonation as well, so on the whole, a sequence of pitch accents and boundary tones would seem to be motivated more by some underlying speaker state or combination of speaker states (e.g. doubt, mood, or interrogativity) than by a single preceding event. The same is true for word order, but because there are many more words in any given language than tonal categories or intonational event types, it is not practical to use a true HMM.

A second reason we should consider using an HMM and not an n-gram model to describe sequences of intonational events is that intonational events are by definition classes of pitch movements that have acoustic features in common. Transitioning from one event to the next, then, is very much like transitioning between the hidden states of an HMM, with each state comprising a distribution of likely values for the different parameters and the individual outputs being the phonetic realizations of the events. What makes this idea even more appealing is that, at any given moment, we do not know why a speaker would be more likely to produce one kind of event (say, a pitch accent) over another (a boundary tone); there are general patterns we could use to make our prediction, like the fact that **a** events are more often followed by **b** events in French than in English because of its particular prosodic structure, but this sort of very rudimentary guessing is the best we can do. If we treat the events as hidden states, however, this uncertainty is built into our model, and we can account for the likelihood of moving between event types by simply adjusting the transition probabilities during training, which using the Baum-Welch algorithm is very easy to do.

A final reason why it is theoretically favorable to treat the intonational events in a Tilt analysis as hidden states in the HMM rather than standalone outputs is that there is clear evidence that two event types may produce pitch movements that are very similar if not identical along one or more of the model's acoustic parameters. In French, for instance, **a** and **b** events both produce rises, and in English, they both produce falls. The events may differ in their duration and amplitude, but as we will see in the data below, their tilt values are often very close, meaning that they have the same basic shape, a similarity that we can model in the HMM by allowing the hidden states to transition to each of the outputs. The added benefit of this idea is that there is no limit on the number of output states we can establish for the model, which allows us to account for the variety of ways the events may be realized in the F0 contour. In an n-gram model, this variation is built into the probability of parameters assigned to each event type (i.e. **a**, **b**, or **c** in the basic Tilt model), so that while the likelihood of encountering a specific sequence of events is handled by the transition probabilities, the actual shape of the output is selected from a single mixture distribution. Using the HMM to model event sequences, however, increases the model's resolution, since each of the outputs will have its own mixture, and we can propose as many outputs as we need for the whole system to be linguistically realistic.

The architecture of the HMM used to evaluate the learner intonation is exemplified in Figure 4.6 below, where solid lines connect the hidden states and dotted lines the states to their outputs. Connections between the **b** and **c** states and the outputs are omitted from the diagram for the sake of simplicity (including them all would have been messy). As described in Taylor's (1999) original paper, the outputs are multivariate Gaussian mixtures that account for each of the five parameters needed for the Tilt analysis, i.e. amplitude, duration, tilt, height, and alignment. The distributions for phonetic data are typically normal, which is why Gaussians are used here, and they are structured as mixtures to allow the outputs to handle more than one type of pitch movement, like high and medium rises (e.g. L*H% and H*H in the ToBI system).

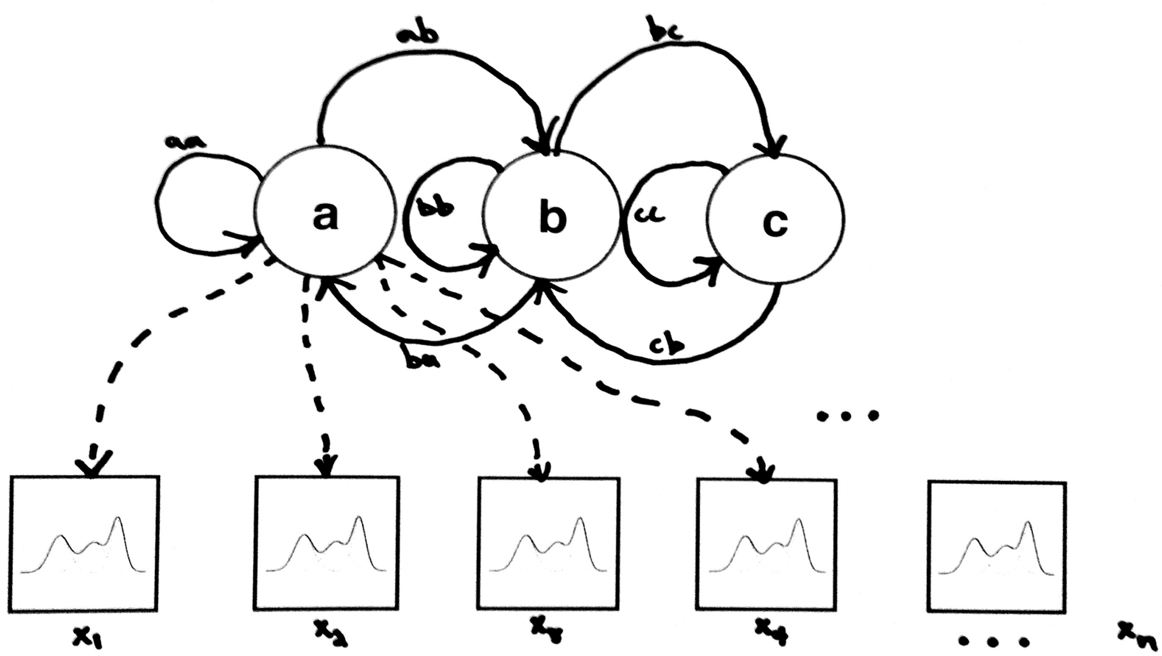


Figure 4.6. *An HMM-based intonational classifier.*

The graphic captures the following features, which are the model's most important characteristics:

1. Each event type (e.g. **a**, **b**, **c**, or **sil**) is an assigned a hidden state;
2. Movement between event types is represented by state transition probabilities;
3. Event types produce pitch contours with tilt parameters;
4. Outputs are theoretically unlimited in number.

As discussed above, the first two features are what distinguish this classifier from the one proposed by Taylor (1999), where the event types are given separate HMMs and then combined via an n-gram probability model. In Taylor's theory (1998), the events are class types that may be realized with sometimes very different phonetic parameters, so the suggestion that they be states in an HMM and not HMMs themselves makes sense. Moreover, the first two features also provide formal linguistic justification for the number of hidden states composing the HMM, a number that is usually determined arbitrarily (Rabiner 1989) and then adjusted to maximize the model's performance. The first three features also mean that the model can capture very language-specific information from the F0 contour, since not only the most likely sequence of events (i.e. the hidden states), but also the most likely parameters for the corresponding pitch movements (i.e. the outputs) are represented by the two types of transition probabilities. This structure gives the model the power to make cross-language comparisons in phrasing, since the transition probabilities model both the likelihood of switching event types mid-sequence and the average length of time spent in continuation. In L1 English, where pitch accents are common and boundary tones are relatively rare, we would expect **a** events to have a high probability of transitioning to **c** events or other **a** events, and in L1 French, where pitch accents are rare, we would expect the opposite, that is for **a** events to transition more often to **c** and **b** events. The low likelihood of encountering pitch accents in French also means that more time will be spent in continuation, so we would also expect the duration parameter for **c** events to be significantly higher than in English. These suggestions are in fact borne out by the results of the classification test described below.

The fourth feature of the model, i.e. that the number of outputs is theoretically unlimited, is a basic property of HMMs and not original to their implementation here. However, the number must be specified for the Baum-Welch algorithm to work, so it is worth giving some thought before training the model on L1 data. One option is for there to be a one-to-one correspondence between states and outputs, which in our classifier would yield three outputs, one for each of the major event types (again, **sil** is omitted from the model because the data were drawn from different guided reading tasks and would thus give biased estimates for the amount of typical silence in speech from the two languages). The benefit of this choice is simplicity, but the drawback is that it probably underestimates the number of possible phonetic realizations for the event types. Even in a language with a relatively simple intonational grammar like French, there are many ways for a single string of tones to surface in the F0 contour, e.g. as sharp rises or level tones for LH*, and even though the Gaussian mixtures in the outputs can handle some variation, using only three of them to model an entire system would be undesirable. A second choice is to start with a predetermined number of outputs, and then steadily increase their number while training the HMM to see which one performs the best in a predetermined classification task, like identifying speech from the L1 corpus. Although this method has no real drawbacks, it is relatively inefficient, and it runs the risk of over-specifying the model's structure by sacrificing linguistic relevance for computational power.

The third option, and the one that is implemented here, is to simply set the number of hidden states, and thus outputs, equal to each language's number of proposed tonal categories. One benefit of the approach is that it mirrors the two levels of representation assumed by the AM framework, e.g. the underlying (events/hidden states) and the surface (event realizations/outputs). It also captures the idea that separate event types can produce very similar phonetic outputs, as is the case in French, where both the **a** and **b** events lead to rises with predominantly identical alignment. Stemming from these two features is the fact that this approach makes the model very language-specific in the sense that its structure adapts to not only the phonetic but also the phonological structure of the speech used for training. Technically, this makes the model more stipulative than Taylor's original implementation (and, in truth, than the Tilt model itself is intended to be), but the increase in precision seems to justify the reduction in objectivity. The one clear drawback with this method is that there is bound to be some error in specifying the number of outputs. There is no precedent in the literature for establishing a tonal inventory based on phonetic and not phonological information—the latter is the basis for ToBI—and even contours with underlying contrastive tone sequences may end up having similar realizations in F0, like H*L% and L+H*L% in English. In an ideal world, there would be a one-to-one correspondence between a tonal category and a particular kind of pitch movement, and we could simply set the number of outputs equal to the size of the tonal inventory. Even though this is not the case, however, the model can still work quite well, since the multivariate Gaussian output distributions can handle some variation within the broad phonetic classes (e.g. rises and falls). Approximating the number of unique pitch movements should be good enough, then, and should increase the model's validity without leaving it vulnerable to overspecification or overfitting.

The model's power as an intonational classifier also increases with its language-specificity, as the more distinct the base models are for the two L1s, the lower the probability of their producing a type-I error (i.e. a false positive) in recognizing the L2. If the opposite were true, and the base models had very broad distributions for the L1 intonational parameters, then it would be much more likely to misclassify the L2, either determining that it is distinct when it is not, or determining that it is more similar to one of the L1s than it really is. In either case, the spurious result would be due to model error and not the inherent variability of the data, which is something we clearly want to avoid.

Before describing how the model may be implemented using real data, a few points regarding its application outside of intonation research should be mentioned. Although its main goal is to quantify linguistically-relevant changes in pitch, the classifier could also be used to study cross-language differences in other suprasegmental features, especially vowel length. The literature presents several methods for classifying languages based on rhythmic typology, but all are based on descriptive rather than inferential statistics, meaning they can only make generalizations about the size and not the sequence of durational phenomena. Even in syllable-timed languages like French that have a relatively uniform distribution of vowel lengths, we might still be interested in knowing the position of long elements relative to short elements, for example whether lengthened syllables can appear outside of phrase-final positions, or whether their distribution is somehow dependent on other factors like speaking rate or pitch range. An HMM-based model would allow us to study this kind of sequential information without losing the ability to describe the distribution of vowel length via mean and standard deviation, and it would also give us a rigorous method for testing the rhythmic similarity of two languages, which would be useful for enriching our knowledge of rhythmic typology.

4.3.4 Model Training

Two HMMs were established for each of the L1s and trained on speech from the corpora described in Chapter 3. Data from the discourse completion task (DCT) were not used for training, since they were heavily scripted and did not elicit natural speech, but data from the guided interviews in the Montpellier-Rognes (MR) and Projet PFC corpora were used. Of the two, the PFC featured more heavily in the training data, as the MR corpus, although drawn specifically from speakers of MF and thus an excellent resource for the classification task, contained a significant amount of speech from highly proficient but non-native speakers and was thus more time-intensive to process. For training the English model, data were drawn from the storytelling task described in Chapter 3, where the learners spoke informally for two to three minutes about a recent social activity, like a family vacation, or an outing with friends. The structure and pragmatic content of the data were similar to those of the guided interviews in the PFC, where interviewers encouraged speakers to tell stories about their hometowns, childhoods, and family histories.

The HMMs were initialized using uniform values for the parameters estimated from the data during preprocessing. Transition probabilities were set to 0.5 so that each event type had the same probability of transitioning to the next type as of returning to itself. Because the Baum-Welch algorithm re-estimates the probabilities during training, it only needs a reasonable starting point for initialization, which Rabiner (1989) suggests should either be stochastic (i.e. random) or completely uniform. More care was taken in estimating the output parameters, however, as these were directly observable in the speech data and thus able to be estimated with a relatively high degree of accuracy. For both the L1 varieties, duration and amplitude were set to 0 for each output. Because the two parameters were normalized during preprocessing, the distribution of their values is centered around their mean, which in terms of z-scores or standard deviations from the mean is equal to 0. The values for alignment and tilt were approximated for each of the outputs based on the general shape of the corresponding pitch movement. In French, for example, tilt for the rising output (there was only one) was set to 1.0, the equivalent of a full rise, and its alignment was set to 0. For English, tilt for the two rising outputs (one for L+H*, and one for L*+H) was also set to 1.0, but their alignments were different, with one being set for 0 and the other for 0.5, indicating a starting position about halfway between the vowel onset and endpoint. It is worth mentioning here that unlike duration, where 0 indicates an average length and not an actual zero-value, the values for alignment indicates the average position of the event's beginning relative to the vowel onset, given in normalized units of time. So, a duration of 0.5 means that the event typically begins half a normalized time unit away from the onset, an alignment of 1.0 a whole time unit away from the onset, and so on. As with the raw values used to study alignment in Chapter 3, negative values represent events that begin in the preceding syllable, which in the data was rare. The reason that the alignment data were not themselves normalized is that doing so would rob the model of its ability to distinguish language varieties based on the parameter—differences in timing of, say, the late-L elbow in French and English rises would be transformed away.

After training was complete, the forward-backward algorithm was used to produce probabilities that the L2 intonation was generated by the L1 HMMs. Data from the storytelling task were used to produce individual estimates for each of the speakers, and then the probabilities were pooled to see how successful the learners were on average at producing French intonation. The average was compared to the same estimates for native French speakers *not* used for model training, and a Welch's t-test was used to compare the means for the two speaker groups. P-values were then calculated to determine if the difference between the groups was statistically significant. The same procedures were also used to determine the likelihood of the L2 speech being generated by the model trained on L1 English. If the learner variety hypothesis is true, then there should be a statistically significant difference between the pooled estimates for the L2 and the estimates for the two L1s. If it is not true, or at least not tenable based on the data gathered here, then there should not be a significant difference between the L2 estimates and one set of the L1 estimates—in other words, the two varieties will be so close as to be statistically indistinguishable. For this conclusion to make sense, the classifier would also need to produce the same results for L1 data being run on its own model, otherwise it would clearly not be specific enough (in an assessment sense) to judge the nativeness of the L2 data. L1 from speakers not included in model training were tested using their respective L1 models; the classifier judged both the French and the English data to be native. In either case, comparing the two sets of L2 estimates gives us a rough but objective idea of how far the learners have advanced in acquiring the French system, something that the current models of L2 intonation cannot reliably do.

4.4 Results

The classifier produced likelihood estimates for each of the speakers' L2 intonation being produced by the two L1 HMMs, which are shown in Table 4.1 (following). Before discussing the implications of the data, there are two important things to mention. First, because separate models were trained for each L1, the per-speaker estimates need not (and most often will not) sum to 1; low values for one typically coincide with higher values for the other, but the trade-off is not required. Second, because the classification task had a binary outcome—the L2 variety either had a good chance of belonging to the L1 or it did not—values close to 0.5 indicate that the algorithm was no better than chance at determining the source variety for the data. Likewise, values close to 0.0 indicate that the model L1 was probably not the source variety, and those close to 1.0 that it was. In terms of answering our research questions, a high score for English

Table 4.1. *Probabilities for L2 intonation being generated by an L1*.

Speaker	English Probability	French Probability	Speaker	English Probability	French Probability
1f	0.51	0.63	11f	0.62	0.21
2f	0.78	0.35	12f	0.78	0.15
3f	0.83	0.29	13f	0.39	0.68
4f	0.45	0.85	14f	0.56	0.42
5f	0.88	0.24	15f	0.73	0.33
6f	0.66	0.44	16f	0.49	0.52
7f	0.35	0.65	17m	0.66	0.42
8f	0.21	0.15	18m	0.39	0.73
9f	0.10	0.90	19m	0.51	0.48
10f	0.65	0.31	20m	0.15	0.23
		All	μ	0.535	0.449

would indicate L1 transfer, a high score for French would indicate proficiency in the L2, and moderate scores for both would indicate a blended intonational system. Just as importantly, the LVH would be strongly confirmed only by low scores for both tests, since a distinct variety must necessarily be classified as something other than English or French. More research will be needed to determine an exact threshold for this distinction, but for our purposes, a likelihood of less than 0.20 will be used as the cutoff point (the score would mean the classifier is only 20% sure that the L2 belongs to the model L1, which given a large enough sample size is most likely sufficiently low to assume phonetic independence).

The most relevant finding is that although only two of the speakers reached native-like probabilities, many of the other speakers came closer than we might have expected based on other phonetic data, like alignment value for the late L-elbow or the scaling of phrase-final rises. The other important result is that for a few of the speakers, their L2 French had a very high probability of being classified as L1 English—in other words, they have made little progress in a acquiring French intonation. Typically, these were also the speakers who produced the fewest AP-final rises, using single H* or L* pitch accents instead and bringing the phonological aspects of their intonation much more in line with what we would expect for L1 English. Considering the alignment and scaling data from Chapter 3, which support Jun and Oh's (2000) findings that learners acquire intonational phonetics later than phonology, this result supports the conclusion that these learners transferred both aspects of the L1 prosodic system to the L2. For comparison, the L1 likelihood estimates for the native speakers of French are included in Table 4.2. The scores are for the most part very high, but some are lower than we might expect. In future experiments, homogeneity in the L1 classification data might be improved by controlling for task

Table 4.2. *Probabilities for L1 French intonational varieties*.

Speaker	French Probability	Speaker	French Probability
1f	0.95	1m	0.91
2m	0.88	2m	0.88
3f	0.91	3m	0.87
4f	0.79	4m	0.96
5f	0.93	5m	0.93
	All	μ	0.901

type (e.g. guided interview vs. DCT) or paralinguistic factors (e.g. speaker mood states) when gathering the training data set.

4.5 Summary

The results from the classification test generally support those in the literature, as well as those from Chapter 3, but the number of learners obtaining relatively low likelihood estimates for L2 French is surprising. As mentioned above, one likely confounder here is duration: because the Tilt model uses time values (albeit normalized ones) in calculating the duration and tilt parameters, differences in the overall variability of tonal duration in L2 French would lower the probability of its being classified as either of the L1s. Learners who speak either more carefully or less fluently in the L2 will bias the data in favor of the learner variety hypothesis (LVH), since those parameters will swing away from the normal values for either of the L1s, and learners who speak confidently and fluidly but not accurately will bias the data against it. This is one area where the subjective judgment of researchers might be very useful in gathering the data before it is preprocessed and labeled, since procedures for controlling variables like speaking rate and speaker mood could be included in the experimental methods, though perhaps not as straightforwardly as we would like.

The data being what they are, however, we can draw some interesting conclusions about the phonetic characteristics of the L2 intonation as it relates to the L1 systems. Contrary to the predictions of the LVH, with the learners' likelihood estimates being so decidedly heterogenous —the scores ranged from 0.15 up to 0.9—it is unlikely that the speakers were all exploiting the same intonational system in their L2. These data confirm the alternative hypothesis that while learner varieties may share some phonological features cross-linguistically, it would be more accurate to say that they are composed of bits and pieces of the native varieties that are realized with varying degrees of success depending on speaker-specific factors like length/strength of exposure to the L2 and prosodic sensitivity. On a related note, despite the large spread of these values, learners tended to fall into one of three categories based on their scores: high English and low French, high French and low English, and both low French and low English. It follows that those in the first category are far from acquisition, those in the second category are close to it, and those in the third category are somewhere in between, with their intonation either being too disfluent to process effectively, or their phonetics actually being distinct from the two L1s. Finally, although we do not need an HMM-based model to demonstrate this point, the fact that no speaker's intonation was classified as being similar to both French and English shows that the phonetic systems of the two languages are in fact distinct—a nice result, but not one that is particularly interesting.

CHAPTER 5

CONCLUSION

This dissertation has presented two methods for measuring learner progress in acquiring non-native or second language intonation. The first was based on contemporary models of intonational phonology and examined how learners implemented L2 tonal categories using L1 phonetics. Results indicated that learners could successfully acquire certain aspects of the L2 intonational system, including phrasing and alignment, but that full acquisition was impossible given the length of exposure to the target language, based on evidence gathered from production tasks completed before, during, and after intensive coursework in the target language, French. The second method was based on Taylor's (1998; 1999; 2000) Tilt intonational model, which was adapted for the purpose of measuring the phonetic similarity of the L1 and L2 varieties. Results from that analysis fell in line with those from above, with the L2 intonation being statistically distinct from the two L1 varieties. The first set of results does not support the learner variety hypothesis (LVH) proposed by Mennen, Chen, and Karlsson (2010) as characterizing L2 intonation, but the second set of results do (cf. p. 164 and 168-9). The discrepancy between the two methods is likely due to a combination of paralinguistic factors influencing the phonetics of the L2 intonation, although it is possible that the phonological analysis was too coarse-grained to account for subtle differences between the varieties.

This chapter offers potential explanations for the results of the previous chapters, and it also suggests several areas for future research that may shed new light on the specific phonological and phonetic nature of learner intonation. It also briefly addresses issues with one of this dissertation's secondary research questions, which is how the methods presented in Chapters 3 and 4 might be used to develop pedagogical materials for improving L2 intonation. To the author's knowledge, there are no computer programs, textbooks, or courses designed for learning intonation, and even though many (if not most) of the available materials mention suprasegmental features like stress and rhythm, they do not go into enough detail to make clear, measurable progress a reality for most learners.

The chapter is organized as follows. Section 5.1 presents evidence for and against the LVH, using data from the previous chapters to determine its theoretical strengths and empirical weaknesses. Section 5.2 addresses several methodological issues raised by the production experiments conducted in Chapter 3, and it makes some small suggestions about how researchers might reduce measurement error when working with speakers of an L2. Section 5.3 addresses one of the primary research questions for the dissertation, which is whether either of the models of L2 intonational phonetics presented in the previous chapters could be used for adapting a perceptual model of SLA like the SLM or PAM to study learner intonation. In doing so, the section also presents a rough formal framework for using phonetic information obtained from production experiments to establish the cross-linguistic equivalence of language-specific intonational categories. Section 5.4 offers a number of simple ways that the phonological and purely phonetic approaches to intonation could be used to develop pedagogical materials, and Section 5.5 summarizes the findings of the previous chapters and offers answers to the research questions outlined in Chapter 1. Finally, because the chapter touches on a broad variety of topics that sometimes only have loose theoretical or practical intersections, directions for future research are suggested at the end of each section rather than being given a section of their own.

5.1 The Learner Variety Hypothesis

The production experiments and analytical methods presented in Chapters 3 and 4 provide some evidence that the LVH holds true for intonation as Klein and Perdue (1997) suggest it does for segments, but perhaps not as clearly as one would like. The first issue is that under an AM-based analysis, the L2 French produced by the speakers in this study contained tonal categories present in both L1 English and L1 French, and the English categories were often complex (e.g. bitonal); the L2 variety also contained no categories not posited for at least one of the L1s (cf. p. 170). Both trends go against the hypothesis, which predicts that the learner variety will be distinct (at least phonologically) from the native varieties, and the clearest explanation for them is simply transfer from the L1 to the L2, which, if we take stock in any of the leading theories of L2 phonology (e.g. the SLM, the PAM, or the perceptual magnet effect model), is exactly what we would expect for tonal categories that the learners perceive to be phonetically indistinguishable. Based on the L2 French data, there really is no reason to assume that non-nativeness in intonational phrasing or phonetics stems from anything other than this kind of transfer. The strongest piece of evidence in favor of this conclusion is the coincidence of low pitch and high intensity in ip-final rises with delayed L-elbows; the shape of the contours supports a bitonal analysis, and the emphasis on the first half of the syllable suggests the presence of lexical stress, both of which are characteristic of the English L*+H pitch accent. The learners in Mennen, Chen, and Karlsson's (2010) study produced mostly simple H* and L* pitch accents, which the authors suggested showed that they were tapping into a simplified but universal intonational system and not the L1 Punjabi and Italian systems. However, if this simplified system were also to contain complex pitch accents, which they say it largely does not, and if its structure is identical to reduced versions of the learner's L1, than the only logical conclusions seems to be that it is a mixture of the L1 and L2 systems, i.e. a variety based on phonological and/or phonetic transfer.

We can also find support for L1 transfer in the L2 intonational phrasing. In general, the only phrase boundaries the learners produced with moderate accuracy were either syntactically (e.g. questions) or pragmatically (e.g. continuation) marked, and they struggled to group words into appropriately-sized APs. The marked boundaries often coincided in terms of positioning (e.g. after a topicalized constituent, or before a major syntactic break) and structure (e.g. rising sequences for questions) with English, in which cases learners often implemented the French tonal categories using their native phonetics, and the APs themselves were often replaced by a series of independent pitch accents. More straightforwardly, the learners really only produced boundaries they would have produced in English, and they did not succeed in producing a majority of the IP- and ip-medial APs. Both are signs that the learners relied on English intonational phrasing to produce the French contours, and when taken with the abundance of bitonal pitch accents would suggest that their variety of French on the whole is more similar to English than to the simplified underlying variety Mennen et al. posit.

Learners also tended to scale phonologically similar rises the same in French as in English, whether they were sentence-medial continuations, ip-final continuations, or IP-final polar questions. There were of course some exceptions to this rule, but overall, the L1 English and L2 French data were remarkably consistent, again suggesting L1 transfer rather than the realization of a universal learner variety. It is not entirely clear how this may be incorporated with the LVH, but one possibility is that the apparent structural simplicity of learner varieties is due to the misperception of L2 tonal categories and resulting transfer of those from the L1 mixed with paralinguistic factors like nervousness and self-doubt that influence how they are realized phonetically.

Still, both Mennen et al. and Klein and Perdue (1997) found evidence of structural similarity in the intonation, morphology, and syntax of the learner varieties, which must be accounted for if the LVH is going to be challenged. In the case of intonation, we can say that the presence of common tonal categories in the L2 English of L1 Punjabi and Italian speakers may simply be due to the phonological similarity of the L1s rather than to the learners exploiting a simplified universal variety. Realistically, L* and H* pitch accents and their basic combinations (LH, HL, etc.) are common throughout world languages (Jun 2005), so it is not surprising that they should appear in an L2, especially if they are common in the target language, as they are French. More surprising would be if the learner variety comprised categories in only one or even neither of the L1s, but again, this seems strongly counterintuitive. A more likely answer is that the processing constraints inherent to speaking a second language limit the amount of information learners can linguistically code (e.g. syntactically, morphologically, and prosodically), and that the constraints are more or less directly related to a combination of factors influencing proficiency, like length and strength of exposure to the L2, age, and general cognitive abilities like working memory (Baddeley 2003), prosodic sensitivity (Clin and Wade-Woolley 09), and phonological awareness (Durgunoğlu, Nagy, and Hancin-Bhatt 2003). Under this analysis, the learner variety would not be distinct per se, but rather the result of limiting the L1 to its most basic components, which many, if not most, languages are likely to share.

Another piece of evidence we must consider in evaluating the LVH, of course, is the result of the Tilt analysis presented in Chapter 4, which found the L2 French to be significantly different (in the statistical sense) from both L1 French and L1 English. One likely explanation for this finding is that the learners produced more variable intonational events in terms of duration in the L2 than in the L1. Duration is a key parameter in the Tilt model, not only on its own, but also as a factor in the tilt parameter itself, and so any significant changes it undergoes in shifting from one variety to the next will lead the classifier to judge them as distinct, even if varieties are otherwise (i.e. phonologically) identical. This is one of the drawbacks of dealing exclusively with phonetic information in the analysis, and it does not necessarily increase the linguistic relevance of the model, since the increase in variability may have been (and in fact very likely was) due to para- or even non-linguistic factors, like the learners' belief states during the DCT, their attitudes toward their L2 French, or even just nervousness. The inverse scenario, where the L2 variety is less variable than the L1 along a particular dimension, was also true. As discussed in Chapter 3, pitch range in the L2 French had both a smaller mean and a smaller standard deviation than range in L1 French, likely due to the same combination of factors as the increase in durational variability in addition to the processing constraints mentioned by Jun and Oh (2000) limiting the learners' ability to prosodically code information in the L2. Thus, changes to the tilt and amplitude parameters did not necessarily reflect changes to intonational structure of the learner variety or its phonetic implementation—rather, they simply reflected the increase in measurement error associated with the L2 speech, which in a general sense can be said to be a worse approximation of the speakers' linguistic competence than their L1 speech (e.g. there is a larger discrepancy between performance and competence).

Of course, the LVH still has merit, albeit more as an attitude toward learner intonation rather than a formal system for analyzing it. Pedagogically, treating the varieties as distinct means means teachers can focus on students' strengths rather than their weaknesses with respect to the target language, and they can work with the structures underlying the variety rather than against them. Moreover, if both researchers and foreign language teachers treat learner varieties as distinct, at least conceptually, then they will be better equipped to understand the articulatory and perceptual challenges faced by speakers of particular L1/L2 combinations. In a general sense, this attitude is no different than the one adopted by proponents of the interlanguage concept in the 1980s (e.g. Krashen 1982; 1984), but in a more specific way, the distinction that a speaker's L2 has properties that cannot be predicted solely by the interaction of the two L1 phonologies may decrease the likelihood of misclassifying speech errors and increase the likelihood of teachers being able to help learners reshape and eventually correct them. This is more a matter of foreign language pedagogy and teacher training than of linguistics, but it is perhaps one way that experimental results from studies of L2 intonation could be brought to the classroom in a meaningful and impactful way.

5.2 Methodological Issues in Pitch Research

The production experiments conducted for this dissertation highlighted some of the methodological difficulties with conducting research on linguistic pitch that should be taken into account not only when considering the results in Chapters 3 and 4, but also when considering new ways to move forward in the field of intonational acquisition. These issues are not normally addressed in the literature, and although they are not likely to affect studies based on large, unguided audio corpora, they may certainly affect laboratory or longitudinal studies, especially since these typically comprise relatively small numbers of speakers; Mennen, Chen, and Karlsson (2010), for example, studied only 6 learners (five females and one male), and Mennen (2004) only 15, with only 55 data points being recorded for each speaker. Any variation, then, in the speakers' moods or belief states will have a larger effect on the outcome variable of interest than it would in a large study, and so it is important that clear methods be established in the experimental procedures for keeping it to a minimum. Procedures like this are common for research in the broader cognitive sciences, but they have not become the norm for research in linguistics, or at least not in research on prosodic acquisition, which limits the reliability of our results and likely leads to the type of contradictory findings that are presented in Chapter 3.

Although it may be difficult to implement, one procedure that might reduce this sort of error would be gathering production data from learners in environments other than phonology laboratories, like households, public spaces, and community centers. Allowing them to relax and socialize with other learners would put less pressure on them to perform well for the tasks, and it would likely increase the chances of them fully exploiting their L2 competency. In scientific terms, this kind of fieldwork might be less precise than traditional lab experiments, since it could not directly test the learners' abilities to produce pitch contours in specific contexts or under controlled circumstances (e.g. the hypothetical scenarios of the DCT used in Chapter 3), but given a sufficient sample size and recording time, enough data could be obtained to make the limitation moot with regards to the procedure's overall power. These methods are often used to collect L1 data, as in the PFC and MR corpora used in this study, but they have rarely been used to gather learner data, and their potential benefits should absolutely make them worth trying.

Another potential workaround for the issue of learner nervousness or doubt during production experiments would be to give participants positive feedback before, during, and after specific tasks, especially those they found to be challenging. In this study, learners were encouraged to ignore errors in pronunciation during the DCT and in diction in the storytelling task, with the goal in both cases being to reliably elicit the smoothest, most expressive pitch contours they could produce. Obviously, this kind of encouragement may help some speakers but have little to no effect on others, but the basic idea of establishing a positive, consequence-free environment in which they can feel comfortable speaking their L2 is crucial for the experiments to be successful. It is suggested, therefore, that researchers should collaborate on developing a set of procedures for establishing these environments in production experiments, and that the procedures should be described in published form along with other methods and materials so that readers have a clearer contextual foundation for evaluating the results. As a practical example, we can consider data collected by Mennen for her study on Dutch learners of Greek (2004). The results were clear—phonetic transfer was bi-directional, and pieces of each L1 intonational system appeared in the speakers' L2—but the experimental protocols are not, such that there is no way of knowing how the researchers interacted with the speakers and what attitudes the speakers had toward the tasks. If the researchers instructed the participants to pay close attention to their intonation, for example, we would expect the results to be different than if they instructed them very clearly to not, especially if they had some conscious knowledge of either language's intonational system. We would also expect the participants' knowledge of the study design and research questions to affect their intonation during the tasks, since speakers in general are more inclined to produce larger pitch movements when speaking carefully (Gussenhoven and Chen 2000), and participants are more likely to speak carefully when they know their speech is being recorded and analyzed for its structural characteristics, especially in an L2. These points are subtle and largely hypothetical, but the core issue here is that intonation is influenced by internal (i.e. speaker-specific) factors as well as external (i.e. language-specific) factors, and it is dangerous to draw conclusions about one without also accounting for the other. If intonational acquisition is going to become a field as broadly developed as segmental acquisition or, more broadly, sociophonetics, we must make it a primary part of our methodology to reduce this kind of variation in our data.

Another issue raised by this dissertation is the frequent use of small sample sizes to make generalizations about the intonational structure of an entire language. Most of the studies cited in Chapter 2 comprise data from 10 or fewer speakers, which may be enough to establish categorial boundaries for their own unique varieties of the target languages, but based on the evidence provided in Chapter 3 would not be enough to do the same for varieties at the community- or language-wide level. There is an important distinction here, too, between the statistical and linguistic significance of results in that researchers often rely on p-values for establishing the validity (i.e. the non-randomness) of their findings, even though the formal implications of the findings themselves may be relatively inconsequential. A case in point is the F-test used in Chapter 3 to determine the similarity of pitch range between speakers, which because of the number of points in the data set and the resulting high statistical power, often found distributions differing by less than .5 ERBs in 80% IQR to be significantly dissimilar. The linguistic relevance of this finding, though, is questionable, since a difference of half an ERB in any context is extremely small, and there is a general agreement in the literature that speakers do not perceive pitch movements smaller than about 1.5 semitones (about 0.6 ERBs) as being important to communication ('t Hart 1981). Even as a single pitch movement, the difference would be barely perceptible, but as a difference in range, it would be much harder, if not impossible, to detect. A basic solution to this problem would be to increase sample size by including more speakers in the corpus instead of gathering more data points from those we already have. The latter is of course important in that it helps us avoid Type I and II errors in the analysis, but the former is necessary for generalizing speaker-specific patterns to a variety or language, and should probably be standard practice. In cases where the sample size is small, it would be very helpful for authors to mention why, since the reasons underlying difficulty of recruitment may impact the inferences we can draw based on their results, e.g. for varieties with only a few remaining speakers, a sample size of 10 might be sufficient, but for those with hundreds of thousands, it would likely not. This suggestion is very simple, but it would serve not only to confirm or disconfirm our results, but also to shed light on areas of methodological weakness in our experiments that could be addressed in future research.

5.3 Intonation and Perceptual Models of SLA

One of the broad goals of this dissertation is to determine whether perceptual models of SLA could be used to study intonational or prosodic and not just segmental acquisition. Although the models have been tested with the acquisition and perception of lexical tone (So and Best 2010, 2011, 2014), Best (p.c.) illustrates the basic issue with extending them to prosody in saying that 'it is more problematic to jump directly to broader levels of prosody (phrasal, sentential), for which there is little to no straightforward evidence of distinctions being strictly contrastive and categorical'.

5.3.1 Evidence for Categorical Scaling

Based on the results in Chapter 3, it seems possible that perceptual models of SLA, like Flege's SLM (1995) or Best's PAM (1995) could be adapted to study intonational acquisition, but not without some refinement. The most obvious problem is that, as they are currently understood, tonal categories do not have consistent cross-linguistic phonetic identities in the same way as segmental categories. In English, for instance, H* tones may be downstepped so much as to be lower in pitch than an L* in the same sentence, and in French, LH* may be scaled higher than a full L+H*H% in English; the categories are strictly relative, and they are really only useful for studying intonation within a single language variety. In order to determine what portion of a speaker's L2 intonational system stems from the misperception of tonal categories—a key goal of perceptual models—the categories must be altered or, more likely, completely recast to avoid inconsistencies in the vein of those described above. Thankfully, researchers have developed purely phonetic models of intonation, like Taylor's Tilt model, which gives us a starting point for considering how such a large-scale restructuring of phonological theory could be effected. Instead of building an inventory of categories based on phonological criteria, then, like whether a contour triggers a change in meaning or creates a particular pragmatic effect, we can build one based on phonetic parameters belonging to broad classes of contours that are consistent across languages. Using the production data from Chapter 3 as an example, we saw that the learners had trouble producing the French LH* contour in all but the most pragmatically marked contexts, relying instead on an alternation between it and their native English L*+H to fill the gaps.

Phonetically speaking, though, we can redefine the contours in terms of alignment and scaling, which are the features primarily responsible for distinguishing them from each other and from other pitch movements in the two languages, and we could construct categories based on where the parameters tended to settle in aggregate cross-linguistic data.

Establishing phonetic boundaries for monotonal pitch accents would be much more difficult, however, and would deal more with the syntagmatic relationships between phonetic realizations of tonal categories in a single utterance than the paradigmatic relationships between their underlying forms. H*, for instance, may be upstepped, downstepped, or held level depending on its surrounding tones and pragmatic context, and its height often drops gradually over the course of a phrase. Dilley (2001) provides a wonderful account of these relationships for English, but her research has not been picked up by the intonational community in general, and it cannot be easily integrated with any of the existing models of intonation, including phonetic ones like Momel/INTSINT, the IPO framework, and the Tilt model. A strong possibility, however, is that by simply redefining pitch accents in terms of their sigma levels, i.e their pitch height in terms of standard deviations away from the mean, we could come close to such a system, since we would not need to relate them back to the syntactic, semantic, or pragmatic meanings they assume in any particular language; this would be the role of the intonational phonology per se, and not of the phonetics. We can see an example of how this might work in the French L2 data, where phrase final rises were generally scaled much lower than in L1 French. Even though the learners had very different individual pitch ranges, they tended to scale AP-final rises to 0σ and ip-final rises to about 0.5σ from the mean—the two were very close in size, regardless of how the other intonational categories were scaled. By contrast, the native speakers of French scaled the same rises to 1σ and 1.75σ on average, high enough for them to stand out from the mean, and far enough from each other to be perceived as categorically distinct pitch movements. On the surface, then, this is a significant problem, since the learners are using a much smaller rise in both cases than the native speakers. English, however, does make use of large rises, particularly in questions, so a hypothetical lack of a 1σ to 2σ rise in the L1 cannot be used to explain the trend in the L2. Also, the learners did in fact use these rises in the L2, but only in certain contexts (e.g. when expressing surprise or asking a yes-no question) and when paired with the non-native alignment of the late L-elbow.

These data point to two preliminary conclusions. First, categorical differences in scaling are not only theoretically possible, as discussed by Ladd (2006), Gussenhoven (2002), and other researchers (Mennen, Chen, and Karlsson 2010), but they are in fact real and demonstrable on the basis of experimental data. Second, learner intonation can be analyzed in terms of the distributional and not just the realizational aspects of these scaling-based categories. Accounts of L2 intonation within the AM framework work from the top down by positing an inventory of intonational categories for the variety and then examining how those categories are realized phonetically, but there is no reason our analysis should not also go the other way, i.e. from the bottom up by developing an inventory of categories based purely on generalizations drawn from phonetic data. If z-score normalization works for lexical tone (Rose 1989), then it is probably robust enough to work for at least some aspects of intonation, if not most, and we could use it to build this inventory in much the same way as the IPA was built for segments. This idea is sketched out in the following subsection, but much more research will be needed before it can be formally integrated with existing intonational theory and then implemented for analyzing L2 intonation.

5.3.2 Establishing Phonetic Intonational Categories

As described in Chapter 4, a parameter-based account of intonation must be robust, flexible, and linguistically realistic. The Tilt model is a good example of such an account, but its main strength, its implementability with HMMs, is also its main limitation as a theory of intonation in that there is no clear way for integrating such sophisticated statistical models with phonological frameworks like ToBI. An argument raised in Chapter 4 was that this incongruence might be reason enough for abandoning the frameworks altogether, but the argument is tenable only if our goal is to conduct a purely phonetic analysis of learner intonation, since the Tilt model has no way of either accessing or modeling linguistic information at other (i.e. underlying) levels of representation, especially those relating to semantics, pragmatics, and information structure. In these cases, a model like ToBI really is the best choice, since it lets us capture generalizations about how the pitch contours are used in context with only a few tones and a handful of simple transformations, like downstepping and upstepping. An ideal solution, then, would be to integrate language-specific ToBI models with a common phonetic intonational grammar, which would not only shed light on how languages employ similar pitch excursions differently, but would also establish which ones are most likely to cause learners problems, e.g. if they are present in the target language but not their native language.

As it turns out, this solution is not far away, at least judging by the preliminary (to this particular end) production data gathered in Chapter 3. Under the assumptions of ToBI, English and French have very different intonational phonologies: the former has well over 10 possible pitch accents that may combine in turn with phrase accents and boundary tones to produce surface contours, while the latter only has one default sequence per phrase. What the languages have in common, though, lies in the phonetic implementation of these categories, with native speakers from both groups producing pitch excursions that vary along three main parametric axes, i.e. tonal scaling, tonal alignment, and slope, which is more or less the generalization motivating phonetic analyses like the Tilt model. Slope is extremely variable, but in a basic sense, it is either positive, negative, or approximately zero, since pitch accents can be grouped into rises, falls, and level tones; although we lose a substantial amount of acoustic information by doing so, we can reduce what would otherwise be a continuous parameter to one that is discrete and limited to only three possible values. In the same way, tonal alignment may be continuous and measurable to very small values—duration is typically presented in milliseconds—but it can be discretized in a number of ways, namely either by associating the underlying tone with a syllable, as required by the AM framework, or by categorizing the position of the tone as either being early, in which case it falls near the vowel onset, or late, in which case it does not. By combining these two parameters alone, we can account for the surface forms of many, if not the majority of, the simple pitch movements in English and French, regardless of their idiosyncratic syntactic, semantic, and pragmatic content. Again, it is the main strength of ToBI to be able to model this information in a meaningful way.

What this basic framework is missing, then, is a way of reducing variation in pitch range and tonal scaling to such a level that they could be treated discretely and categorically. In one way, we already have such a tool in z-score normalization, which results in both the literature (Rose 1987; Ladd 2008) and in Chapter 3 of this dissertation has shown to be effective in accounting for interspeaker variation in pitch range within a single language variety. The solution, then, is a relatively simple one, and it is that we can categorize pitch excursions by defining them in terms of the sigma value of their average scaled height, which when combined with the values for slope and alignment suggested above, would give us a basic, three-parameter system for classifying tonal categories according to their phonetic realization. The production data in Chapter 3 support this idea, as native speakers for both languages tended to produce rises at the 1σ and 2σ levels. These were continuations and questions in French, and focused constituents and questions in English.

Based on the framework sketched above, we can imagine establishing cross-linguistic equivalence of tonal categories by thinking of them in terms of the average value of their core phonetic parameters instead of the paradigmatic and syntagmatic relationships that tie them together in their source languages. With respect to the SLM and PAM, what this suggestion means is that we could develop perceptual experiments to test the learners' abilities to discriminate pitch excursions in the L2 regardless of their syntactic, semantic, or pragmatic function. More research is clearly needed to explore whether this is a viable option for categorizing variation in pitch and, by extension, examining learner perception of L2 tonal categories, but on the surface, it appears to be a very good one, and seems much better than relying solely on continuous data to compare the intonational phonetics of different languages.

5.4 Intonational Acquisition and Foreign Language Pedagogy

To the author's knowledge, the pedagogical materials used in foreign language courses at the university level (or below) are not generally designed to improve L2 intonation. They may touch on other aspects of prosody, especially when features are contrastive in one language but not in the other, as is the case with stress in English and French, but for the most part, they do not approach intonation, either as a formal component of the L2 grammar, or as a practical aspect of the learner's pronunciation that may be improved. There are, of course, exceptions to this rule—any course in Mandarin Chinese must include information about tone and, to an extent, intonation—but for the most part, it holds true, and we can hardly hold curriculum designers at fault for its persistence, given the formal and empirical difficulties with studying L2 intonation demonstrated by the chapters above. This section presents a few ideas about how we might move forward in addressing these issues, with the caveat that because so little work has been done in the area, any suggestions we might make can only be speculative, however much they might be supported by the results of this dissertation.

5.4.1 Instructional Goals

Any course designed to improve L2 intonation should have clear instructional goals based on specific milestones in the acquisitional process. The data here suggest that these could be organized around improving the phonetic aspects of learner intonation, like alignment and scaling, that have a subtle but measurable effect on foreign accent, and which seem to be so difficult for learners to acquire under normal pedagogical circumstances (i.e. as a result of classroom instruction). It would also be worth focusing on aspects of intonational phonology that, though easier to acquire, are important not only for developing prosodic proficiency, but also communicative proficiency in general. With regards to L1 English and L2 French, there are three clear areas that an instructional program should target: phrasing, tonal alignment, and tonal scaling. The first is the easiest to address, in that both passive (e.g. multiple choice or sentence labeling) and active (e.g. extemporaneous speaking) tasks could be used to help late beginning and intermediate speakers understand where phrase boundaries should be marked in French. The other two would be trickier, since neither alignment nor scaling are readily measured by ear, meaning that some sort of additional instructional technology would be needed to assess learner progress in meeting the established goals. Nonetheless, a goal for students working on alignment might be to reduce the number of late-aligned rises in their speech, or it might be to bring the average latency of the late-L elbow closer to zero; both methods would be easy to assess, although perhaps less so to teach. Similarly, an instructional goal for scaling could be for learners to raise their overall pitch range to within native values, or to raise the height of specific contours, like the continuation and question rises, which could be monitored by the students during at-home practice sessions and then assessed formally in class by recording samples of their L2 speech. The key to all of these goals is that intonation must be viewed as a measurable linguistic phenomenon rather than a paralinguistic phenomenon influenced by immeasurable variables and subject to unpredictable variation.

5.4.2 Intonation Pedagogy and Computers

The clearest way to develop materials to help learners actively improve their L2 intonation is through the use of computers, since native speaker judgments, which could be also be used to evaluate learner progress, are subjective and thus difficult to consistently replicate. Because pitch tracking algorithms are now widely available in both standalone versions and as part of larger audio software packages, it would be easy to program a simple learning environment that would allow speakers to see pitch contours in real time as they spoke in the L2 (this is in fact already possible with software like Praat, but the programs tend not to be very user-friendly and often require the use of ad-hoc scripts to perform specific functions). The benefits of such a tool would be too numerous to describe here in detail, but in a very simple way, receiving visual feedback from their speech would lighten the load on learners' psychoacoustic systems and allow them to focus on the articulatory processes behind the pitch movements rather than straining, sometimes unsuccessfully, to hear them in the first place.

A simple implementation of this possibility would be to give learners visual feedback on a single phonetic parameter of pitch contour they produced. Using the speakers in this dissertation's corpus as an example, we could imagine showing them a visual representation of the timing of the late-L elbow in their phrase-final rises (either the number in milliseconds or the elbow's position in the syllable) to improve alignment. If enough individual data were gathered, we could also train more variable parameters in much the same way, using simple suggestions like *Go higher!* or *Almost there!*, for instance, to encourage speakers to expand their range and reach levels within L1 norms for the scaling of specific tonal targets. In both cases, as long as we could give the program clear target values for the L1, a goal that may become much more of a possibility if a sigma-based cross-linguistic grammar of pitch excursions could be established, then designing a user-friendly, interactive, and intuitive interface for helping learners reach them would be relatively straightforward. Regular practice like this would not replace classroom instruction, study abroad programs, immersions courses, or tutoring as a primary pedagogical method, but it would add an invaluable and cost-effective tool to the materials and methods that already exist, and it would allow learners to receive valuable, perhaps otherwise unobtainable, feedback on their L2 intonation. This principle has been applied to the instruction of segmental phonology and phonetics for example in the voice recognition components of large-scale software packages like Rosetta Stone, so it makes good sense to extend it to both intonation and prosody in general.

5.5 Summary

This dissertation has addressed a small but important gap in the literature on intonational acquisition. Its findings support existing models of the process, demonstrating that learner varieties of intonation are in many ways indeed different from native varieties, and they also highlight aspects of intonational theory that will either need to be extended or revised for it to predict how, when, and why learners may struggle to acquire particular aspects of another language's system. The main results of the experiments in Chapters 3 and 4 are as follows. Learner varieties of French are not phonetically distinct from either L1 English or L1 French, and their structure seems to be heavily influenced if not determined by phonological and phonetic L1 transfer. Based on production data gathered from students participating in study abroad programs, it seems likely that full acquisition of an L2 intonational system is highly improbable, but steady and measured progress is entirely possible, especially within the domains of intonational phrasing, tonal alignment, and, to a lesser extent, tonal scaling. These results are supported a Tilt-based intonational variety classifier, which showed that learner intonation has a high probability of being classified as belonging to either (but crucially not both) of the L1s. Finally, although further research must be done to determine both the theoretical and the practical viability of the suggestion, the production data also provide preliminary evidence in favor of a cross-linguistic phonetic intonational grammar, whereby continuous parameters like pitch and slope may be discretized and used to establish the acoustic equivalence of tonal categories in separate languages.

REFERENCES

Adank, P., Smits, R., and Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America, 116*(5), 3099-3107.

Adler, M., and Elhadad, M. (2006). An unsupervised morpheme-based HMM for Hebrew morphological disambiguation. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics,* 665-672.

Armstrong, M. (2010). Intonational encoding of pragmatic meaning in Puerto Rican Spanish interrogatives. *Proceedings of Speech Prosody 2010*.

Arvaniti, A., and Garding, G. (2007). Dialectal variation in the rising accents of American English. *Laboratory Phonology*, *9*, 547-576.

Arvaniti, A., Ladd, D. R., and Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics, 26*(1), 3-25.

Arvaniti, A., Ladd, D. R., and Mennen, I. (2006). Phonetic effects of focus and 'tonal crowding' in intonation: Evidence from Greek polar questions. *Speech Communication, 48*(6), 667-696.

Auran, C. (2004). *Prosodie Et Anaphore Dans Le Discours En Anglais Et En Français: Cohésion Et Attribution Référentielle*. Doctoral Dissertation. Université de Provence.

Backman, N. (1977). Learner intonation--A pilot study. *Los Angeles Second Language Research Forum*. Los Angeles, California.

Backman, N. (1979). Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin*, *4*(2), 239-266.

Barkat, M., Ohala, J., and Pellegrino, F. (1999). Prosody as a distinctive feature for the discrimination of Arabic dialects. *Proceedings of Eurospeech, 99*, 395-398.

Baumann, S., and Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*(10), 1636-1657.

Beckman, M. E., and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(01), 255-309.

Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. S. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, *32*(6), 1-29.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 289-304. Springer.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, 167-224.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Ed: Strange, Winifred. York Press.

Best, C. T., Faber, A., and Levitt, A. (1996). Assimilation of non−native vowel contrasts to the American English vowel system. *The Journal of the Acoustical Society of America*, *99*, 2602.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*, 775.

Best, C. T., McRoberts, G. W., and Sithole, N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345-360.

Bigi, B., and Hirst, D. (2012). Speech phonetization alignment and syllabification (SPPAS): A tool for the automatic analysis of speech prosody. *Proceedings of Speech Prosody* 2012, 1-4.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis.* Lawrence Erlbaum and Associates. Mahwah, NJ.

Black, A. W. (1997). Predicting the intonation of discourse segments from examples in dialogue speech. *Computing Prosody*, 117-128. Springer.

Bolinger, D. (1951). Intonation: levels versus configurations. *Journal of the International Linguistic Association*, *7*(3), 199-210.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse.* Stanford University Press.

Bolinger, D., and Bolinger, D. L. M. (1986). *Intonation and its parts: Melody in spoken English*. Stanford University Press.

Borràs-Comes, J., Costa-Faidella, J., Prieto, P., and Escera, C. (2012). Specific neural traces for intonational discourse categories as revealed by human-evoked potentials. *Journal of Cognitive Neuroscience, 24*(4), 843-853.

Borras-Comes, J., Vanrell, M. d. M., and Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association, 44*(01), 1-20.

Brandmeyer, A., Desain, P. W. M., and McQueen, J. M. (2012). Effects of native language on perceptual sensitivity to phonetic cues. *Neuroreport, 23*(11), 653-657.

Braun, B., Kochanski, G., Grabe, E., and Rosner, B. S. (2006). Evidence for attractors in English intonation. *The Journal of the Acoustical Society of America*, *119*, 4006.

Brown, A. M., Gaskill, S. A., Carlyon, R. P., and Williams, D. M. (1993). Acoustic distortion as a measure of frequency selectivity: Relation to psychophysical equivalent rectangular bandwidth. *The Journal of the Acoustical Society of America*, *93*(6), 3291-3297.

Büring, D. (2007). Semantics, intonation and information structure. *The Oxford Handbook of Linguistic Interfaces.* Oxford University Press. Oxford, UK.

Busà, M. G., and Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. *Proc.17th ICPhS*, 380-383.

Campione, E., Flachaire, E., Hirst, D., and Véronis, J. (1997). Stylization and symbolic coding of F0: A quantitative model. *Intonation: Theory, Models and Applications, International Conference.*

Campione, E., Hirst, D., and Véronis, J. (2000). Automatic stylization and symbolic coding of F0: Implementations of the INTSINT model. *Intonation: Analysis, Modeling and Technology, Text, Speech and Language Technology*, *15*, 185-208.

Campione, E., and Véronis, J. (1998). A multilingual prosodic database. *ICSLP*, *98*, 3163-3166.

Caroline Féry, and Frank Kügler. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680-703.

Cenoz, J., and Jessner, U. (2000). *English in Europe: The acquisition of a third language* Multilingual Matters, 30.

Cenoz, J., and Jessner, U. (2001). *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives.* Multilingual Matters, 31.

Chai, W., and Vercoe, B. (2001). Folk music classification using hidden Markov models. *Proceedings of International Conference on Artificial Intelligence*, *6*(6.4)

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning* Cambridge University Press.

Chen, A. (2003). Language dependence in continuation intonation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1069-1072.

Chen, A. (2006). Interface between information structure and intonation in Dutch Wh-questions. *Speech Prosody 2006, International Conference*.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, *59*(2), 367-409.

Chen, A., Gussenhoven, C., and Rietveld, T. (2002). Language-specific uses of the effort code. *Speech Prosody 2002, International Conference.*

Chen, A., Gussenhoven, C., and Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, *47*(4), 311-349.

Chen, A., and Mennen, I. (2008). Encoding interrogativity intonationally in a second language. *Speech Prosody 2008, International Conference.*

Connell, B., and Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology*, (1), 1.

Cruttenden, A. (1982). How long does intonation acquisition take? *Papers and Reports on Child Language Development*, *21*, 112-118. Stanford University Department of Linguistics.

Cruttenden, A. (1997). *Intonation.* Cambridge University Press. Cambridge, UK.

Cummins, J. (2000). Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. In M. R. Childs & R. M. Bostwick (Eds.) *Learning through two languages: Research and practice.* Second Katoh Gakuen International Symposium on Immersion and Bilingual Education, 34-47. Katoh Gakuen, Japan.

Daly, N., and Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics, 5*(1).

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*(1), 491-504.

Delais-Roussarie, E., and Post, B. (2008). Unités prosodiques et grammaire de l'intonation: Vers une nouvelle approche. *Actes Des Journées d'Étude Sur La Parole*, *JEP-TALN*, *8.*

Delattre, P. (1963). Comparing the prosodic features of English, German, Spanish, and French. *Iral-International Review of Applied Linguistics in Language Teaching*, *1*, 193-210.

Delattre, P. (1966). Les dix intonations de base du français. *French Review*, 1-14.

Delattre, P., Olsen, C., and Poenack, E. (1962). A comparative study of declarative intonation in American English and Spanish. *Hispania*, 233-241.

Delattre, P., Poenack, E., and Olsen, C. (1965). Some characteristics of German intonation for the expression of continuation and finality. *Phonetica, 13*(3), 134-161.

Delin, J., and Zacharski, R. (1997). Pragmatic determinants of intonation contours for dialogue systems. *International Journal of Speech Technology, 1*(2), 109.

Derwing, T., Thomson, R. I., Foote, J. A., and Munro, M. J. (2012). A longitudinal study of listening perception in adult learners of English: Implications for teachers. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 68*(3), 247-266.

Di Cristo, A. (1998). Intonation in French. *Intonation Systems: A Survey of Twenty Languages*, 195-218.

Di Cristo, A., and Hirst, D. J. (1986). Modeling French micromelody: Analysis and synthesis. *Phonetica, 43*(1-3), 11-30.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*(2), 249-274.

Díaz-Campos, M. (2006). The effect of style in second language phonology: An analysis of segmental acquisition in study abroad and regular-classroom students. *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 26-39.

Dilley, L. C. (2005). *The Phonetics and Phonology of Tonal Systems.* Doctoral Dissertation. MIT.

Dilley, L. C. (2010). Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica, 67*(1-2), 63-81.

Dilley, L. C., and Brown, M. (2007). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics, 35*(4), 523-551.

D'Imperio, M., and Michelas, A. (2014). Pitch scaling and the internal structuring of the intonation phrase in French. *Phonology, 31*(01), 95-122.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America, 67*(1), 253-261.

Dolores Ramírez Verdugo. The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *Journal of Pragmatics, 37*.

Dusterhoff, K. E., Black, A. W., and Taylor, P. A. (1999). Using decision trees within the Tilt intonation model to predict F0 contours. *Eurospeech 99*, Budapest.

Dusterhoff, K., and Black, A. W. (1997). Generating F0 contours for speech synthesis using the Tilt intonation theory. *Intonation: Theory, Models and Applications,*

Earle, M. A. (1975). *An acoustic phonetic study of Northern Vietnamese tones.* Speech Communications Research Laboratory.

Eckert, H., and Laver, J. (1994). *Menschen und ihre Stimmen.* Beltz Psychologie Verlags Union.

Eddy, S. R. (1996). Hidden Markov models. *Current Opinion in Structural Biology, 6*(3), 361-365.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics, 14*(9), 755-763. Oxford, UK.

Els, T., and Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal, 71*(2), 147-155.

Eriksson, A., Grabe, E., and Traunmüller, H. (2002). Perception of syllable prominence by listeners with and without competence in the tested language. *Speech Prosody 2002, International Conference.*

Estruch, M. (2000). Évaluation de l'algorithme de stylization mélodique Momel et du système de codage symbolique INTSINT avec un corpus de passages en catalan. *Travaux Interdisciplinaires Du Laboratoire Parole Et Langage d'Aix-En-Provence (TIPA), 19*, 45-62.

Fant, G. (1975). Non-uniform vowel normalization. *Speech Transmission Laboratory Quarterly Progress and Status Report, 16*(2-3), 1-19.

Fernandez-Cruz, R. (2000). *L'Analyse Phonologique Et Acoustique Du Portugais Parlé Par Des Communautés Noires De l'Amazonie.* Phd Dissertation. Université de Provence.

Féry, C. (2014). Final compression in French as a phrasal phenomenon1. *Perspectives on Linguistic Structure and Context: Studies in Honor of Knud Lambrecht, 244*, 133.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277.

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*, 692.

Flege, J. E., and Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, *5*(04), 323-347.

Flege, J. E., Munro, M. J., and MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*, 3125.

Flege, J. E., Takagi, N., and Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, *38*(1), 25-55.

Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Fougeron, C. é., and Jun, S. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics, 26*(1), 45-69.

Fox, R. A., Flege, J. E., and Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *The Journal of the Acoustical Society of America, 97*, 2540.

Freed, B. F. (1995). *Second language acquisition in a study abroad context.* John Benjamins Publishing.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent. *Second Language Acquisition in a Study Abroad Context*, *9*, 123-148.

Freed, B. F., Segalowitz, N., and Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(02), 275-301.

Fujisaki, H. (1992). The role of quantitative modeling in the study of intonation. *Proceedings of the International Symposium on Japanese Prosody*, 163-174.

Fujisaki, H., and Hirose, K. (1982). Modeling the dynamic characteristics of voice fundamental frequency with application to analysis and synthesis of intonation. *Proceedings of 13th International Congress of Linguists*, 57-70.

Fujisaki, H., Ljungqvist, M., and Murata, H. (1993). Analysis and modeling of word accent and sentence intonation in Swedish. *Acoustics, Speech, and Signal Processing, 1993. ICASSP*-93., 2, 211-214.

Fujisaki, H., and Ohno, S. (1995). Analysis and modeling of fundamental frequency contours of English utterances. *Eurospeech.*

Gilles, P. (2005). Regionale prosodie im deutschen. *Linguistik–Impulse and Tendenzen.in: Günther,* S.Et.Al (Eds.)*: Linguistic–Impulse and Tendenzen,*

Giordano, R. (2005). Analisi prosodica e transrizione intonativa in INTSINT. Leoni and Giordano (Eds), 231-256.

Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1), 103-138.

Gordon, R. L. (2010). *Neural and behavioral correlates of song prosody.* Florida Atlantic University.

Grabe, E. (1997). Comparative intonational phonology: English and German. *Intonation: Theory, Models and Applications.*

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics, 26*(2), 129-143.

Grabe, E., Kochanski, G., and Coleman, J. (2003). Quantitative modeling of intonational variation. *Proc.of SASRTLM*, 45-57.

Grabe, E., and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology, 7*(515-546)

Grabe, E., Nolan, F., and Farrar, K. J. (1998). IVie-a comparative transcription system for intonational variation in English. *ICSLPi.*

Grabe, E., and Post, B. (2002). Intonational variation in the British isles. *Speech Prosody 2002, International Conference.*

Grabe, E., Post, B., Nolan, F., and Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*(2), 161-185.

Grabe, E., Rosner, B. S., García-Albea, J. E., and Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375-401.

Grosz, B. J., and Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *ICSLPi.*

Grover, C., Jamieson, D. G., and Dobrovolsky, M. B. (1987). Intonation in English, French and German: Perception and production. *Language and Speech, 30*(3), 277-295.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*, 2711.

Gussenhoven, C. (1999). Discreteness and gradience in intonational contrasts. *Language and Speech, 42*(2-3), 283-305.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Speech Prosody 2002, International Conference.*

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge University Press.

Gussenhoven, C., and Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. *Interspeech,* 91-94.

Gussenhoven, C., and Rietveld, T. (1997). Empirical evidence for the contrast between L* and H* in Dutch rising contours. *Intonation: Theory, Models and Applications,*

Gussenhoven, C., and Rietveld, T. (1998). On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics, 26*(4), 371-380.

Gussenhoven, C., and Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech, 43*(2), 183-203.

Gut, U. (2000). *Bilingual acquisition of intonation : A study of children speaking German and English.* Tübingen. M. Niemeyer, 2000.

Hakuta, K., Bialystok, E., and Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science, 14*(1), 31-38.

Hallé, P. A., and Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop /1/clusters. *The Journal of the Acoustical Society of America, 121*(5 PT1), 2899.

Hallé, P. A., Chang, Y., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

He, X., van Heuven, V. J., and Gussenhoven, C. (2012). The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge. *Second Language Research, 28*(3), 283-318.

Hendriks, H. (2005). *The structure of learner varieties.* Mouton de Grutyer: The Hague.

Hermes, D. J., and Rump, H. H. (1994). Perception of prominence in speech intonation induced by rising and falling pitch movements. *The Journal of the Acoustical Society of America, 96*(1), 83-92.

Hirschberg, J. (2004). Pragmatics and intonation. *The Handbook of Pragmatics*, 515-537.

Hirschberg, J., Nakatani, C., and Grosz, B. (1995). Discourse structure in spoken language: Studies on speech corpora.

Hirschberg, J., and Pierrehumbert, J. (1986). The intonational structuring of discourse. *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics,* 136-144.

Hirschberg, J., and Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics,*

Hirst, D. (1988). Tonal units as constituents of prosodic structure. the evidence from English and French intonation. *Autosegmental Studies on Pitch Accent*, (11), 151.

Hirst, D. (2000). ProZed: A multilingual prosody editor for speech synthesis. IEE Seminar on State of the Art in Speech Synthesis (4).

Hirst, D. (2012). The analysis by synthesis of speech melody: From data to models. *Journal of Speech Sciences, 1*(1), 55-83.

Hirst, D. J. (2005). Form and function in the representation of speech prosody. *Speech Communication, 46*(3), 334-347.

Hirst, D., Cho, H., Kim, S., and Yu, H. (2007). Evaluating two versions of the Momel pitch modeling algorithm on a corpus of read speech in Korean. *Interspeech,* 1649-1652.

Hirst, D., Di Cristo, A., and Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. *Prosody: Theory and experiment* (pp. 51-87) Springer.

Hirst, D., and Espesser, R. (1993). Automatic modeling of fundamental frequency using A quadratic spline function.

Hirst, D. J. (2007, August). A Praat plugin for Momel and INTSINT with improved algorithms for modeling and coding intonation. In *Proceedings of the XVIth International Conference of Phonetic Sciences,* 12331236.

Howard, M., Lemée, I., and Regan, V. (2006). The L2 acquisition of a phonological variable: The case of /1/ deletion in French. *Journal of French Language Studies, 16*(1), 1.

Huang, B. H., and Jun, S. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech, 54*(3), 387-414.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57.

Jilka, M. (2000). *The Contribution of Intonation to the Perception of Foreign Accent.* Dissertation.

Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Jun, S. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15*(02), 189-226.

Jun, S. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press: Oxford.

Jun, S., and Fougeron, C. (2000). A phonological model of French intonation. *Intonation,* 209-242. Springer.

Jun, S., and Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14*(1), 147-172.

Jun, S., and Oh, M. (2000). Acquisition of second language intonation. *Proceedings of International Conference on Spoken Language Processing, 4,* 76-79.

Kang, S., Ahn, H., and Hong, M. (2012). The acquisition of L2 English focus by Korean learners. 언어*,* (1), 1.

Kislova, O., Rusalova, M., and Morozov, V. (2011). Characteristics of EEG in subjects with high level of identification of emotions in speech. *Neuroscience Letters, 500*, 40-41.

Klein, W., Plomp, R., and Pols, L. C. (1970). Vowel spectra, vowel spaces and vowel identification. *The Journal of the Acoustical Society of America, 48*, 999-1009.

Knight, R., and Nolan, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association, 36*(01), 21-38.

Kochanski, G. (2009). Using mimicry to learn about phonology. *Oxford University Working Papers in Linguistics, Philology, and Phonetics*, 9.

Krashen, S. D. (1984). Immersion: Why it works and what it has taught us. *Language and Society, 12*(1), 61-64.

Kubozono, H. (1987). The organization of Japanese prosody. *Linguistics and English Language PhD Thesis Collection*. University of Edinburgh.

Kügler, F. (2005). *Swabian and upper saxon intonation patterns*. Unpublished PhD Thesis, Universität Potsdam.

Kügler, F. (2009). A model for the quantification of pitch accent realization1. *Phonology and Phonetics*, 405.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850-11857.

Ladd, D. R. (1994). Constraints on the gradient variability of pitch range, or, pitch level 4 lives!'. *Papers in Laboratory Phonology III: Phonological Structure and Phonological Form,* Ed.P.Keating, 43-63.

Ladd, D. R. (2008). *Intonational phonology* Cambridge University Press.

Ladd, D. R., and Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*(3), 313-342.

Ladd, D. R., Silverman, K. E., Tolkmitt, F., Bergmann, G., and Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *The Journal of the Acoustical Society of America*, *78*, 435.

Laniran, Y. (1990). High tone raising in Yorùbá. *Journal of the Acoustical Society of America*, *87*, S46.

Laniran, Y. O., and Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics, 31*(2), 203-250.

Lee, S. (2014). The realization of French rising intonation by speakers of American English. *Speech Prosody 2014, International Conference*.

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America, 125*, 1138.

Lin, C., and Wang, H. (2005). Language identification using pitch contour information. *Proc. ICASSP*, *(1)*601-604.

Ling, L. E., Grabe, E., and Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*(4), 377-401.

Liu, F., and Xu, Y. (2006). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*(2-4), 70-87.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long−term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96*(4), 2076-2087.

Llanes, À., and Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System, 37*(3), 353-365.

Louw, J., and Barnard, E. (2004). Automatic intonation modeling with INTSINT. *Proceedings of the Pattern Recognition Association of South Africa*, 107-111.

Mannel, C., and Friederici, A. D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Developmental Science, 14*(4), 786-798.

Marandin, J., Beyssade, C., Delais-Roussarie, E., Rialland, A., and de Fornel, M. (2004). *The meaning of final contours in French*. Ms Available at Http://www.Llf.cnrs/Gens/Marandin.

Martin, P. (2012). The autosegmental-metrical prosodic structure: Not fit for French? *Speech Prosody, International Conference*.

Mattingly, I. G. (1966). Synthesis by rule of prosodic features. *Language and Speech, 9*(1), 1-13.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 207-263.

McClelland, J. L., and Siegler, R. (2001). *Mechanisms of cognitive development: Behavioral and neural perspectives.* Psychology Press.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32*(4), 543-563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. *Trends in Linguistics Studies and Monographs, 186*, 53.

Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition, 13*(1), 33-40.

Mennen, I., Chen, A., and Karlsson, F. (2010). Characterizing the internal structure of learner intonation and its development over time. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech.*

Mennen, I., Schaeffler, F., and Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, *36*(02), 303-329.

Mennen, I., Schaeffler, F., and Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. *16th International Congress of Phonetic Sciences*, 1769-1772.

Mennen, I., Schaeffler, F., and Docherty, G. (2008). A methodological study into the linguistic dimensions of pitch range differences between German and English. *Proceedings of the 4th Conference on Speech Prosody*, *Campinas.*

Mennen, I., Schaeffler, F., and Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German). *The Journal of the Acoustical Society of America*, *131*(3), 2249-2260.

Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S., and Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research*, *26*(1), 13-41.

Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language*, *9*(3), 257-288.

Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. *Speech Prosody 2004, International Conference,*

Mertens, P., Beaugendre, F., and d'Alessandro, C. R. (1997). Comparing approaches to pitch contour stylization for speech synthesis. *Progress in speech synthesis*, 347-363. Springer.

Mertens, P., and d'Alessandro, C. (1995). Pitch contour stylization using a tonal perception model. *Proc. 13th International Congress of Phonetic Sciences*, *4,* 228-231.

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.

Miller, J. D. (1989). Auditory−perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America*, *85*(5), 2114-2134.

Miyazaki, K., and Ogawa, Y. (2006). Learning of absolute pitch by children. *Music Perception, 24*(1), 63-78.

Moore, B., and Glasberg, B. (1986). The role of frequency selectivity in the perception of loudness, pitch and time. *Frequency Selectivity in Hearing*, 251-308.

Moore, B. C., and Glasberg, B. R. (1983). Suggested formulae for calculating auditory−filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, *74*, 750.

Moore, B. C., and Moore, B. C. (1982). *An introduction to the psychology of hearing.* Academic Press: London.

Moore, C. B., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America, 102*(3), 1864-1877.

Mora Gallardo, E. (1996). *Caractérization Prosodique De La Variation Dialectale De l'Espagnol Parlé Au Venezuela.* PhD Dissertation. University de Lille III.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *Tesol Quarterly, 25*(3), 481-520.

Myers, E. B., and Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience, 24*(8), 1695-1708.

Najim, Z. (1995). *Prosodie De l'Arabe Standard Parlé Au Maroc: Analyse Historique, Sociolinguistique Et Expérimentale.* PhD Dissertation. Université de Provence.

Nakagawa, S., Ueda, Y., and Seino, T. (1992). Speaker-independent, text-independent language identification by HMM. *ICSLP, 92,* 1011-1014.

Nesterenko, I. (2006). *Analyse Formelle Et Implémentation Phonétique De l'Intonation Du Parler Russe Spontané En Vue d'Une Application à La Synthèse Vocale.* PhD Dissertation. Université de Provence.

Ni, J., and Hirose, K. (2006). Quantitative and structural modeling of voice fundamental frequency contours of speech in Mandarin. *Speech Communication, 48*(8), 989-1008.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences,* 771-774.

O'Brien, I., Segalowitz, N., Freed, B., and Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*(04), 557-581.

Ọdẹ́jọbí, Ọ. A., Beaumont, A. J., and Wong, S. H. S. (2006). Intonation contour realization for standard Yorùbá text-to-speech synthesis: A fuzzy computational approach. *Computer Speech and Language, 20*(4), 563-588.

Ọdẹ́túnjí A. Ọdẹ́jọbí, Beaumont, A. J., and Sylvia, S. H. Intonation contour realization for standard Yorùbá text-to-speech synthesis: A fuzzy computational approach. *Computer Speech and Language, 20*, 563-588.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica, 40*(1), 1-18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*(1), 1-16.

Oxford, R. L. (1994). *Language learning strategies: An update.* ERIC Clearinghouse on Languages and Linguistics, Center for Applied Linguistics.

Paulmann, S., Jessen, S., and Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia, 50*, 1609-1620.

Perdue, C. (1993). *Adult language acquisition: Volume 2, the results: Cross-linguistic perspectives.* Cambridge University Press.

Pierrehumbert, J. (1981). Synthesizing intonation. *The Journal of the Acoustical Society of America, 70*, 985.

Pierre−humbert, J. (1979). Intonation synthesis based on metrical grids. *The Journal of the Acoustical Society of America, 65*, S131.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation.* PhD thesis. Massachusetts Institute of Technology.

Pitrelli, J., Beckman, M., and Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proc. ICSLP, 1* 123-126.

Pols, L. C., Tromp, H. R., and Plomp, R. (2005). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America, 53*(4), 1093-1101.

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.

Post, B. (2002). French tonal structures. *Speech Prosody 2002, International Conference.*

Post, B., and Delais-Roussarie, E. (2011). French ToBI. *Workshop on Romance ToBI, Satellite Workshop PaPI.*

Prieto, P., Shih, C., and Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics, 24*(4), 445-473.

Prom-On, S., Xu, Y., and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America, 125*, 405.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257-286.

Ramus, F., Nespor, M. and Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America, 105*, 512.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265-292.

Rao, K. (2011). Application of prosody models for developing speech systems in Indian languages. *International Journal of Speech Technology, 14*(1), 19-33.

Rao, R. (2011). Intonation in Spanish classroom-style didactic speech. *Journal of Language Teaching and Research, 2*(3), 493-507.

Reddy, V. R., and Rao, K. S. (2013). Two-stage intonation modeling using feedforward neural networks for syllable based text-to-speech synthesis. *Computer Speech & Language*, *27*(5), 1105-1126.

Reynolds, D. A. (1995). Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication, 17*(1), 91-108.

Reynolds, D. A., Quatieri, T. F., and Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing, 10*(1), 19-41.

Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication, 6*(4), 343-352.

Rose, P. (1991). How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency? *Speech Communication, 10*(3), 229-247.

Rose, P.How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency? *Speech Communication, 10*, 229-247.

Rouas, J., Farinas, J., Pellegrino, F., and André-Obrecht, R. (2003). Modeling prosody for language identification on read and spontaneous speech. *Acoustics, Speech, and Signal Processing, 2003. ICASSP. 2003 IEEE International Conference, 6*(1), I-40-3.

Rubin, J. (1975). What the 'good language learner' can teach us. *TESOL Quarterly*, 41-51.

Rump, H., and Hermes, D. J. (1994). Some control experiments on a model for prominence perception. *The Journal of the Acoustical Society of America, 96*(5), 3349-3349.

Rump, H., and Hermes, D. J. (1996). Prominence lent by rising and falling pitch movements: Testing two models. *The Journal of the Acoustical Society of America, 100*(2), 1122-1131.

Schwenk, H. (2007). Continuous space language models. *Computer Speech and Language, 21*(3), 492-518.

Scuffil, M. (1982). *Experiments in comparative intonation: A case-study of English and German* Walter de Gruyter.

Segalowitz, N., and Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(02), 173-199.

Seliger, H. W. (1977). Does practice make perfect?: A study of interaction patterns and l2 competence. *Language Learning, 27*(2), 263-278.

Selkirk, E., and Tateishi, K. (1991). Syntax and downstep in Japanese. *Interdisciplinary approaches to language,* 519-543. Springer.

Silverman, K., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price, P., and Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody. In *Proceedings of the Second International Conference on Spoken Language Processing,* 867-879.

Sharon Wilkinson. (2002). The omnipresent classroom during summer study abroad: American students in conversation with their French hosts. *The Modern Language Journal, 86*(2), 157-173.

Shen, W., and Reynolds, D. A. (2007). Improving phonotactic language recognition with acoustic adaptation. *Interspeech,* 358-361.

Shriberg, E., Ladd, D. R., Terken, J., and Stolcke, A. (1996). Modeling pitch range variation within and across speakers: Predicting F0 targets when 'speaking up'. *Proceedings of the 4th International Conference on Spoken Language Processing,* 1-4.

Simões, A. R. (1996). Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. *Hispania*, 87-95.

Snow, D., and Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*(12), 1025-1058.

Snow, D. (2007). Polysyllabic units in the vocalizations of children from 0 ; 6 to 1 ; 11: Intonation-groups, tones and rhythms. *Journal of Child Language, 34*(4), 765-797.

So, C. K., and Best, C. T. (2008). Do English speakers assimilate Mandarin tones to English prosodic categories? *Interspeech 2008,* 1120.

So, C. K., and Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics PSiCL, 47*, 133.

So, C. K., and Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(02), 195-221.

So, C. K., and Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273-293.

Soltau, H., Saon, G., and Kingsbury, B. (2010). The IBM Attila speech recognition toolkit. *Spoken Language Technology Workshop (SLT), 2010 IEEE,* 97-102.

Sproat, R. W. (1997). *Multilingual text-to-speech synthesis.* Kluwer Academic Publishers.

Steedman, M. (1991). Surface structure, intonation, and 'focus'. *Natural Language and Speech,* 21-38. Springer.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649-689.

Stevens, S. S., Volkmann, J., and Newman, E. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*, 185.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629-654.

Taylor, P. (1994). The rise/fall/connection model of intonation. *Speech Communication, 15*(1), 169-186.

Taylor, P. (1998). The Tilt intonation model. *ICSLP 1998*, paper 0827.

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *The Journal of the Acoustical Society of America, 107*, 1697.

Taylor, P. A. (1993). Automatic recognition of intonation from F0 contours using the rise/fall/ connection model. *Eurospeech 1993*.

Taylor, P. A., and Black, A. W. (1994). Synthesizing conversational intonation from a linguistically rich input. *Second ISCA/IEEE Workshop on Speech Synthesis.*

Taylor, P., Caley, R., Black, A. W., and King, S. (1999). Edinburgh speech tools library. *System Documentation Edition*, *1*, 1994-1999.

Teixeira, J. P., Freitas, D., and Fujisaki, H. (2004). Prediction of accent commands for the Fujisaki intonation model. *Speech Prosody 2004, International Conference*.

Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In *Prosody: Theory and Experiment*, 89-127). Springer Netherlands.

Ulbrich, C. (2002). A comparative study of intonation in three standard varieties of German. *Speech Prosody 2002, International Conference.*

Vaissière, J. (2002). Cross-linguistic prosodic transcription: French vs. English. *Problems and Methods of Experimental Phonetics.in Honour of the 70th Anniversary of Pr.LV Bondarko,*

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech, 38*(3), 253-265.

Vanrell, M. d. M. (2011). The phonological relevance of tonal scaling in the intonational grammar of Catalan. Unpublished Doctoral Dissertation, Universitat Autònoma De Barcelona, Bellaterra, Spain.

Vargas, F. S., and Delais-Roussarie, E. (2011). Exploring L2 learner intonation to develop a phonological analysis of final rises in French. *Phonetics and Phonology in Iberia (PaPI)*.

Vargas, F. S., and Delais-Roussarie, E. (2012). Acquiring phrasing and intonation in French as second language: The case of yes-no questions produced by Mexican Spanish learners. *Speech Prosody 2012.*

Venditti, J., and van Santen, J. (2000). Japanese intonation synthesis using superposition and linear alignment models. *Proceedings ICSLP.*

Verdugo, D. R. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness, 15*(3), 141-159.

Visser, I., and Speekenbrink, M. (2010). depmixS4: An R-package for hidden Markov models. *Journal of Statistical Software*, *36*(7), 1-21.

Vogel, S., Ney, H., and Tillmann, C. (1996). HMM-based word alignment in statistical translation. *Proceedings of the 16th Conference on Computational Linguistics-Volume 2,* 836-841.

Wakita, H. (1977). Normalization of vowels by vocal-tract length and its application to vowel identification. *Acoustics, Speech and Signal Processing, IEEE Transactions on, 25*(2), 183-192.

Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach.* Cambridge University Press.

Ward, G., and Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language,* 747-776.

Warren, P. (2005). Issues in the study of intonation in language varieties. *Language and Speech, 48*(4), 345-358.

Weaver, C. T., and Fry, C. L. (1974). *Intonation and fundamental frequency of infants' and parents' vocalizations during social interaction.* ERIC. Manuscript.

Welby, P. (2004). The structure of French intonational rises: A study of text-to-tune alignment. *Speech Prosody 2004, International Conference.*

Welby, P. S. (2003). *The Slaying of Lady Mondegreen, being a Study of French Tonal Association and Alignment and their Role in Speech Segmentation.* PhD Dissertation. The Ohio State University.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics, 34*(3), 343-371.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics, 15*(4), 399-420.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis.* Oxford University Press.

Wennerstrom, A. K. (1997). *Discourse Intonation and Second Language Acquisition: Three Genre-Based Studies.* PhD Dissertation. University of Washington, Seattle.

Wennerstrom, A. (2011). Rich pitch: The humorous effects of deaccent and L+H* pitch accent. *Pragmatics and Cognition, 19*(2), 310-332.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77*(1), 54-59.

Whalen, D., Levitt, A. G., and Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18*(03), 501-516.

Willems, N., and Collier, R. (1988). A synthesis scheme for British English intonation. *The Journal of the Acoustical Society of America, 84*, 1250.

Wiltshire, C. R., and Harnsberger, J. D. (2006). The influence of Gujarati and Tamil L1s on Indian English: A preliminary study. *World Englishes, 25*(1), 91-104.

Xu, Y., and Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159-197.

Yeung, H. H., Chen, K. H., and Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123-139.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry*, 65-100.

Young, S., Evermann, G., Gales, M., Hein, T., Kershaw, D., Moore, G., Valtchev, V. (2013). *The HTK Book for Version 3.3 (April 2005).*

Zellers, M., and Post, B. (2009). Fundamental frequency and other prosodic cues to topic structure. *Proceedings of IDP 2009.*

Zhou, P., Su, Y., Crain, S., Gao, L., and Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: A view from Mandarin Chinese. *Journal of Child Language, 39*(4), 687-730.

Zimmermann, M. (2011). The grammatical expression of focus in West Chadic: Variation and uniformity in and across languages. *Linguistics, 49*(5), 1163-1213.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America, 33*, 248.

APPENDIX A

PARTICIPANT BIOGRAPHICAL INFORMATION

A.1 Students (English L1/French L2)

ID	Gender	Age	Years French	Other Languages
1	F	19	3	0
2	F	20	3	0
3	F	21	4	2
4	F	21	2	1
5	F	20	4	0
6	F	19	5	0
7	F	19	5	0
8	F	19	4	0
9	F	20	3	0
10	F	21	2	1
11	F	22	5	0
12	F	19	4	0
13	F	19	6	0
14	F	20	4	0
15	F	20	4	0
16	F	19	3	0
17	M	19	3	1
18	M	21	3	0
19	M	21	5	0
20	M	20	2	0

APPENDIX B

DISCOURSE COMPLETION TASK

QUESTIONNAIRE RÉDUIT

1. INTONATION DÉCLARATIVE

1.1 De type neutre

Énoncé d'une unité tonale

(1a1). Imagine que tu adores les mandarines; c'est ton fruit préféré. Quelqu'un te demande : Tu préfères les poires ou les mandarines ? Qu'est-ce que tu réponds ? .

Les mandarines.

Énoncé de plusieurs unités tonales

(1a3). Regarde la photo et, en disant son prénom, dis ce que fait Marie.

Marie mange une banane.

Énumération

(1b1). Dis les jours de la semaine.

Lundi, mardi, mercredi, jeudi, vendredi, samedi, dimanche.

1.2 De type non neutre

Focalization contrastive

(1d1). Tu entres dans un magasin où la vendeuse est un peu sourde. Tu lui dis que tu voudrais quelques oranges, mais elle ne comprend pas bien et elle te demande si ce sont des citrons que tu veux. Dis-lui que non, que ce sont des oranges que tu veux.

Non ! Ce sont des oranges que je veux.

Exclamative
(1d2). Tu es invité(e) à dîner chez une amie. Quand tu arrives tu sens de bonnes odeurs de cuisine. Comment le dis-tu à ton amie ?
Qu'est-ce que ça sent bon !

Déclarative catégorielle
(1d3). Avec la voisine, tu parles d'amis communs qui veulent acheter un appartement mais vous n'êtes pas d'accord sur l'endroit où ils veulent l'acheter. Toi, tu es sûr qu'ils vivront à Limoges. Ta voisine dit que non, qu'ils iront certainement vivre à Bordeaux. Dis-lui, convaincu, qu'ils vivront à Limoges.
Mais non, ils vivront à Limoges !

Déclarative douteuse
(1d4). Un ami t'a demandé d'acheter un cadeau pour quelqu'un que tu connais à peine et tu crains de ne pas avoir fait le bon choix. Dis à ton ami que ce que tu as acheté, ça ne va peut-être pas plaire à cette personne.
Peut-être que ça ne lui plaira pas…

Déclarative d'évidence
(1d5). Tu vas voir la voisine pour lui annoncer que Marie, une amie commune, est enceinte. Ta voisine te demande de qui. Tu trouves ça plutôt bizarre qu'elle ne le sache pas parce que tout le monde sait que c'est de son petit ami Julien ! Qu'est-ce que tu lui réponds ?
Mais de Julien, bien sûr !

2. INTERROGATIVES ABSOLUES

2.1 De type neutre

Énoncés d'une unité tonale
(2a1). Tu entres dans un magasin où tu n'as jamais été avant, et tu demandes s'ils ont des mandarines.
Vous avez des mandarines ?

Disjonction
(2b1). Tu as acheté de la glace à la vanille et à la noisette pour ta fête. Demande aux invités s'ils veulent de la glace à la vanille ou à la noisette.
Vous voulez de la glace à la vanille ou à la noisette ?

Énumération
(2b3). Un ami a besoin de ton aide. Tu peux y aller lundi, mardi, jeudi ou vendredi. Demande-lui s'il veut que tu viennes lundi, mardi, jeudi ou vendredi.
Quel jour est-ce que tu veux que je vienne, lundi, mardi, jeudi ou vendredi ?

<u>2.2 De type non neutre</u>

Questions confirmatives

(2d6). Jean a dit qu'il viendrait dîner. Tu lui demandes de le confirmer.

Tu viendras dîner, non ?

(2d10). Tu sais que l'heure du dîner est déjà passée. Tu vois arriver Marjolaine qui court à la cuisine pour trouver quelque chose à manger, et pour te moquer un peu d'elle, tu lui demandes si elle a faim.

Tu as faim ?

Questions impératives

(2e1). Tu as tes neveux à la maison, et ils font beaucoup de bruit ; alors tu ne peux pas entendre la télévision. Tu leur demandes s'ils ne peuvent pas rester tranquilles.

Vous ne pourriez pas rester tranquilles ?!

(2e5). Demande à tes petits neveux s'ils veulent un bonbon.

Vous voulez un bonbon ?

3. INTERROGATIVES PARTIELLES

<u>3.1 De type neutre</u>

Énoncés d'une unité tonale

(3a2). Il te faut aller à Paris et tu veux acheter un cadeau pour une personne que tu connais à peine et avec laquelle tu veux garder de bonnes relations. Tu demandes à un ami ce qu'il te conseillerait.

Qu'est-ce que tu lui offrirais ?

(3a3). Un ami qui a des appartements en location te raconte qu'il a finalement loué une maison qui était restée vide depuis longtemps. Demande-lui à qui il l'a louée.

À qui tu l'as louée ?

Énoncé de plus d'une unité tonale

(3b1). Ta fille de 14 ans te dit qu'elle veut aller à une fête ce soir. Demande-lui, un peu impérativement, où elle va, comment elle s'y rendra et quand elle va rentrer.

Où est-ce que tu vas, comment tu vas y aller et à quelle heure tu vas rentrer ?

<u>3.2 De type non neutre</u>

Questions impératives

(3e1). Tu demandes à ton frère de te rendre un service, mais tu n'es pas trop sûr qu'il le fera parce que tu lui avais déjà demandé plusieurs fois et il ne t'a jamais aidé. Demande-lui, déjà un peu fâché, quand il t'aidera.

Quand est-ce que tu m'aideras ?!

(3e2). Il t'importe beaucoup que quelques amis viennent dîner chez toi. Tu leur demandes, presque implorant (parce qu'ils t'ont déjà dit qu'ils ne peuvent pas venir) pourquoi ils ne viendront pas.

Pourquoi vous ne venez pas ?

4. INTERROGATIVES ECHO

4.1 De type neutre

Question écho absolue

(4a2). On t'a donné l'heure, mais tu n'as pas bien entendu. Tu penses avoir compris qu'il est une heure. Tu demandes s'ils ont dit qu'il est une heure.

(Vous avez dit qu')il est une heure ?

Question écho partielle

(4b1). On t'a demandé où tu allais, mais tu ne sais pas si tu as bien compris. Demande si c'est bien ça qu'ils t'ont demandé.

(Vous m'avez demandé) Où je vais ?

Disjonction

(4c1). On t'a demandé comment tu es **venu**, mais tu n'as pas compris s'ils ont demandé ça ou plutôt comment tu étais **entré**. Demande s'ils t'ont demandé l'une ou l'autre chose, c'est-à-dire comment tu es **venu** ou comment tu es **entré**.

(Vous m'avez demandé) Comment je suis venu ou comment je suis entré ?

4.2. De type non neutre

Absolue anti-expectative

(4d1). On te raconte que ton ami Jean présente sa candidature à la mairie. Tu ne peux pas le croire, parce que Jean, la politique, ça ne l'intéresse pas du tout, et tu demandes si c'est bien ça.

Qu'est-ce que vous dites, Jean présente sa candidature à la mairie ?

Partielle anti-expectative

(4d2). Ta voisine te raconte qu'elle a dîné dans un restaurant et qu'elle a commandé du lapin aux oignons. Complètement convaincue, elle affirme qu'ils lui ont donné du chat au lieu de lapin. Tu ne peux pas le croire. Demande-lui (très étonné) ce qu'elle dit qu'ils lui ont donné.

Qu'est-ce que tu dis qu'ils t'ont donné ?

5. INTONATION IMPÉRATIVE

Ordre

(5a3). Tu es dans le parc avec ta nièce Marjolaine. Elle court et quitte le parc. Tu as peur parce qu'il y a beaucoup de voitures dans l'avenue qui entoure le parc. Dis-lui de venir.

Viens !

Prière

(5b2). Tu veux aller au cinéma avec un ami. Il te dit qu'il lui faut travailler, mais tu sais qu'il peut le faire plus tard. Comment est-ce que tu ferais pour le convaincre ? Dis-lui de venir.

Allez, viens !

6. VOCATIFS

(6a1). Tu entres dans la maison de ton amie Marjolaine, mais quand tu es dedans, tu ne la vois pas. Tu penses qu'elle doit être dans sa chambre. Appelle-la.

Marjolaine !

(6a2). Tu attends dix secondes, mais personne ne vient. Tu penses qu'elle est peut-être en haut et tu l'appelles avec insistance.

Marjolaine !!

Intonation Questionnaire (Short-Form)

1. INTONATION DÉCLARATIVE

<u>1.1 Neutral</u>

Single tonal unit

(1a1). Imagine that you love mandarin oranges—they're your favorite fruit. Someone asks you, "In general, do you prefer pears, or mandarins?" How would you respond?.

Mandarins.

Multiple tonal units

(1a3). Look at the photo, and say what Marie is doing (using her name).

Marie is eating a banana.

Counting

(1b1). Say the days of the week.

Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday

1.2 Non-neutral

Contrastive focus

(1d1). You just walked into a store where the salesperson is hard of hearing. You say you'd like to buy some oranges, but she doesn't quite understand and asks if you'd like to buy some lemons instead. You say no, and that you wanted to buy oranges.
No! I want to buy oranges.

Exclamative

(1d2). A friend invites you to dinner, and when you arrive, you smell some good smells coming from the kitchen. How would you tell your friend?
That smells good!

Categorical declarative

(1d3). You're chatting with your neighbor about some mutual friends who want to buy an apartment, but you disagree on where you think they should live. You're convinced they should live in Atlanta, but your neighbor thinks they should definitely live in Savannah. Tell her where you think they should move.
No, they should live in Atlanta!

Declarative with doubt

(1d4). A friend asked you to buy a present for someone you hardly know, and you think you've made the wrong choice. Tell him that what you bought might not be right for the other person.
I'm not sure they'll like it…

Déclarative d'évidence

(1d5). You're going to see your neighbor to tell her that Marie, a mutual friend, is pregnant. When you deliver the news, your neighbor asks you who the father is. You think it's weird she doesn't know, because everyone else knows that it's Julian. How would you respond?
Well it's Julian, of course !

2. TOTAL QUESTIONS

2.1 Neutral

Single tonal unit

(2a1). You walk into a grocery store you've never been to and would like to know if they sell mandarins. What do you say?
Do you sell mandarins?

Disjunction

(2b1). You're throwing a party, so you buy some ice cream. When your guests show up, you ask them if they want chocolate or vanilla.
I've got some ice cream—do you want chocolate, or vanilla?

Enumeration

(2b3). One of your friends needs your help studying for a test, but you're only free on Monday, Tuesday, Thursday, and Friday. How would you ask which day is best?

What day should I come: Monday, Tuesday, Thursday, or Friday?

<u>2.2 Non neutral</u>

Confirmatory question

(2d6). John said he was coming to dinner, but you're not sure, so you call him to confirm.

Hi John! You're coming to dinner, right?

(2d10). Dinner time has passed, but you see Mike running to the kitchen to look for some food. You decide to make fun of him a little by asking if he's hungry.

Are you hungry?

Imperative question

(2e1). You have your nephews over to visit, and they're making a ton of noise, so much so that you can't hear the television. You ask them to be quiet.

Can't you two be quiet?!

(2e5). You also ask them if they'd like some candy.

Do y'all want some candy?

3. PARTIAL QUESTIONS

<u>3.1 Neutral</u>

Single tonal unit

(3a2). You need to buy a present for someone you hardly know, so you ask a mutual friend what you think she'd like.

What should I buy him?

(3a3). Your friend has been trying to sublease his room for a long time, and he finally found someone to take it. Ask him who he rented it to.

Who did you rent it to?

Multiple tonal units

(3b1). Your 14-year-old daughter just told you she wants to go to a party tonight. Ask her, a bit strictly, where she's going, who's taking her, and when she'll get back.

Where are you going, how are you getting there, and when will you get back?

<u>3.2 Non-neutral</u>

Imperative questions

(3e1). You ask your brother to do you a favor, but you're not sure he will because you've already asked him a few times, and he never helped you out. Ask him, a little angrily, if he's ever going to help you out.

When are you going to help me out?!

(3e2). It's really important to you that your good friends come over to dinner. You ask them, almost begging, since they already told you "no", why they won't come.
Why aren't you going to come?

4. ECHO QUESTIONS

<u>4.1 Neutral</u>

Absolute

(4a2). You just asked someone for the time, but you didn't quite hear him. You think he said 1:00, but you're not quite sure, so you ask him again.
It's 1:00?

Partial

(4b1). You think your friend just asked you where you went, but you're not quite sure, so you ask if that's what she said, just to make sure.
(You asked me) where did I go?

Disjunction

(4c1). Someone at a party asked you how you got there, but you're not sure if they meant to ask you how you got there, or how you got it in to the house. Ask her which one she meant, just to make sure.
(You asked me) how did I get here, or how did I get in?

<u>4.2 Non-neutral</u>

Absolute unexpected

(4d1). Someone just told you your friend John's running for mayor, but you don't believe him, because the John you know has never been interested in politics. So, you ask again, just to make sure.
What did you say? John is running for mayor?

Partial unexpected

(4d2). You and your friend are talking about what she had for dinner at a local Chinese restaurant. She says she ordered the orange chicken, but she's convinced what they actually served her was cat, which you have a hard time believing. Ask her (very surprisedly) to tell you one more time what she thinks they brought her.
What did you say they served you?

5. IMPERATIVE INTONATION

<u>5.1 Command</u>

(5a3). You're in the park with your nice Mary, and she's just run away from the playground and out into the parking lot. You're worried she might get hit by a car, so you tell her to come back.
Come back!

5.2 Request

(5b2). You're trying to get your friend to go to the movies with you, but he says he's got too much homework to do. You think he'll have plenty of time tomorrow to finish his work, so you tell him to come along anyway.

Come on, let's go!

APPENDIX C

TILT MODEL PROCEDURES AND CODE

Pitch data for all speakers were parameterized in C++ using the dedicated Tilt/RFC programs of the Edinburgh Speech Tools (EST) library (Taylor et al. 1999). The **tilt_analysis()** executable program takes two inputs: a text file containing information about intonational event types and durations, and an F0 pitch contour. Text files were generated using a short script in Praat that extracted event type and duration from the intonational event tier of the TextGrid; the figure below, borrowed from the EST online manual, gives an example of such an input, with the left boundary (i.e. beginning) of each event listed in the first column and the event type in the second. As discussed in Chapter 4, the **c** event type is important because it marks time between pitch accents and boundary tones in a phrase, and the **sil** type is important because it effectively marks phrase boundaries.

```
0.290  146 sil
0.480  146 c
0.620  146 a
0.760  146 c
0.960  146 a
1.480  146 c
1.680  146 a
1.790  146 sil
```

Figure b1. *Input file format for Tilt analysis in EST.*

In studies involving larger corpora than those used here, it may be necessary to implement the automatic event labeler described by Taylor (1999) to segment the sound files, as hand labeling was quite time-consuming. The F0 contour was generated from the raw .wav file using EST's **pda()** pitch detection algorithm, which was applied manually to each of the sound files in the corpus.

The output text contains the information in the input, plus a tilt parameter calculated for the **a** and **b** event types (tilt is unimportant for **c** events and by definition does not exist for those marked **sil**). Sample output, again borrowed from the EST online manual, is included below.

```
intonation_style tilt
#
0.29 26     phrase_start ; ev.F0 115.234 ; time 0.29 ;
0.53 26     a ; int_event 1 ; ev.F0 118.171 ; time 0.53 ; tilt.amp 21.8602 ;
tilt.dur 0.26 ; tilt.tilt -0.163727 ;
0.77 26     a ; int_event 1 ; ev.F0 112.694 ; time 0.77 ; tilt.amp 27.0315 ;
tilt.dur 0.32 ; tilt.tilt -0.446791 ;
1.53 26     a ; int_event 1 ; ev.F0 100.83 ; time 1.53 ; tilt.amp 7.507 ;
tilt.dur 0.22 ; tilt.tilt -0.296317 ;
1.79 26     phrase_end ; ev.F0 92.9785 ; time 1.79 ;
```

Figure b2. *EST output file with intonational event boundaries and parameters*.

The output was (inelegantly) converted to a comma-separated value file using the find and replace feature in Microsoft Word (letters were replaced with empty space, and semicolons with commas). The file was then read into R using the **read.csv()** command and saved as a data frame, where one column was added using the **cbind()** command to indicate alignment relative to vowel onset of each event’s F0 peak.

The HMM language model for the intonational variety classifier was implemented in R using the **depmixS4** (Visser and Speekenbrink 2010) library. The code creates an HMM with ten hidden states, roughly corresponding to the number of discrete pitch accents and boundary tones proposed for French. Silence is not modeled, since, for the most part, it is not linguistically relevant; however, if a larger corpus were used, it might speed up the classification process to add a hidden state the **sil** event type, which that would effectively allow the model to be trained on continuous speech instead of on intonational phrases segmented by hand. The states are connected by standard transition probabilities, while the output distributions are modeled by multivariate Gaussian mixtures. The outputs are necessarily multivariate because the Tilt model specifies more than one acoustic parameter per event (i.e. duration, amplitude, tilt, and alignment all covary).

```
###################################################################
##An HMM-based intonational variety classifier based on Taylor's Tilt model of intonation
##by Scott Lee, The University of Georgia (2014)
###################################################################
rm(list=ls()) #this clears the workspace to make sure no leftover variables are floating
around.
graphics.off(); #close all graphics windows, in case there are still some open from previous
sessions
library(depmixS4); #loads the depmixS4 package for handling the HMMs
##########################################################################

#imports text files containing Tilt parameters as R data frames
#(E)nglish, (F)rench, and (L)earner data are included
Evalues <- read.table('~/Documents/OneDrive/School/Dissertation/Evalues.txt',
            col.names = c('type', 'amp', 'dur', 'tilt'))
Fvalues <- read.table('~/Documents/OneDrive/School/Dissertation/Fvalues.txt',
            col.names=c('type','amp','dur','tilt'))
Lvalues <- read.table('~/Documents/OneDrive/School/Dissertation/Lvalues.txt',
            col.names = c('type', 'amp', 'dur', 'tilt'))

#initializes an HMM for each of the L1 varieties; the learner variety is not included
#because we are only interested in the probability of its being generated
#by the L1 varieties and not in its own unique structure (we could, however,
#do this if we were interested in estimating valus for its parameters)
Emodel <- depmix(list(amp~1,dur~1,tilt~1),data=Evalues,nstates=10,
         family=list(Gaussian(),Gaussian(),Gaussian()))
Fmodel <- depmix(list(amp~1,dur~1,tilt~1),data=Fvalues,nstates=10,
         family=list(Gaussian(),Gaussian(),Gaussian()))
```

```
#fits the HMM to the data in the 'values' data frame
Emodelfit <- fit(Emodel)
Fmodelfit <- fit(Fmodel)

#defines two new HMMs using the L2 values as outputs; this is just so the
forwardbackward()
#function has a depmix object to call when we compute the probability of the
#L2 values with respect to the each of the L1s
L2Emodel <- depmix(list(amp~1,dur~1,tilt~1),data=Lvalues,nstates=10,
            family=list(Gaussian(),Gaussian(),Gaussian()))
L2Fmodel <- depmix(list(amp~1,dur~1,tilt~1),data=Lvalues,nstates=10,
            family=list(Gaussian(),Gaussian(),Gaussian()))

#sets the parameters of the L2 model to those of the L1 models so we can
#get forward probabilities for the L2 response observations
setpars(L2Emodel,values=getpars(Emodel));
setpars(L2Fmodel,values=getpars(Fmodel));

#the final step in the process: using the forward-backward algorithm;
#forward probabilities are contained in the 'alpha' column of the output
Eprob <- forwardbackward(L2Emodel)['alpha'];
Fprob <- forwardbackward(L2Fmodel)['alpha'];

######################################################################
######################################################################
```

Figure b3. *R code for HMM implementation*.