Chapter 21

Part II: Experimental methods and paradigms for prosodic analysis*

Pilar Prieto
ICREA-UPF

Submitted to: *Handbook of Laboratory Phonology*, ed. by Abigail Cohn, Cécile Fougeron, and Marie Huffman.
Publisher: Oxford University Press (Oxford Handbooks).

0. Introduction

There is a long tradition of experimental research in the field of prosody, as different aspects of speech production and perception related to prosody have often been part of traditional laboratory and phonetics investigation. However, in recent years, the development of a set of laboratory tools to investigate language performance and its neurocognitive basis has prompted a new experimental approach to prosody research, in addition to the linguistic approaches that have traditionally been used. This new approach encompasses a wide range of methodological paradigms such as acoustic analysis of speech productions, direct measurement of articulator movements, judgments and reaction times obtained during identification and discrimination tasks, measurements of brain activity and patterns of attention in babies.

In this chapter we provide an overview of some of the experimental methods and paradigms that are currently used for the phonetic and phonological analysis of prosody. Given the large amount of literature on prosody research from the linguistics, speech, and psycholinguistic communities it would be impossible to provide an exhaustive list of relevant research reports. Instead, selected examples of such work and the methodological paradigms used will be provided. Importantly, we will show how these methodological advances have contributed in various ways to our understanding of a large range of issues in the field of speech prosody, as experimental findings have been able to empirically test the various predictions posed by different models of prosody.

* I am grateful to G. Elordieta, S. Frota, M. Grice, C. Gussenhoven, D. Mücke, B. Post, M. Swerts, and to the editors A. Cohn, C. Fougeron, and M. Huffman, for their comments on an earlier version of this manuscript. This research has been funded by projects FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 CSD2007-00012” (awarded by the Ministerio de Ciencia e Innovación) and by project 2009 SGR 701 (awarded by the Generalitat de Catalunya).

1. Acoustic Analysis

Production studies have been widely used for phonetic and phonological analyses of prosody. There is a long tradition of using acoustic analysis of speech productions under various elicitation conditions in the field or in the laboratory. Though most such studies work with laboratory speech and within the strict demands of corpus design and experimental control (**see Part II, chapter 21 on Corpus Design**), other studies are increasingly working with large databases of read and spontaneous speech (**see Part II, chapter 21 on Data Types**).

In the last two decades, the increasing availability of freeware to carry out acoustic analyses, in particular the *Praat* program (Boersma & Weenink 2009), has made it much easier to carry out acoustic analyses even without access to a specialized phonetics laboratory –for detailed information on acoustic analysis, see **Chapter 20**. In most of these production experiments, the analysis of the abovementioned acoustic parameters is manually performed. However, corpus analyses increasingly resort to automatic procedures such as automatic segmentation procedures (for the analysis of duration), automatic detection of F0 turning points, and even automatic prosodic labelling –see **Cole&Hasegawa_chap.19**.

In prosody research, acoustic parameters such as fundamental frequency (F0), duration patterns, and intensity or amplitude patterns have been widely investigated. There is a long tradition of research that has been concerned with the **acoustic characteristics** of diverse prosodic phenomena such as stress prominence or prosodic boundary phenomena (see the classical experiments by Fry 1955, 1958 on word stress in English). Another issue that has been the focus of both production and perception research is that of intonational meaning. One of the crucial issues in intonational phonology is how phonetic elements encode **intonational contrasts**. Since the development of the Autosegmental Metrical approach to intonation (Pierrehumbert 1980, Beckman & Pierrehumbert 1986, and others), **tonal alignment** has been shown to play a central role in encoding intonational contrasts. For a thorough review about work done on tonal alignment and tonal association, see **Arvaniti/D'imperio_chap.9**. Detailed acoustic studies have also served to develop predictive models of phonetic realization. For example, an important goal of intonation research has been to develop **predictive models of tonal alignment** (Silverman & Pierrehumbert 1990 for English; Prieto, van Santen & Hirschberg 1995 for Spanish). Overall, this work has shown that L and H targets are independently aligned relative to the syllable and that the accentual rise is neither of fixed slope nor of fixed duration, that is, the **fixed-rise time hypothesis** cannot be maintained (Arvaniti, Ladd & Mennen 1998, Prieto, van Santen & Hirschberg 1995, among others).

2. Articulatory Analysis

An important line of investigation within the Articulatory Phonology framework has used kinematic data of articulator gestures obtained with the magnetometer, or EMMA (see **Chapter 20**) to study the **intragestural dynamics** of boundary-adjacent lengthening phenomena (Byrd & Saltzman 1998). This work interpretes boundary-adjacent lengthening as a local slowing of the gestures in the immediate vicinity of sufficiently strong prosodic boundaries at multiple levels. Thus just as the syllable edges influence intergestural timing, other types of prosodic boundaries have been shown to influence it too (for a review, see Byrd 1996 and Krakow 1999).

In recent work, researchers have started paying attention to the **coordination between tonal gestures** (measured as F0 turning points) **and oral constriction gestures**. Recent work by D'Imperio et al. (2007), Mücke et al. (2006), Mücke et al. (2009a), Prieto et al. (2007), and Mücke et al. (2009b) has investigated tonal-oral constriction alignment patterns for three different languages (Italian, German, Catalan respectively) by using the magnetometer (EMMA). This work shows that the temporal coordination between pitch movements and articulatory gestures is in many cases stronger than that between acoustic events and F0. Interestingly, there is some variation as to the articulatory landmark which serves as an anchor for the tonal target. For example, in German nuclear LH accents, the H peaks co-occurred with the intervocalic C target, whereas in prenuclear accents peaks co-occurred with the target for the following vowel (accent shift, Mücke et al. 2009). In Catalan rising pitch accents it was the consonantal peak velocity rather than the maximum constriction for the consonant which served as the landmark (Prieto et al. 2007). Such an apparently small alignment difference in the articulatory anchor type may be used by speakers to make phonological distinctions, as in Neapolitan, where H in L*+H (questions) aligns with the maximum constriction, and H in L+H* (statements) with peak velocity (see D'Imperio et al. 2006). Recently, Gao (2008) has formalized the coordination patterns between tonal and vocalic gestures within the Articulatory Phonology framework, proposing that the different alignment patterns can analyzed as differences in phasing between supraglottal and tonal events.

Another research area in which **facial** and **gestural articulatory analysis** has been performed is that of **visual prosody**. In recent years researchers have analyzed quantitatively the head and facial movements that accompany speech and analyzed the visual correlates of prominence and focus, question intonation, as well as the audiovisual expression of affective functions such as uncertainty. For example, Cavé et al. (1996) have analyzed the production of eyebrow movements and their association with tonal rises in French. They found that in only 71% of the cases was there an association, while 38% of the eyebrow movements occurred while the subject was not speaking. Thus, eyebrow movements may serve as back-channel signals and play a role in turn-taking during conversation.

3. Categorization: Identification and Discrimination Tasks

In the past few decades there has been a significant increase in the number of studies that take a psycholinguistic or cognitive approach to prosody research. Common research paradigms consist of behavioral experiments in which subjects are presented with stimuli and asked to make conscious decisions about them. Such experiments can take the form of an identification test, a discrimination test, a gating task, etc.

One method that comes from the study of consonantal contrasts and which has been applied to the study of tonal and intonational contrasts across languages is the well-known **Categorical Perception Paradigm** (henceforth CP paradigm; see Liberman, Harris, Hoffman, & Griffith, 1957; (see **Iverson_chap22** for a thorough review of this paradigm). The CP paradigm involves, first, an **identification/classification task** in which the listeners have to categorize stimuli taken from a continuum, and secondly, a **discrimination task** in which listeners are asked to judge pairs of stimuli as being either the same or different. The solid lines in Figure 1 show the ideal S-shape functions of responses for the identification task, i.e. an abrupt shift from one category to the other. The dashed line shows the ideal function for the discrimination task: if perception is categorical, discrimination between stimuli should be more accurate between categories (a peak of discrimination is obtained at this point) than within them.

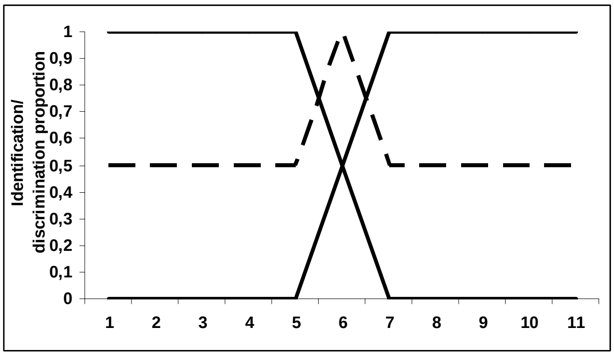

Figure 1. Idealized identification and discrimination functions –solid lines and dashed line respectively.

The CP paradigm has been applied to both **tonal languages** (Francis et al. 2003, Francis & Ciocca 2003) and **intonational languages**, both for *boundary tones* (Remijsen & van Heuven 1999, Post 2000, Schneider & Linfert 2003, Falé & Hub Faria 2006, Prieto, Torres-Tamarit & Vanrell 2008) and for *pitch accents*, in terms of either differences in peak alignment (Kohler 1987, D'Imperio & House 1997, Chen 2003, Grice & Savino 2007, Dilley submitted, Gili-Fivela 2009) or differences in pitch height (Ladd & Morton 1997, Vanrell 2007, Prieto, Torres-Tamarit & Vanrell 2008). Yet the application of this paradigm to intonation research has met with mixed success and there is still a need to test the convergence and degrees of adequacy of this particular experimental method (see Gussenhoven, 2004, 2006, Ladd & Morton 1997, Chen 2003). Although claims have been made of categorical perception for a particular contrast, in the majority of cases no discrimination peaks appear in the crossover of categories revealed by the identification test (Falé & Hub Faria 2006, Remijsen & van Heuven 1999, Prieto, Torres-Tamarit & Vanrell 2008). Only three studies present clear evidence of categorical perception, with a clear discrimination peak in the expected position (Kohler 1987, Vanrell 2007, Schneider & Linftert 2003). Chen (2003) claims that the use of **reaction time (RT) measures** is essential in conjunction with the results of identification tasks to help interpret whether tonal categorical perception effects are linguistically real. Mean RTs are longer for across-category stimuli (that is, ambiguous stimuli) and shorter when the stimulus corresponds to an unambiguous category. The importance of the reaction time data has been confirmed in a number of studies (Vanrell 2007, Falé & Faria 2005, Savino & Grice 2008, Prieto, Torres-Tamarit & Vanrell 2008, among others). Other tasks have also been proven to be successful to discover whether two pitch contours are phonologically distinct, such as the **imitation task** (Pierrehumbert & Steele 1989) and the so-called **semantic task.** Gussenhoven & Rietveld (2000) undertook a perception experiment with the high rise pitch configuration (H* HH%) and the low rise configuration (L* HH%) in Dutch. Listeners were asked to rate "perceived surprise" with a continuum of stimuli with different pitch ranges. Listeners showed a different pattern of responses for both categories, lending support to the hypothesis that the two pitch contours are categorically distinct in Dutch.

3.2. The Gating Paradigm

The **Gating Paradigm** methodology was initially developed to study online word recognition (Grosjean 1980, 1996). The goal of this online speech processing task is to study the speakers' online performance with an **identification task** when only part of the speech signal is available. In this experimental paradigm, target sequences are cut into smaller pieces. These gated stimuli are presented to subjects in a sequential order of increasing duration. Subjects are asked to identify the target unit and rate their level of confidence in their own answers. Two independent measures are important, namely, the isolation point (IP, or the location where correct identification is achieved and maintained over fragments), and the recognition point (RP). The RP will be determined by the rating of the confidence level given by subjects, which will be reached when a stimulus is first rated as 'sure' and this rating is maintained throughout the sentence.

In prosody research, gating tasks have been used to study the **contribution of stress** information in **spoken word recognition** or **word spotting**. For example, Lindfield, Wingfield & Goodglass (1999) tested the role of stress on word recognition by using a word onset gating technique that allowed subjects to hear only the full prosodic pattern of a word (number of syllables and syllabic stress), deprived of segmental information beyond that contained in the onset gate. Their results suggest that word prosody is represented in the mental lexicon and is indeed used by listeners in spoken word recognition.

In the last decade, the gating paradigm has used to investigate the contribution of early intonational features in the **processing of interrogative meaning** (Face 2005 for Spanish, Falé & Faria 2005 for European Portuguese, Vio & Colas 2006 for French, Petrone 2008 for Neapolitan Italian). In these studies, listeners had to identify sentence type when presented with auditory speech stimuli that were gated in specific sentence locations. In general, findings indicate that listeners are very accurate in identifying the sentence type early in the utterance, and that the first posttonic syllable tends to be a good isolation point. These studies demonstrate that early prosodic information provides listeners with enough cues to access and recognize sentence type well before the end of the utterance.

3.3. The Priming Paradigm

Psychologists have developed several ways of probing for association among representations in memory, and one of them is **priming tasks** (see **Schiller_chap22**). This procedure allows the comparison of reactions to a target stimulus when the presentation of the target is primed or is not primed by the immediately prior presentation of another stimulus. For example, when adult speakers of English are asked to perform a **self-paced reading task** (that is, to read aloud a series of words presented one by one on a video screen), they respond more quickly if they hear a semantically related word just before seeing the target word. Psychologists interpret priming results as evidence that representations of semantically related words are associated in memory.

The **cross-modal priming paradigm** allows moment-by-moment activation of word meanings to be accurately tracked during ongoing spoken sentence comprehension as subjects are exposed to visual cues. Cross-modal priming has been employed in research to test the effects of lexical stress on spoken language processing and word recognition. For example, Cooper, Cutler & Wales (2002) conducted a study in which participants were asked to perform a **lexical decision task**. They were presented with a visual target (e.g., a card reading “MUSIC”) after a one-syllable-long fragment. Responses were faster when the prime’s stress matched the target word (e.g., *mu* from *music*) than when it did not (e.g., *mu* from *museum*).

Recent research has demonstrated that prosodic perception is very sensitive to multimodal dimensions, that is, audiovisual integration. One of the most compelling demonstrations of multimodal speech perception is known as the **McGurk effect**

(McGurk & MacDonald 1976). This effect has been extensively replicated across conditions: the experiments involve discrepant audible and visible utterances which are dubbed so as to be synchronously produced. The results show that what hearers perceive is strongly influenced by the visual component, demonstrating that many times the visual component overrides the audio component. A particular area which has used priming techniques is that of **visual prosody**. Swerts & Krahmer (2008) investigated the importance of visual cues to prominence. Their perception experiment probed the relation between auditory and visual cues by means of testing reaction time to congruent and incongruent stimuli, that is, stimuli where auditory and visual cues to prominence were occurring on the same word (congruent) or on different words (incongruent). Their results showed that participants can more easily determine prominence when the visual cue occurs on the same word as the auditory cue, while displaced visual cues hinder prominence perception.

The **facial affect decision task** is one of the latest approaches to gathering new information on how **emotional prosody** is processed for meaning and then integrated with other communicative events. This technique analyzes emotional congruity **priming** effects in perceptual processing. For example, Pell (2005) undertook a study with this task to see how implicit activations of prosodic stimuli representing an emotion were related to specific facial expressions. The subjects had to listen to auditory stimuli over headphones and at the same time judge a face on a computer screen. Auditory stimuli were spoken in different emotional tones and facial stimuli were portraits of an actor conveying either "true" emotional expressions (e.g., happy or sad) or "false" emotional expressions (resembling a "grimace" or facial expression that does not represent basic emotions). The results indicated that the prime-target relationship induced both error and latency responses to facial expressions, revealing that emotional prosody biases online processing of "emotional faces".

3.4. The Eye-Tracking Paradigm

Monitoring **eye movements** while a person is silently reading written text is a procedure that was originally used to infer the processes that underlie skilled reading. There is growing evidence that in silent reading, readers tend to project a default prosodic structure (what Fodor 2002 calls the implicit/silent prosody hypothesis) onto the written words that can then influence syntactic processing, just as 'real' prosody does when we listen to speech. In these experiments, eye movements are tracked by two miniature cameras that record the size of the pupil and the corneal reflection for each eye. By computing the position of the eyes with respect to the position of the markers, one can infer, in real time, the direction and fixations of the participant's eye gaze on a predefined area.

A rapidly expanding community of psycholinguists is now using the **eye-tracking paradigm** to study the role of prosody in spoken language comprehension. Dahan, Tanenhaus & Chambers's (2002) study provides a good example of investigation of the role accentuation and deaccentuation play in the processing of information structure. They used the action-based version of the so-called **visual world paradigm** (Tanenhaus et al. 1995), which contains of a visual display consisting of four black and white line drawings representing four distinct objects, arranged within a grid. Participants had to perform a **word recognition task** and pick up one of the competing objects (for example, *candle* vs. *candy*) while their eye movements were monitored. They were prompted by a series of instruction sequences such as *Put the necklace below the candle*. *Now put the CANDLE above the square*. The target in the second instruction was either accented or unaccented, and referred to either the picture mentioned in the first sentence or to a previously unmentioned picture. The pattern of eye fixations to the target objects demonstrated that deaccented nouns were initially biased toward a given and anaphoric (mentioned) entity, whereas accented nouns were biased toward a nonanaphoric new entity.

More recently, other studies have successfully used the eye-tracking paradigm to study the role of pitch accent type in online processing of information structure. Using the same technique as Dahan, Tanenhaus & Chambers (2002), Chen, Den Os & De Ruiter (2007) found that rise-falls create a strong bias towards newness, whereas rises and also deaccentuation create a strong bias towards givenness. Watson, Gunlogson & Tanenhaus (2008) used the same paradigm to investigate whether pitch accent type in English (namely L+H* vs H*) can bias listeners toward interpreting a temporarily ambiguous noun as referring to a discourse-given or discourse-new entity. Their results show that the interpretive domains of both pitch accents ovelap, that is, L+H* creates a strong bias toward contrast referents, whereas H* is compatible with both new and contrast referents. For a review of the eye-tracking paradigm applied to prosody research, see Watson et al. (2006, 2008).

3.6. The Head-Turn Preference Procedure

The Head-Turn Preference Procedure (HPP) has been shown to be a valuable technique for testing attentional patterns in babies (see classic reference Kemler et al. 1995, and **Maye_chap22** for a review of this paradigm). This procedure records the summary patterns of the babies' eye fixations and does not require very detailed eye-tracking technology. Though the procedure is of limited applicability, since infants have not yet learned the lexicon of their language, it can indirectly tell us significant facts about the acquisition of prosody.

Several types of cues to **word** and **phrase recognition** have been experimentally studied, such as phonotactic cues, stress patterns, prosodic boundary cues, with many studies showing that infants at a very early age are sensitive to these prosodic cues. For example, Jusczyk, Cutler & Redanz (1993) used the HPP procedure to examine the potential role that sensitivity to predominant stress patterns of words might play in lexical development. In English, the majority of words have stressed initial syllables, and the authors demonstrated that by 9 months of age American infants listen significantly longer to words with strong/weak stress patterns than to words with weak/strong stress patterns.

4. Neurophysiological and Neurobehavioral Methods

With the advent of modern neuroimaging techniques such as **event-related potentials** (ERPs) and **functional magnetic resonance imaging** (fMRI), there has been increasing interest in investigating the neural mechanisms involved in the processing of prosody. These techniques have also been brought to bear on many of the questions addressed above.

4.1. Event Related Potentials (ERPs)

ERPs can be reliably measured using electroencephalography (EEG), a procedure which measures electrical activity of the brain and which allows for the non-invasive measuring of brain activity during cognitive processing –see **Schiller_chap22** and **Idsardi&People_chap22** for a thorough review of this technique. One important and robust response used with event-related potentials is the so-called **magnetic mismatch negativity (MMN)**, which is elicited when the auditory perceptual system detects a mismatch between a neural representation of a frequently repeated stimulus (the standard) and a stimulus deviating in at least one parameter (the deviant). This is the so-called **mismatch paradigm**. Since its discovery, this well-known MMN index of automatic acoustic change detection has also been found to be a sensitive indicator of long-term memory traces for native language sounds (e.g., phonemes, syllables) and native lexical words. When comparing MMNs to words and meaningless pseudowords, researchers have detected larger amplitudes for words than for meaningless items. This is interpreted as a neurophysiological signature of word-specific memory circuits/cell assemblies activated in the human brain in a largely automatic and attention-independent fashion. Other studies have found evidence that the MMN reflects automatic syntactic and semantic processing commencing as early as ~100ms after relevant information becomes available in the acoustic input —(see Shtyrov & Pulvermuller 2007 for a review of ERP work related to language functions).

A number of ERP studies have focused on various aspects of prosody processing. For example, some studies have investigated the neural correlates of intonational phrase boundaries, which elicit a specific component in event-related brain potentials, the so-called closure positive shift. Although there is general agreement on the disambiguating role played by intonational phrase (IP) boundaries, the role of lower phrase boundaries seems to be still an open issue, as does the role of word boundaries. Li & Yang (2009), on the basis of EEG measures, investigated whether prosodic boundaries at different levels could evoke the closure positive shift reflecting prosodic boundary perception: they found that listeners were very sensitive to both intonational phrase boundaries and phonological phrase boundaries.

4.2. Brain Imaging Techniques (fMRI)

Functional Magnetic Resonance Imaging (fMRI) is a type of MRI scan that measures the hemodynamic response to neural activity in the brain. Since the early 1990s, fMRI has come to dominate the brain mapping field due to its low invasiveness and lack of radiation exposure. One of the advantages of this technique is that it provides a high spatial resolution, but on the other hand it has poor temporal resolution compared with EEG.

By using the same mismatch paradigm, fMRI studies have been recently applied to the analysis of brain activation localization patterns found during tonal and intonation processing. Gandour et al. (2003) examined neural responses to the discrimination of differences in illocutionary force (questions vs. statements) and emotional valence (happy vs. angry vs. sad) in Chinese utterances in Chinese and English speakers. In both groups of subjects, discrimination of illocutionary force compared to a passive listening baseline led to widespread increased neurological signal in both hemispheres, suggesting bihemispheric processing of intonation. In general, recent studies confirm a right hemisphere dominance during the processing of intonation contrasts (Gandour et al. 2003, Friederici & Alter 2004, Meyer et al. 2004, Fournier et al. to appear).

5. Conclusion

In the last two decades, the field of prosody has witnessed a significant growth in interdisciplinary research that integrates behavioral experimentation with neurophysiological and neuroimaging studies. As we have seen, a wide range of methodological paradigms are now available for prosody research, including acoustic and articulatory analyses of speech productions, judgments and reaction times obtained during identification, discrimination, gating and priming tasks, and measurements of brain activity, eye movements, and infant attention patterns. Many laboratory phonologists have started to use these diverse and complementary methods to address questions in prosody research, overcoming their former reliance of phonological research on a small range of methods. As noted earlier, even though our review is hardly exhaustive, the chapter has attempted to illustrate a selection of methods that are representative of experimental approaches to the study of prosody. We have selected some representative articles for each paradigm in order to provide a glimpse into some of the issues of current interest in the field, such as the categorical or gradient perception of intonation, the target vs. movements approach to intonation, and the use of prosodic information in the recognition and processing of lexical representations, syntax, and discourse. We believe that the full exploitation of these methodological advances will provide important answers and will most likely lead to even more improved experimental paradigms.

References

Arvaniti, Amalia, Ladd, D. Robert, and Mennen, Ineke (1998). 'Stability of tonal alignment: the case of Greek prenuclear accents', *Journal of Phonetics* 26: 3-25.

Arvaniti, Amalia, and Ladd, D. Robert (2009). 'Greek wh-questions and the phonology of intonation', *Phonology* 26: 43-74.

Beckman, Mary, and Pierrehumbert, Janet (1986). 'Intonational Structure in Japanese and English', *Phonology Yearbook* 3: 255-310.

Boersma, Paul, and Weenink, David. (2009). *Praat: doing phonetics by computer* (Version 5.1.01) [Computer program]. Retrieved 26 February 2009, from http://www.praat.org/

Byrd, Dani (1996). 'A phase window framework for articulatory timing'. *Phonology* 13: 139-169.

Byrd, Dani, and Saltzman, E. L. (1998). 'Intragestural dynamics of multiple phrasal boundaries'. *Journal of Phonetics* 26: 173–199.

Cavé, Christian, Guaitella, Isabelle, Bertrand, Roxane, Santi, Serge, Harlay, Françoise, and Espesser, Robert (1996). 'About the relationship between eyebrow movements and F0 variations', *Proceedings of ICSLP 1996*, Philadelphia, 2175-2179.

Chen, Aoju (2003). 'Reaction Time as an Indicator of Discrete Intonational Contrasts in English', *Proceedings of Eurospeech*, 97-100.

Chen, Aoju, Den Os, Els, and De Ruiter, Jan P. (2007). 'Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech', *The Linguistic Review* 24(2): 317-344.

Cooper, Nicole, Cutler, Anne, and Wales, Roger (2002). 'Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners', *Language and Speech* 45 (3), 207-228.

Dahan, Delphine, Tanenhaus, Michael K., and Chambers, Craig G. (2002). 'Accent and reference resolution in spoken language comprehension', *Journal of Memory and Language* 47: 292–314

D'Imperio, Mariapaola (2000). *On defining tonal targets from a perception perspective*. Ph. D. Dissertation, Ohio State University.

D'Imperio, Mariapaola, and House, David (1997). 'Perception of questions and statements in Neapolitan Italian'. *Proceedings of Eurospeech* 1997, vol. 1, 251-254. Rhodes, Greece.

D'Imperio, Mariapaola, Espesser, Robert, Loevenbruck, Helène, Menezes, Caroline, Nguyen, Noël, and Welby, Pauline (2007). 'Are tones aligned with articulatory events? Evidence from Italian and French', in J. Cole & J.I. Hualde (eds.), *Laboratory Phonology 9: Phonology & Phonetics*. Berlin/New York: Mouton de Gruyter, 577-608.

Dilley, Laura C. (submitted). 'The role of F0 alignment in distinguishing categories in American English intonation', *Journal of Phonetics*.

Face, Timothy L. (2005). 'F0 peak height and the perception of sentence type in Castilian Spanish', *Revista Internacional de Lingüística Iberoamericana* 2(6): 49-65.

Falé, Isabel, Faria, Isabel Hub (2005), 'A glimpse of the time course of intonation processing', *Proceedings of the 9th European Conference on Speech Communication and Technology*. Lisboa, 2377-2380.

Falé, Isabel, Faria, Isabel Hub (2006). 'Categorical perception of. intonational contrasts in European Portuguese', in R. Hoffmann & H. Mixdorff (eds.), *Proceedings of Speech Prosody*. Dresden: TUDpress Verlag der Wissenschaften GmbH, 69-72.

Francis, Alexander, and Ciocca, Valter (2003). 'Stimulus presentation order and the perception of lexical tones in Cantonese', *Journal of Acoustical Society of America* 114: 1611--1621.

Francis, Alexander, Ciocca, Valter, and & Kei Chit Ng., Brenda (2003). 'On the (non)categorical perception of lexical tones', *Perception & Psychophysics* 65: 1029-1044.

Fodor, Janet D. (2002). 'Psycholinguistics cannot escape prosody'. *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, France, 83-88.

Fournier, Rachel, Gussenhoven, Carlos, Jensen, Ole, and Hagoort, Peter (to appear). 'Lateralization of tonal and intonational pitch processing: An MEG study'. *Brain Research*.

Friederici, Angela D., and Alter, Kai (2004). ‘Lateralization of auditory language functions: A dynamic dual pathway model’. *Brain and Language* 89: 267-276.

Fry, D. B. (1955). ‘Duration and Intensity as physical correlates of linguistic stress’. *Journal of the Acoustical Society of America* 27(4): 765-768.

Fry, D. B. (1958). ‘Experiments in the perception of stress’, *Language and Speech* 1: 126-152.

Gao, Man (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. Ph.D. dissertation, Yale University.

Gandour, Jack, Wong, Donald, Dzemidzic, Mario, Lowe, Mark, Tong, Yunxia, and Xiaojian, Li (2003). ‘A Cross-Linguistic fMRI Study of Perception of Intonation and Emotion in Chinese’. *Hum. Brain Mapping* 18:149 -157.

Gandour, Jack (2007). ‘Neural substrates underlying the perception of linguistic prosody’, in C. Gussenhoven & T. Riad (eds.), *Tones and Tunes*, *Volume 2: Experimental Studies in Word and Sentence Prosody*. Berlin/New York: Mouton de Gruyter, 3-25.

Gerrits, E. & M.E.H. Schouten (2004). ‘Categorical perception depends on the discrimination task’, *Perception & Psychophysics* 66(3), 363-376.

Gili-Fivela, Barbara (2009). ‘From production to perception and back: An analysis of two pitch accents’, in S. Fuchs, H. Lœvenbruck, D. Pape & P. Perrier (eds.), *Some aspects of speech and the brain*. Germany: Peter Lang GmbH, 363-405.

Grosjean, F. (1980). ‘Spoken word recognition processes and the gating paradigm’, *Perception and Psychophysics* 28 (4), 267-283.

Grosjean, F. (1996). Gating, *Language and Cognitive Processes* 11 (6), 597-604.

Gussenhoven, Carlos, and Rietveld, Toni (2000). ‘The behavior of H* and L* under variations in pitch range in Dutch rising contours’, *Language and Speech* 43: 183-203.

Gussenhoven, Carlos (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos (2006). ‘Experimental Approaches to Establishing Discreteness of Intonational Contrasts’, in Suhoff, S., Lenertovà, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., and Schliesser, J. (eds.), *Methods in Empirical Prosody Research*. Berlin: Mouton de Gruyter, 321-334.

Jusczyk, P.W., Cutler, Anne, and Redanz, N.J. (1993). ‘Preference for the predominant stress patterns of English words’, *Child Development* 64: 675–687.

Kemler Nelson, D.G., Jusczyk, Peter W., Mandel, Denise R., Myers, James, Turk, Alice, and Gerken, Louann (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development* 18, 111-116.

Kohler, Klaus (1987). ‘Categorical pitch perception’, in U. Viks (ed.), Proceedings of the *11th International Congress of Phonetic Sciences,* vol. 5, 331-333). Tallinn.

Kohler, Klaus (2006). ‘Paradigms in Experimental Prosodic Analysis: From measurements to function’, in Sudhoff et al. (eds.), *Methods in Empirical Prosody Research*, 123-152. Berlin/New York: Mouton de Gruyter.

Krakow, R. A. (1999). ‘Physiological organization of syllables: A review’, Journal of Phonetics 27: 23-54.

Kroll, Judith F., Gerfen, Chip, and Dussias, Paola E. (2008). ‘Laboratory Designs and Paradigms: Words, Sounds, and Sentences’, in L. Wei & M. Moyer (eds.), *The Blackwell guide to research methods in bilingualism*. Cambridge, MA: Blackwell Publishers, 108–131.

Ladd, D. Robert (1996). *Intonational phonology*. Cambridge: Cambridge University Press, second edition 2008.

Ladd, D. Robert & Morton, Rachel (1997). ‘The perception of intonational emphasis: continuous or categorical?’, *Journal of Phonetics* 25: 313-342.

Li, W., Yang, Y. (2009). ‘Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences’, *Neuroscience* 158(4): 1416-1425.

Lindfield, Kimberly C., Wingfield, Arthur & Goodglass, Harold (1999). ‘The Role of Prosody in the Mental Lexicon’. *Brain and Language* 68, 1-2: 312-317.

Liberman, Alvin M., Harris, Katherine S., Hoffman, Howard S., & Griffith, Belver C. (1957). ‘The discrimination of speech sounds within and across phoneme boundaries’. *Journal of Experimental Psychology* 54(5): 358-368.

Massaro, Dominic (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. The MIT Press, Cambridge, MA.

Meyer, Martin, Steinhauer, Karsten, Alter, Kai, Friederici, Angela D., & von Cramon, D. Yves (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language* 89: 277-289.

McGurk, H. & MacDonald, J. (1976). ‘Hearing lips and seeing voices’, *Nature* 264(5588): 746–748.

Mücke, Doris, Grice, Martine, Becker, Johannes, Hermes, Anne, and Baumann, Stefan. (2006). ‘Articulatory and Acoustic Correlates of Prenuclear and Nuclear Accents’. *Proceedings Speech Prosody* 2006, Dresden. 297-300.

Mücke, Doris, Grice, Martine, Becker, Johannes, and Hermes, Anne (2009). ‘Sources of variation in tonal alignment: evidence from acoustic and kinematic data’. *Journal of Phonetics* 37 (3): 321–338.

Mücke, Doris, Nam, Hosung, Prieto, Pilar, and Goldstein, Louis (2009). ‘Coupling of Tone and Constriction Gestures in Catalan and German’. Paper presented at Phonetics and Phonology in Iberia, PaPI 09, Las Palmas de Gran Canaria, 17-19 June 2009.

Post, Brechtje (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Pell, Marc D. (2005). 'Nonverbal emotion priming: evidence from the 'facial affect decision task'', *Journal of Nonverbal Behavior*, 29 (1), 45-73.

Pierrehumbert, Janet (1980). *The Phonetics and Phonology of English Intonation*, Ph.D. dissertation, Massachusetts Institute of Technology.

Pierrehumbert, Janet, and Beckman, Mary (1988): Japanese Tone Structure. Cambridge, Massachusetts: MIT Press.

Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.

Pierrehumbert, Janet, Beckman, Mary, and Ladd, D. Robert (2000). 'Conceptual foundations in phonology as a laboratory science', in N. Burton-Roberts, P. Carr, and G. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press: 273–304.

Petrone, Caterina (2008). *From targets to tunes: Nuclear and prenuclear contribution in the identification of intonation contours in Italian*. Ph.D. Dissertation, Laboratoire de Parole et Langage, Université de Provence.

Prieto, Pilar, van Santen, Jan, and Hirschberg, Julia (1995). 'Tonal Alignment Patterns in Spanish', *Journal of Phonetics* 23: 429 451.

Prieto, Pilar, D'Imperio, Mariapaola, and Gili-Fivela, Barbara (2005). Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* 48: 359-396

Prieto, Pilar, Mücke, Doris, Becker, Johannes, and Grice, Martine (2007). 'Coordination Patterns between pitch movements and oral gestures in Catalan', in J. Trouvain & W.J. Barry (eds.), *Proceedings of the XVIth International Congress of Phonetic Science*s, Saarbrücken, Germany: Pirrot GmbH: Dudweiler, 989-992.

Prieto, Pilar, and Torreira, Francisco (2007). 'The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish', *Journal of Phonetics* 35.4: 473-500.

Prieto, Pilar, Torres-Tamarit, Francesc J., and Vanrell, Maria del Mar (2008). 'The role of tonal scaling in distinguishing intonational categories in Catalan'. Paper presented at the *Third TIE Conference on Tone and Intonation*, Lisbon, 15-17 September 2008.

Remijsen, Bert, and van Heuven, Vincent, (1999). 'Categorical pitch dimensions in Dutch: diagnostic test', in J.J. Ohala, Y. Hasagawa, M. Ohala, D. Granville & A.C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences.* San Francisco: University of California, 1865-1868.

Repp, B. H. (1984). 'Categorical perception: Issues, methods, findings', in N. J. Lass (ed.), *Speech and Language, Advances in Basic Research and Practice*. Orlando: Academic Press, Inc, 243-335).

Savino, Michelina, and Grice, Martine (2007). 'The role of pitch range in realising pragmatic contrasts –the case of two question types in Italian'. in J. Trouvain & W.J. Barry (eds.), *Proceedings of the XVIth International Congress of Phonetic Science*s, Saarbrücken, Germany: Pirrot GmbH: Dudweiler, 1037-1040.

Savino, Michelina, and Grice, Martine (2008). Reaction Time in the Perception of Intonational Contrasts in Italian. Paper presented at the *Third TIE Conference on Tone and Intonation*, Lisbon, 15-17 September 2008.

Schneider, Katrin & Linftert, Britta (2003). 'Categorical perception of boundary tones in German', in D. Recasens, M.J. Solé & J. Romero (eds.), *Proceedings of 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions Pty Ltd, 631-634.

Silverman, Kim, & Pierrehumbert, Janet (1990). 'The timing of prenuclear high accents in English', in J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 72–106.

Shtyrov, Y. & Pulvermuller, F. (2007). 'Language in the mismatch negativity design: motivations, benefits and prospects', *Journal of Psychophysiology*, 21(3-4):176-187.

Swerts, Mark, and Krahmer, Emiel (2005). 'Audiovisual prosody and feeling of knowing'. *Journal of Memory and Language*, 53(1), 81–94.

Swerts Mark, and Krahmer, Emiel (2008). Facial expression and prosodic prominence: effects of modality and facial area. *Journal of Phonetics*, 36:2(2008), p. 219-238.

Vanrell, Maria del Mar (2007). 'A tonal scaling contrast in Majorcan Catalan interrogatives', in G. Elordieta & M. Vigário (eds.), *Journal of Portuguese Linguistics* (special issue on Prosody of Iberian Languages) 6.1: 147-178.

Vio, Monique & Colas, Annie (2006). 'Pitch cues for the recognition of yes-no questions in French', *Journal of Psycholinguistic Research* 35.5: 427-445.

Watson, Duane G., Gunlogson, Christine A., & Tanenhaus, Michael K. (2006). 'Online methods for the investigation of prosody', in S. Sudhoff, D. Lenertov, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schlieer (eds.), *Methods in Empirical Prosody Research*. Walter de Gruyter: New York, 259-282.

Watson, Duane G., Tanenhaus, Michael K., & Gunlogson, Christine A. (2008). 'Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*'. *Cognitive Science* 32: 1232-1244.