Daniel Hirst

Speech Prosody: From Acoustics to Interpretation

Prosody, Phonology and Phonetics

Series Editors

Daniel J. Hirst, CNRS Laboratoire Parole et Langage, Aix-en-Provence, France

Hongwei Ding, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China

Qiuwu Ma, School of Foreign Languages, Tongji University, Shanghai, China

The series will publish studies in the general area of Speech Prosody with a particular (but non-exclusive) focus on the importance of phonetics and phonology in this field. The topic of speech prosody is today a far larger area of research than is often realised. The number of papers on the topic presented at large international conferences such as Interspeech and ICPhS is considerable and regularly increasing. The proposed book series would be the natural place to publish extended versions of papers presented at the Speech Prosody Conferences, in particular the papers presented in Special Sessions at the conference. This could potentially involve the publication of 3 or 4 volumes every two years ensuring a stable future for the book series. If such publications are produced fairly rapidly, they will in turn provide a strong incentive for the organisation of other special sessions at future Speech Prosody conferences.

Daniel Hirst

Speech Prosody: From Acoustics to Interpretation

Daniel Hirst
Laboratoire Parole et Langage
Aix-Marseille Universite, CNRS
Aix-en-Provence, France

Paper in this product is recyclable.

To all my Hirsts

Yvette -- (me)

Lenny--Audrey

Clara   Anna

Acknowledgements

Contents

Introduction

The study of speech prosody, today, has become a research area which has attracted interest from researchers in a great number of different related fields including academic linguistics and phonetics, conversation analysis, semantics and pragmatics, sociolinguistics, acoustics, speech synthesis and recognition, cognitive psychology, neuroscience, speech therapy, language teaching… and the list is open.

The field of speech prosody is, in fact, so vast today that it is particularly difficult for any one person to keep up to date on research in all relevant areas. This led to the creation in 2000 of **SProSIG**, a *Special Interest Group on Speech Prosody*, as a working group of the *International Speech Communication Association* (**ISCA**[1]).

In August 2012 the council of the *International Phonetic Association* (**IPA**)[2] voted to support the group, so that SProSIG became the first Special Interest Group to be supported by both ISCA and IPA. This was in keeping with one of the primary motivations behind the founding of the Special Interest Group, which was to provoke interactions between the two principal communities of researchers interested in the field of speech prosody: i.e. engineers and computer scientists on the one hand and linguists and phoneticians on the other hand.

One of the main functions of SProSIG has been to organise a series of *International Speech Prosody Conferences*, which have attracted researchers from all the different areas of research. So far, there have been conferences in France (Aix-en-Provence, 2002), Japan (Nara, 2004), Germany (Dresden, 2006); Brazil (Campinas, 2008), the USA (Chicago, 2010), China (Shanghai, 2012), Ireland (Dublin, 2014), again in the USA (Boston, 2016), in Poland (Poznań, 2018), again in Japan (Tokyo, 2020: because of the covid-19 epidemic this was held as a virtual on-line conference) in Portugal (Lisbon, 2022) and, most recently, in the Netherlands (Leiden, 2024). There is also a

---

[1] http://www.isca-speech.org/iscaweb/

[2] http://www.langsci.ucl.ac.uk/ipa/

Speech Prosody mailing list[3], where you can read and post announcements and discussions on the subject of the scientific study of speech prosody.

In this book, I present a personal overview of Speech Prosody, and in particular the different areas in which I have been especially interested over the last decades. After an introductory chapter on speech prosody and its relation to text and speech (Chapter 1), I look successively at the acoustics of speech prosody (Chapter 2), the transcription of prosody (Chapter 3), the relationship between lexical and non-lexical prosody (Chapter 4), the phonology of speech prosody (Chapter 5), prosodic structure (Chapter 6), the modelling of rhythm and of melody (Chapters 7 and 8), and the central question of the multiple and sometimes quite mysterious ways in which prosody contributes to the interpretation of an utterance (Chapter 9), a question which I also raise more briefly in the first Chapter. After this, in the final chapter (Chapter 10), I briefly outline the directions which I believe may prove the most productive and fruitful for future work in the field of Speech Prosody.

---

[3] Speech Prosody Special Interest Group <sprosig@listserv.utep.edu>

Chapter 1
Speech Prosody, the Missing Link

1.1 What is speech prosody?

As a native speaker of English who has been living and working in France for the last 50 years, I often wonder what I could answer, if asked to summarise what I actually know today about the prosodic differences between the French and English languages, two languages which I obviously know rather well. I could certainly give a talk or write an essay on the subject, but how many of the differences which I might describe could be called established facts?

And, of course, if the question asked concerned two languages with which I am much less familiar, such as Korean and Finnish, for example, my answer would be even more reserved, since these are two languages that I do not speak or understand although they are, in fact, languages in which I have been interested for several years and that I have often heard spoken.

So what, then, can I possibly hope to say about the prosody of the several thousand languages that are said to exist in the world today[1], the vast majority of which I have never even once heard spoken?

This seemingly impossible task is precisely what faces us when we try to answer a question like: *What is speech prosody?*, particularly if the question is not: what can we say about this topic, but: what do we really know?

Having worked in the field of speech prosody for nearly as long as I have lived in France, I find it a sobering thought that our actual objective knowledge of the subject is still rather limited. We probably know quite a lot about segmental phonology today, but despite abundant publications and numerous theories, it seems that we still *know* rather little about speech prosody.

When we talk about speech prosody, we are of course talking about speech. So we can start off by asking what we already know about speech.

---

[1] Ethnologue (Lewis, 2009) [online version: https://www.ethnologue.com 27th edition, February 2024] maintains that there are 7164 known living languages in the world which are in use today.

Speech, or spoken language, is the medium by which we humans usually communicate using sounds. When we think of speech, we probably think of words; we can say that speech is made up of words. This is reflected in the vocabulary of a number of languages: for example the Italian for both *word* and *speech* is "parola" (plural "parole") and the French for *speech* is also "parole", although the word is not pronounced the same as in Italian[2]. So the link between speech and words seems fairly intuitive.

1.2 Speech and Writing

Written language is also made up of words, but of course writing is not the same thing as speech. The comparison of written and spoken language is, in fact, particularly instructive. People who study speech and people who study written language generally come from different backgrounds and belong to different scientific communities. They generally work in different university departments; they go to different conferences and they publish in different journals - so that in fact they actually rarely meet each other[3] .

Somewhat infuriatingly for people working on speech, people working on written language often refer to this as *Natural Language*, to distinguish it from artificial languages, such as programming languages like C, Java, Python etc. In contrast, for speech scientists, the only really *natural* language is spoken. Very many languages in the world have no written form at all, and of course the *invention of writing* is much more recent than the beginning of spoken language. In the same way, for a child, a spoken language is something which does not need to be *learned*. Children *acquire* language in the same way that they acquire other natural abilities like walking. Reading and writing, on the other hand, are *skills* which need to be *learned*, much as other skills such as riding a bicycle or swimming. And of course, those children who do learn to read and write do so much later than they learn to speak.

The fact that the majority of linguists working on what they call *Natural Language* are concerned with text rather than with speech is, no doubt, due to the fact that the majority of them are interested in the automatic analysis of language by computers and that, up until fairly recently, the only language material that could easily be analysed by a computer was text. This, of course, was partly a question of computer memory. One hour of speech, recorded Mono in standard CD quality (16 bits, 44100 Hz) will require over 300 megabytes (MB) of computer memory; which is why a standard 700 MB CD will hold just over an hour of stereo recordings. If the same amount of

---

[2] Italian: /pa'role/ French: /pa'ʁɔl/.

[3] Hopefully, this is less and less true today - as mentioned in the Introduction, this lack of communication between the two communities was one of the principal motivations behind the creation of the Speech Prosody Special Interest Group *SProSIG*

speech material is transcribed as written text, its size will probably be no more than a few thousand kB of memory, which is a gain of a factor of over 1000 compared to the memory space taken up by speech. It is only in recent years that such a huge difference in storage requirements has ceased to be a major problem for people working with language.

Text, then, is much easier than speech for computers to store. It is also, for much the same reason, easier to input and output as well as to manipulate. For humans, though, the opposite seems to be the case. Processing *text* for human beings normally requires using the eyes for input (reading), and the hands for output (writing, typing). Both the hands and the eyes are extremely valuable resources for human beings; and they are required for numerous other tasks in everyday life. We need to use both our hands and eyes for so many tasks that it is virtually impossible, and in some cases highly dangerous, to read and write at the same time as we are doing something else like, for example, driving a car. By contrast, the input and output of spoken language requires only the ears and the mouth, so that using them for speaking or hearing is only a problem when we are trying to do something specific, like eating or listening to music, for example, at the same time.

Not only is spoken language less demanding on our resources than written language, it is also richer. We get more information when we listen to a spoken message than when we read the same words. When given a choice, in fact, most people will prefer oral communication to written communication.

One of the important things that speech tells us, but that text *doesn't* tell us, is the identity of the person who is speaking. Most of us are capable of recognising a familiar voice after only a couple of syllables of speech - "Hello" said over the telephone, for example, is usually enough for us to know who is talking. This is obviously a very important function. But even if we abstract away from the speaker's identity, though, there are still many other things that we get from a spoken message that we don't get from a written one. It is a truism that *the way we say something* is just as important as, and sometimes even more important than, *what we say*. And, to a considerable extent ,we can say that *the way we say something* is due to the prosody of what we say.

Prosody, then, can be seen, in a sense, as the missing link between spoken and written language. Put succinctly, and quasi mathematically:

(1)   Speech = Text + Prosody[4]

---

[4] An interesting everday application of this formula occurs when watching films or tv series in a language one doesn't speak and following the text with subtitles. Here the text is provided by the subtitles and the prosody is provided by the actual speech. Someone I know will happily spend several hours watching tv series in Korean, which they don't speak at all, and following the text with the subtitles. In fact, they far prefer listening to the original Korean with subtitles to watching the same series dubbed in English.

So from this, then, we can derive a first working definition of prosody as:

(2)   Prosody = Speech - Text

Text has the advantage over speech that it is a permanent object that we can examine many times, whereas a spoken message, before the advent of speech recording, was notoriously ephemeral. As the famous Latin proverb (often erroneously attributed to Horace but in fact anonymous) puts it:

(3)   a.   Verba volant, scripta manent.

      b.   spoken words fly away but writing remains

But if speech is essentially text plus prosody, then the study of prosody provides a way of studying precisely what it is that makes speech more valuable than text.

Written language is, decidedly, a rather bad way of communicating the *way* we would say something. Put another way we can say that writing doesn't, in general, represent the prosody of a message, just the words.

There is, though, a partial exception to this and this is the use of punctuation. Punctuation can be seen as a very rudimentary way of representing some of the most basic functions of prosody.

If we take a simple spoken utterance, such as the two syllables of the word "OK", we can make a distinction between the message itself and the way in which we pronounce it. The same message "OK" could be pronounced in hundreds of different ways - for example it could be pronounced *slower* or *faster*, *higher* or *lower*, *louder* or *softer* or any combination of these.

Most of these different pronunciations will have the same meaning, and the same punctuation. For some of them, though, we may want to use a different punctuation and when we do that, we may well consider that the meaning of the utterance is no longer exactly the same. Take the following utterances for example:

(4)   a.   OK.

      b.   OK!

      c.   OK…

In each case, we have the same *word,* represented by the same two letters, "O" and "K". Changing the punctuation from a full stop to an exclamation mark or to the suspension dots, while not radically changing the meaning of the utterance, nonetheless adds a distinctive *flavour* to the way in which the utterance is to be interpreted, expressing, perhaps, greater conviction on the part of the speaker. Notice, though, that if we change this to:

(5)   OK?

replacing the full stop by a question mark changes the interpretation of what we say quite radically. Instead of *affirming* something to be true, as in the first three examples, we are now *asking* the listener whether the *statement* "OK" is true or not.

Punctuation, then, provides a fairly limited way of communicating the way in which we want an utterance to be interpreted. It is significant that in the history of writing, punctuation actually developed rather later than the other symbols which we use; and this is true not only for our Western alphabetic writing system but also for all the major writing systems of the world. Writing has always been first and foremost a way of communicating the *words* of an utterance. It is only in much more recent times that punctuation has been used: first to separate parts of an utterance into chunks, showing readers essentially where they should insert a pause when reading, and then, later, to add more subtle relationships between the chunks.

The punctuation symbols we use today in European languages were not standardised until after the introduction of printing. Today, similar symbols, with some variants, are used in printed texts in the majority of alphabetic writing systems in the world, as well as in most non-alphabetic writing systems such as Chinese, Japanese and Korean, although with the notable exception of Thai, which uses practically no punctuation (including no spaces between words) except for [,], separating items in lists.

Here is an example of a sentence in English:

(6)   Words, phrases and clauses are often separated by punctuation; but this is not always the same in every language: some languages, for example, don't use semi-colons.

and its translation into Russian:

(7)   Слова, фразы и предложения часто разделяются знаками препинания; Но это не всегда одно и то же на всех языках: например, в некоторых языках точка с запятой не используется.

Greek:

(8)   Οι λέξεις, οι φράσεις και οι προτάσεις συχνά χωρίζονται με σημεία στίξης. αλλά αυτό δεν είναι πάντα το ίδιο σε κάθε γλώσσα: ορισμένες γλώσσες, για παράδειγμα, δεν χρησιμοποιούν ερωτηματικά.

Arabic:

غالبًا ما يتم فصل الكلمات والعبارات والجمل بعلامات الترقيم ؛ لكن هذا ليس هو نفسه دائمًا في كل لغة: بعض اللغات ، على سبيل المثال ، لا تستخدم الفاصلة المنقوطة.

(9)

Hindi:

(10)   शब्द, वाक्यांश और खंड अक्सर विराम चिह्न द्वारा अलग किए जाते हैं; लेकिन यह हर भाषा में हमेशा समान नहीं होता है: कुछ भाषाएं, उदाहरण के लिए, सेमी–कोलन का उपयोग नहीं करती हैं।

Chinese:

(11)   单词，短语和从句通常用标点符号分隔开，但并不是每个语言都有相同的情况：有些语言，例如，不使用分号。

Japanese:

(12)   単語、句、節は、通常、句読点で区切られています。しかし、これは全ての言語でいつも同じというわけではありません。一部の言語では、たとえば、セミコロンは使いません。

Korean:

(13)   일반적으로 단어와 문장 또는 절과 같은 단위들의 구분에는 구두점이 사용되지만, 그 내용과 용례는 언어마다 다르게 나타날 수 있다. 그 예로, 일부 언어들의 경우에는 세미콜론; 이 전혀 사용되지 않는다는 사실을 들 수 있겠다.

and Thai:

(14)   คำ, วลี, และประโยคมักจะแยกโดยวรรคตอน แต่นี่ก็ไม่เสมอเหมือนกันในทุกภาษา เช่น บางภาษาไม่ใช้จุดคู่กึ่ง

It seems probable that it was the development of the use of written language in more communicative situations which led to the need to provide some minimal prosodic annotation. The earliest written messages were essentially either records of financial transactions or monumental commemorations of dead heroes or historical events. For messages like this, the need for punctuation was obviously correspondingly limited.

Once writing began to be used for more interactive communication, such as personal letter-writing for example, then the communicative pressure led to the progressive introduction of such devices as the question mark and the exclamation mark and then, later, to the marking of emphasis through the use of underlining or, in printed texts, by using italic or bold type. At

the same time, the increased complexity of syntactic structures transcribed in writing led to the development of a hierarchy of text separators, ranging from the minor separation implied by a comma, to the major separation implied by a full stop with intermediate values expressed with the semi-colon and colon.

Recent developments in communication technology such as internet chat and mobile SMS messages have, as was to be expected, added further pressure for devices for prosodic annotation. The first, and most well-known of these, written :-) as a sideways representation of a "smiling face" or "smiley", together with :-(to express the opposite, was originally posted to the Carnegie Mellon University computer science board by Scott Fahlman in a 1982 email[5]

```
19-Sep-82 11:44    Scott E  Fahlman            :-)

I propose that the following character sequence for joke markers:

:-)

Read it sideways.  Actually, it is probably more economical to mark
things that are NOT jokes, given current trends.  For this, use

 :-(
```

This rapidly became popular although the interpretation seems to have switched rapidly from "I'm joking/not joking" to "I'm pleased/displeased". The two markers, more often simplified to :) and :(were quickly supplemented by other symbols such as ;) for a wink, :x for a kiss, :O for a shocked face and :/ for a bored or puzzled face. These and other symbols have since become quite a fad and have been baptised *emoticons* or *emoji*, defined as "the textual portrayal of a writer's face or mood". On some web forums or instant messengers, the text symbols are automatically replaced by small graphical images as in Figure 1.1.

1.3 Prosody at work

We saw, then, that prosody can modify the way in which we interpret an utterance. Sometimes the modification can be rather minor, as in the 'unfinished' interpretation of OK…. In other cases, the modification can be more substantial. Suppose, for example, that someone pronounces the following sentence with a bored tone of voice:

---

[5] http://en.wikipedia.org/wiki/Emoticon

Fig. 1.1 A list of graphical emotiwcons as provided by Yahoo Instant Messenger

(15)   This book is really exciting.

The *prosodic information* (boredom) is, in this case, directly opposite to the information of the sentence which a listener can obtain by combining the information from the lexical items "this", "book", "is" "really" and "exciting" with the information obtained from the syntactic analysis of the sentence, that tells the listener how the words fit together.

The interpretation obtained by combining lexical information with syntactic information is something we can call the *literal* interpretation of an utterance. We could also call it the *semantic* interpretation, although that term, being more technical, generally implies a specific theory of semantics, which is something I do not wish to go into here.

Here, the literal interpretation of the sentence provides a meaning that allows the listener to understand that the speaker has a positive reaction to the book he is looking at, that he finds it exciting. And yet the prosody of the utterance tells us exactly the opposite. It is almost as if the speaker had used different words; instead of "exciting", he might have said "boring" and in that case he would have expressed *literally* what in our example is being expressed prosodically.

Of course, what is interesting about example (15) is that the speaker doesn't actually *say* that he thinks the book is boring; he just *implies* that by his "tone of voice". This is an opportunity which speakers readily seize, to criticise without having to take the responsibility for their criticism.

A teacher who recommends one of his students by saying:

(16)   He's very hardworking...

can, if challenged, deny having said anything critical at all about the student.

The prosody of an utterance, then, can modify its literal interpretation. In examples (15) and (16), it does so by adding a global overtone to the message which was contradictory to the information derived from the lexical items "exciting" and "hardworking".

Prosody can also affect the literal interpretation of an utterance by contributing to the syntactic interpretation , which as I suggested above, is, after the lexical information, the second major component of a literal interpretation. Example (17) can be interpreted in two different ways:

(17)   She asked the man who lived there.

The word "who" can be interpreted as a *relative pronoun*, so that (17) is interpreted as equivalent to:

(18)   The man she asked was the one who lived there.

"Who" could, alternatively, be interpreted as an *interrogative pronoun*. In that case the sentence would be interpreted as equivalent to:

(19)   She asked the man to tell her who was the person who lived there.

The syntactic interpretation of the utterance is determined by the interpretation of the pronoun "who" and it seems that prosody, here too, contributes to that interpretation. To simplify, we can say if the word "who" is prominent, or *accented*, then it is likely to be interpreted as interrogative, while an unaccented "who" will tend to be interpreted as a relative pronoun. This is, or course, not specific to the word "who", but is part of a larger pattern of the way in which words in English are accented.

Ambiguities like example (17), where the prosody contributes to the syntactic interpretation of the utterance, are actually very common in English and presumably in all languages. It is often the case that punctuation will resolve the ambiguity as in the case of the following often quoted example:

(20)   Woman without her man is nothing.

which an English professor reputedly once wrote on a blackboard asking his students to punctuate it[6]. The male students are said to have written:

---

[6] The original author seems to have been Herb Wheaton (1912-1952) publisher of Hokah Chief, quoted in "Without Stuttering", who wrote:
A comma is a small mark, yet its importance is very great. Read the following sentence: "Woman, without her, man would be but a savage beast."
Remove the commas and read it again.
cf. http://msgboard.snopes.com/message/ultimatebb.php?/ubb/get_topic/f/95/t /000576/p/1.html 2003 September 12. die daagliks phosdex

(21)   Woman, without her man, is nothing.

while the female students wrote:

(22)   Woman: without her, man is nothing.

Here, the different punctuation reveals a different syntactic structure for the two interpretations of the sentence, interpretations which would also correspond to different ways of saying the sentence.

There are, however, cases where the literal interpretation is more complex and where a difference of interpretation corresponding to a difference in the prosody, is not necessarily reflected in the punctuation.

In the following examples:

(23)   a.   Mr Jones wasn't in his office so Mary phoned him.

      b.   Mr Jones wasn't in his office so Mary phoned the president.

The default interpretation would seem to be that in (23a.), Mary phoned Mr Jones whereas in (23b.) she didn't phone Mr Jones but she phoned the president. This interpretation depends, though, on the accentuation of the two sentences. In the default version, the pronoun "him" is not accented but the noun "president" is. In fact, if we accent the pronoun "him", the sentence is likely to be interpreted as referring to a different person, probably indicated by a pointing gesture. Similarly, if the noun "president" is not accented, it can be interpreted as referring to Mr Jones, who is the president.

The following (24) is an authentic example which I heard on the French radio news, back in 1977[7]. My discussion of this example refers to the French original but the example seems to work in essentially the same way in English.

(24)   a.   Il semble que les policiers sont à deux doigts d'arrêter Spaggiari, mais il faudra qu'ils fassent vite pour trouver la cachette de l'ancien parachutiste.

      b.   *It seems that the police are on the point of arresting Spaggiari, but they'll have to act quickly to find the hiding place of the former paratrooper.*

For the written version of the utterance, there are two possible interpretations. In one of these, the police are looking for two people: one named

---

[7] *France Inter*, Informations 13-14, 1977 March 12. I remember distinctly that I was driving my car when I heard this example on the radio and stopped the car to note down the sentence as well as the time and date.

Spaggiari and another who is a former paratrooper. In the second interpretation, Spaggiari and the former paratrooper are the same person. In fact, in the spoken version, I had no hesitation in choosing the second interpretation, even though at that time I did not know that Spaggiari, a famous bankrobber, had been a paratrooper. This means that the information that Spaggiari and the former paratrooper were the same person, in other words that the second expression was being used as an anaphoric reference to the first, was somehow conveyed by the prosody of the utterance[8]. This example illustrates to some extent the complexity of the way in which prosody can contribute to the interpretation of utterances, a subject to which we will return at much greater length in Chapter 9.

1.4 How many words?

We use prosody to interpret utterances at every moment when we are listening to speech and we are, of course, quite unconscious of doing so. It is, in fact, quite rare that we become aware of the ambiguity of many of the utterances that we hear every day.

The following example, demonstrates not only how common utterances are often ambiguous, but more so, that they can be ambiguous in a surprisingly massive way. And to show furthermore that this ambiguity is not simply a question of some vague *attitude* or *tone of voice*, the example I give is from the realm of mathematics, more precisely from that of ordinary numbers. The example I give is in French, since the French language happens to lend itself particularly well to the demonstration; but the relevance of the example to numbers in *any* language will be apparent.

In French, the following nine words each correspond to a number which is given beneath the word as a gloss.

(25)   *Sept, cent, vingt, cinq, mille, six, cent, trente, neuf*
      7   100 20   5   1000 6   100 30   9

but these same nine words could also be interpreted as representing a list of just three numbers instead of nine:

(26)   *sept cent vingt, cinq mille six, cent trente-neuf*
    720   5006   139

or the whole sequence of nine words could represent just one single number:

---

[8] Specifically, this interpretation was favoured by a major pitch accent on the word "vite" (*quickly*) followed by a de-accenting of the whole final phrase "pour trouver la cachette de l'ancien parachutiste". (*to find the hiding place of the former paratrooper.*)

(27) *sept   cent vingt cinq mille six-cent trente-neuf*
      725639

We might wonder just how many other ways we could interpret the sequence of numbers of example (25). The answer is fairly astonishing - there are precisely another 253 other ways in which we could interpret the same sequence of words. There are, in fact, exactly 2^{n-1}. ways to interpret a sequence of n numbers, so for 9 numbers, there are 2^8 different ways, a total of 256 in all.

1.5 Summary and Perspectives.

In this chapter we have had a glimpse at some of the different ways in which prosody contributes to meaning. I shall often use the word prosodic *functions* in the rest of this book to refer to the way in which prosody contributes to the overall interpretation of an utterance, by contrast with prosodic *forms* which I shall be looking at in more detail from Chapter 2 onwards.

Most of the functions of prosody in speech are nearly universal. Prosody seems to be used in every language in the world to do some, and often all, of the following things:

- to express different speech acts, e.g. statement, question etc: example (4) **OK**
- to express speaker states, i.e. the speaker's attitudes and emotions: examples (15) **exciting** and (16) **hardworking**
- to make some words more prominent than others: example (17) **who**
- to help identify different syntactic structures: example (20) **woman**
- to express different anaphoric relations between phrases: examples (23) **president** and (24) **Spaggiari**
- to separate one block of words from another: example (25) **numbers**

Besides these, there seem to be a virtually unlimited number of other functions, including several concerned with the organisation of discourse and dialogue.

But there is one more function which I believe deserves our particular attention. In the vast majority of languages of the world, a change in the way that we pronounce some sounds, a change in what I have called *prosodic form*, like in the examples that we have been looking at in this chapter, can be used to actually distinguish words in the language. This lexical prosodic function is not strictly universal, since I will argue that there are languages, and, I claim, French is one, which possess *no* lexical prosody at all. The prosody of words will be dealt with in detail in Chapter 4 and I believe it holds a key to our understanding and analysis of prosodic form. But before we investigate lexical prosody, since there is no standard way of *writing*

prosody, we need first to leave the world of functions, meaning and interpretation for a closer look, in the following chapters, at prosodic forms, looking first in Chapter 2 at the physical manifestations of speech prosody and then in Chapter 3 at the ways in which prosodic forms have been transcribed in the literature.

Chapter 2
Looking at Sounds

2.1 Sounds

Speech is conveyed by sound. The sounds of speech, like all sounds, are physical events that can be recorded and analysed. The scientific analysis of sounds is called acoustics.

Acoustics was, for a long time, a subject that required expensive and complicated equipment. For this reason, it was practised and taught only in specialised universities and technological research centres. Today, with the development of micro-computers, it is possible for anyone with access to a computer and to an internet connection to obtain sophisticated software capable of performing analyses which, just twenty years ago, would only have been possible in a small number of specialised laboratories throughout the world.

In this chapter, we shall be looking at speech sounds, in particular at the prosody of speech sounds, analysed by means of computer software. And when I say "looking" at speech sounds, I mean that literally; while sounds are produced to be heard, modern technology allows us to examine them in a more static way and to actually see what a sound looks like, once it has been converted into a computer file.

All of the analyses described in this chapter can be performed on a home computer (desktop or laptop) using software which can be downloaded free of charge via internet. The analysis will, of course, involve the visual display of sounds but since what you are studying is actually sound, it is very important for you to be able to listen to the different examples at the same time as you look at the visual displays. You are encouraged to perform the analyses yourself and to listen carefully to the examples given, as well as carrying out similar analyses of your own recordings, as described below.

2.2 Using software for acoustic analysis

A number of good software programmes for acoustic analysis are available today. All the examples analysed in this chapter, however, will assume that you are using the *Praat* software package Boersma & Weenink (1992/2024), which was developed by Paul Boersma and David Weenink from the Institute of Phonetics of the University of Amsterdam.

Although this is, of course, not the only program which is available for the acoustic analysis of speech, it does have a number of advantages over most of its competitors. In particular, Praat is:

- *comprehensive* If anything, Praat could be accused of doing too much! See below for an inventory of some of Praat's functions
- *programmable* The software includes its own powerful interpreted scripting language
- *multiplatform* Praat is available on all the major (and many minor) computing platforms. This is particularly useful for teamwork since it makes it possible to share both data, annotations and scripts.
- *open source* Praat is released under the GPL licence (GNU General Public Licence)
- *actively updated* Since the last major update [6.4] released in November 2022, until the current version (July 2024) there have been 14 minor revisions. There are on average about two updates every month.
- *actively supported* There is a Praat-users group and mailing list[1] with currently over 2900 subscribers. Paul Boersma is particularly reactive, answering users' questions and introducing new features. The group has full archives (over 10 300 messages) dating back over the last twenty years.
- *well documented* Besides the extensive documentation within Praat, there are a number of third-party references providing additional documentation for Praat.[2].
- *free of charge* No comment!

Besides a full set of acoustic analysis tools, the software also includes tools for:

- listening experiments
- speech synthesis (PSOLA resynthesis, formant synthesis and articulatory synthesis)
- statistics (multidimensional scaling, principal component analysis, discriminant analysis)
- high quality graphics
- feedforward neural networks

---

[1] https://groups.io/g/Praat-Users-List
[2] See for example Styler (2011/2022), Beňuš (2021). and (in Spanish) Correa Duarte (2014)

• discrete and stochastic Optimality Theory

If, despite all this, you choose to use other software, it should be relatively straightforward to adapt the explanations given in the rest of this chapter.

2.2.1 Downloading and installing Praat.

The Praat program can be downloaded from the web page:
http://www.praat.org.
Since the program is very frequently updated, it is advisable to check periodically that the version you have is (reasonably) up to date. This will be particularly important if you decide to make use of the possibility of using Praat *scripts*, especially if you use scripts written by somebody else, since the script may well rely on some recently introduced feature of the software.
Under the heading *Download Praat* you will find the following options:

• Macintosh
• Windows
• Linux, FreeBSD
• SGI, Solaris, HPUX
• the source code

which is a refreshing change from all those sites which assume that everybody in the world is using Windows!
Download the appropriate version of the software for your operating system and store it on your computer as you would with any other program. Note that any options which you define when you are using Praat will be remembered by your computer, so that when you install a more recent version of the program, you will find that all of your familiar options are still available.
In the rest of this chapter I give a brief introduction to using the Praat program to analyse sounds. Praat is, as we saw above, in fact an extremely complex set of programs. This is sometimes a little daunting for new users. What I try to do in this presentation, is to introduce you gradually to various functions of Praat related to the study of prosody. Don't expect to understand everything about the program immediately. Take time to assimilate the different functions - experiment by yourself. Once you have finished working through this chapter, you should be in a position to go on and use the integrated Help manual to find out more about the program (there are currently more than 1000 manual pages!!).
Obviously, the first thing you need to do with Praat is to provide it with a speech recording that you can take a look at. You can start with the record-