

**Practice material for English Phonetics 1:  
Basic prosody and connected speech**

**Štefan Beňuš**

## Contents

1.	Introduction.....	3
1.1.	Motivation.....	4
1.2.	Material.....	5
1.3.	Notes on the approach taken in the book.....	7
2.	Crash course to basic suprasegmental aspects of fluent speech .....	11
2.1.	Weak forms.....	11
2.2.	Connected speech .....	13
2.3.	Prosody: Prominence and chunking .....	16
3.	Commentaries to individual excerpts .....	20
3.1.	Weak forms.....	20
3.1.1.	s3_m2_120.6: what you do is you go down towards them and when you get...	20
3.1.2.	s4_m2_60.8: just ignore the atomic bomb your road basically goes slightly to the left and then straight up.....	23
3.2.	Connected speech aspects.....	26
3.2.1.	s4_m2_376.2: it crosses over it and goes .....	26
3.2.2.	s3_m2_117.5: okay do you want me to go under A-gray e- grade eggs.....	30
3.3.	Exercises with sample answers.....	37
3.3.1.	s3_m2_73.8: and then it goes in between the two speech pens .....	38
3.3.2.	s4_m2_200.2: they're quite close together so where you were in the middle of the fantasy god you just go straight down .....	43
4.	Conclusion .....	51

## 1. Introduction

The goal of the study material presented in this book is to offer a guide to exploring the patterns and habits English native speakers use when speaking. The book is divided into four sections. In this first introductory section, I present the motivation behind the book in greater detail in subsection 1.1. The description of the corpus and material selected for sample analyses together with the list of files provided for readers and analyzed in this book are included in subsection 1.2. The last subsection 1.3 presents a description of the approach to the analysis of spontaneous speech taken in this book. It is instrumental for framing the material in the subsequent sections and provides a broader context for the readers' engagement and hands-on exploration in these sections.

Section 2 presents a very brief crash course to the basic suprasegmental aspects of fluent speech. It assumes that the readers have a good grasp of the theoretical issues related to the aspects of prosody and connected speech in English from the lectures, textbooks, and other material. Additionally, some familiarity with the segmental aspects, especially the characterization of the activity of articulators is also assumed. Hence, the readers are strongly encouraged to consult other resources such as course texts, lecture notes, and supplementary material to gain fuller understanding of these concept if needed. The aim of Section 2 is **not** to provide a condensed form of the theoretical aspects of the suprasegmental aspects and replace these other course material. I should warn that the description in this section is neither complete nor sufficient. Rather, this crash course provides a road map to one way of how the material in this book might be approached. The section is divided into three subsections covering weak forms (subsection 2.1), connected speech aspects such as assimilations, elisions and linking (subsection 2.2), and the prosodic structure (subsection 2.3).

Section 3 is the core of the book and presents the commentaries and analyses of six individual excerpts; two in each of the three subsections. The following approach was employed in this section. For each excerpt I first sketch a basic prosodic analysis (described below in the crash course section). The first subsection 3.1 concentrates on weak forms and their detailed discussion is presented after the prosodic structure is analyzed. The second subsection 3.2 expands the coverage and after the discussion of the prosodic structure and weak forms adds the analysis of assimilations, elisions, and linking. This is to facilitate hands on practice with analyzing these phenomena as they are covered during the classes. Finally, the third subsection

3.3 presents two excerpts with a complete analysis putting together all the phenomena covered in this book.

Section 4 includes brief concluding remarks.

Importantly, in no way is the goal of the book to provide some definite analyses for you to ‘learn’. Rather, I invite you to engage with speech, compare your own awareness of speech patterns with those in my commentaries, be curious, and ask (yourself or your instructor during the seminars and lectures) if alternative analyses or understandings are possible. In essence, many situations will lend themselves to different analyses, or, in other words, the evidence in front of you might be interpreted differently. If you support your own analysis with reasonable arguments, this is what is very valuable and in fact preferable to providing a ‘correct’ analysis and not understanding the underlying principles governing such analysis. In this process, I believe you learn vital skills usable in almost any line of your future careers such as finding patterns in data, applying understanding to novel situations, forming hypotheses and checking them with data (evidence), using computer software for supporting your conclusions, and hopefully many others.

## **1.1. Motivation**

There is a relatively wide variety of available resources for practicing the pronunciation of individual words in English. Students can nowadays use various online resources and applications portable to computers or smartphones. They might include either online English dictionaries or various computer assisted pronunciation learning tools that enable recording students voices and giving feedback on their match between their and the model pronunciations. These online dictionaries and applications also usually provide both the sound files to use for imitating and improving thus student’s pronunciation as well as the transcription of the words in the form of International Phonetic Alphabet (IPA).

However, resources describing the prosodic and connected speech phenomena such as weak forms, assimilations, or reductions of the dictionary citation form of the words when they are produced in real continuous speech are much more difficult to find. Therefore, the primary goal of this resource material accompanying classes of English phonetics and phonology is to help students increase awareness of these processes and provide practice in describing them both using IPA transcription options but more importantly verbal descriptions.

It is hoped that when students engage with the real recordings using Praat, read the descriptions of the phenomena provided in the text, and finally try exercises to describe these

phenomena on their own and compare with the sample answers, the students acquire the understanding of these phenomena that will enable them to analyze novel speech excerpts in this informed way. Moreover, this understanding is, in my opinion, instrumental, in the students' efforts at limiting their non-native accent when speaking English. The dissecting of speech into individual chunks, phrases and words allows the students to practice and form new speech habits at every unit of analysis, progressively going from words, to junctures, to chunks and finally to entire utterances.

## **1.2. Material**

All the speech material was recorded as part of the Maptask Corpus (Beňuš et al. 2010) designed for research in the pronunciation patterns of Slovak speakers of English. The recordings come from the maptasks section of the corpus, in which two interlocutors seating back to back to prevent their visual contact have to jointly and collaboratively replicate a path from the start point to the end point depicted on the map of one interlocutor (Direction giver) to the map of the another interlocutor (Direction follower). The path passes various objects on the map (such as 'football pitch', 'black X', 'coast beach', 'still pool', etc.) and there are slight variations in the objects and their positions between the maps of the two interlocutors to increase the difficulty of the task and prompt spontaneous spoken interaction. Hence, speakers have to use natural communicative strategies seeking real pragmatic goals and resolving the differences between the maps in order to achieve a successful completion of the task. This design results in a very spontaneous-like interactional dialogues resembling speech in normal situational context and limits the effects of (read) laboratory speech (e.g. see review in Xu 2010) and the observer's paradox (Labov 1972) referring to the wish of the experimenter to collect natural speech data from subjects but also the unwitting and necessary influence of the experimenter and the recording situation on the collected data.

All the excerpts used in the demonstrations come from a single maptask dialogue between two Australian female speakers who participated in data collection in 2009. They both worked as English lecturers in Slovakia at that time. The reason to include Australian English in the excerpts analyzed in this book was mainly to increase exposure to a lesser known variety of English and thus to balance the British and American varieties that are dominant in the experience of most students of English Phonetics in Slovakia.

The table below shows file names included in the accompanying material to this book, text transcript of what was said, and broad IPA transcription of British English including only

the strong forms and the dictionary transcription of individual words. Hence, this IPA transcription is NOT the transcription of what the person actually said but the initial stage for our discussion, in which we discuss differences between this abstract non-existent word-by-word pronunciation and the actual fluent speech of the native speakers. The files marked in bold will be analyzed in this book. The remaining files are included for the students to attempt and practice their own analyses along the lines of the sample analyses provided in the book.

File ID	Text	IPA (broad, dictionary, British)
<b>s3_m2_73.8</b>	<b>and then it goes in between the two speech pens</b>	<b>ænd ðen it gəʊz ɪn bi'twi:n ðə tu: spi:tʃ pɛnz</b>
s3_m2_103.0	do you have a coast beach	du: ju: hæv ə kəʊst bi:tʃ
s3_m2_117.5	ok do you want me to go	'əʊ'keɪ du: ju: wɒnt mi: tu: gəʊ
<b>s3_m2_120.6</b>	<b>what you do is you go down towards them and when you get</b>	<b>wɒt ju: du: ɪz ju: gəʊ daʊn tə'wɔ:dz ðəm ænd wɛn ju: gɛt</b>
s3_m2_192.3	the top left corner of the football pitch	ðə tɒp left kɔ:nə əv ðə 'fʊtbɔ:l pi:tʃ
s3_m2_249.2	no black x	nəʊ blæk eks
s3_m2_356.8	closer to the edge of the paper than yours	'kləʊsə tə ðɪ edʒ əv ðə 'peɪpə ðæn jɔ:z
s3_m2_387.4	ok let me do the finger hooking around thing	əʊ'keɪ let mi: du: ə 'fɪŋgə 'hʊkɪŋ ə'raʊnd θɪŋ
<b>s3_m2_403.9</b>	<b>and I'm just gonna draw it to hook around the black X</b>	<b>ænd aɪm dʒʌst 'gɒnə drɔ: ɪt tə hʊk ə'raʊnd ðə blæk eks</b>
s3_m2_437.5	it goes through the middle of the coast beech	ɪt gəʊz θru: ðə 'mɪdl əv ðə kəʊst bi:tʃ
s3_m2_462.8	do you have a still pool? Do you have a still pool anywhere? So my still pool might be where your atomic bomb is	dʊ jʊ hæv ə stɪl pu:l dʊ jʊ hæv ə stɪl pu:l 'eniweə səʊ maɪ stɪl pu:l maɪ bi: weə jɔ: ə'tɒmɪk bɒm ɪz
s3_m2_498.1	I have got wide bits of metal	aɪ hæv gɒt waɪd bɪts əv metl
s3_m1_47.5	and the giant peach	ænd ðə dʒaɪnt pi:tʃ
s4_m2_43.6	where's the atomic bomb	weəz ðɪ ə'tɒmɪk bɒm

s4_m2_60.8	<b>just ignore the atomic bomb your road basically goes slightly to the left and then straight up</b>	dʒʌst ɪɡnə: ðɪ ə'tɒmɪk bɒm jɔ: rəʊd 'beɪsɪkli ɡəʊz 'slɑ:tli tə ðə left ænd ðen streɪt ʌp
s4_m2_174.2	what you need to do is stop before you get above the football pitch	wɒt jʊ ni:d tə du: ɪz stɒp bɪ'fɔ: ju: ɡet ə'bʌv ðə 'fʊtbɔ:l pɪtʃ
s4_m2_200.2	<b>they're quite close together so where you were in the middle of the fantasy god you just go straight down</b>	ðeɪ ə kwɑ:t kləʊs tə'geðə səʊ weə ju: wɜ: ɪn ðə 'mɪdl əv ðə 'fæntəsi ɡɒd ju: dʒʌst
s4_m2_323.5	and it goes maybe a centimeter	ænd ɪt ɡəʊz meɪbi ə ,sentɪ'mi:tə
s4_m2_376.2	<b>it crosses over it and goes</b>	ɪt 'krɒsɪz 'əʊvə ɪt ænd ɡəʊz
s4_m2_460.1	in the top right hand corner of the page	ɪn ðə tɒp raɪt hænd kɔ:nə əv ðə peɪdʒ
s4_m2_475.5	hang on hang on not north west north east	hæŋ ɒn hæŋ ɒn nɒt nɔ:θ west nɔ:θ i:st
s4_m2_487.8	you go to the very bottom corner left corner	jʊ ɡəʊ tə ðə veri 'bɒtəm kɔ:nə left kɔ:nə
s4_m2_495.1	have you got wide bits of metal	hæv ju: ɡɒt waɪd bɪts əv metl

### 1.3. Notes on the approach taken in the book

The pronunciation aspects of connected speech are naturally more challenging than the pronunciation patterns when practicing the dictionary forms of individual words. There are several aspects of the discussion in this book that I would like to highlight and encourage you to always bear in mind when you engage with the actual sound clips.

First, you need to bear in mind that whenever we are trying to isolate a single aspect of pronunciation for instructive and pedagogical reasons; i.e. to describe what native speakers do in this particular situation and context, that aspect of pronunciation is *never isolated* from other suprasegmental aspects of speaking arising from the situational context. To give an example, consider the instruction from the describer 'go to the left and then straight up' and the realization of linking the words 'and' and 'then'. There are several options including the weak form with the elision of 'd' [ən], nasal place assimilation when the alveolar [n] becomes dental [ɲ] anticipating the dental place of articulation of the following [ð], or the full realization of [ænd]

or even the strong form of [ænd]. How these two words will be linked together by the speaker depends crucially on the way she chunks this information, or in other words, where she decides to make a prosodic boundary. She might say this as one chunk, or may split this into two (or more) chunks and thus ‘go to the left // and then straight up’, or ‘go to the left and // then straight up’ are both possible, although the first is more likely than the second. With the prosodic boundary between ‘and’ and ‘then’, the chances of elision and assimilation are much lower than without the break. In addition to the chunking decisions, the placement of accent, or prominence on the word ‘then’ also affect the realization of the linking with the previous ‘and’. And similarly many other decisions.

Hence, we are trying to understand the choices the speaker has at her disposal and should be always aware that the particular suprasegmental aspects that we currently discuss are always intertwined with other linguistic and paralinguistic aspects of speaking and never isolated from the situational context. For this reason, the discussion of each excerpt will begin with the basic analysis of prosody (chunking into units and identifying the prosodically highlighted words), then we will proceed with the aspects of connected speech.

The first point described above is closely linked to the second aspect that you should be aware of when reading the descriptions and improving your understanding of speaking behavior. Our goal is definitely *not* to say what ‘correct’ and what ‘incorrect’ pronunciation is. In other words, the approach of the book is *not prescriptive*. Rather, the aim is to build your awareness of the speech patterns like elision or assimilation and improve your understanding how they are realized in a particular prosodic context. This, in turn, will enable you to analyze a novel excerpt of speech or dialogue on your own and describe in your own words what is most likely happening in the speaker’s mouth and why. In other words, our approach is *descriptive* in that we are trying to understand how the skilled behavior of speaking is effortlessly adapted to the situation similarly to how other forms of skilled behavior (sport, playing an instrument, handicraft, etc.) manage the same. Hence, rather than saying what a speaker should do, we are trying to understand the choices she has at her disposal and the reasons leading her to pick some of them.

This descriptive approach is fundamentally different for us as non-native speakers from some other aspects of speaking. For example, take the problem of word stress. With word stress there is a ‘correct’ solution. Specifically, we can check with a dictionary which syllable receives



the primary and which, if any, the secondary stress.<sup>1</sup> With sufficient perseverance and devotion in practicing, we can work on changing those habits that we, being non-native speakers, have developed erroneously. For example, many Slovaks would say ‘event’ as [‘i:vent] instead of [ɪ’veɪnt]. Hence, the resource is available in the dictionary, feedback is provided by the instructor when the mistake is identified, and with sufficient practice of the word with correct stress placement in many contexts and situations, non-native speakers may replace the problematic habit with a new one that is more native-like.

However, with connected speech aspects we do not have this gold standard and our non-native habits reside in linking a particular realization of words in particular context, which is not ‘wrong’ per se, just not what the majority of native speakers would do in such context. Hence, to break these types of non-native habits, we have to understand the link between the realization and context, and only then we can go on to practice replacing the old habit affected by native language interference with one that is more native-like.

Finally, the third aspect, again closely linked and briefly mentioned in the discussion of the previous two aspects above, is the role of *social context* and *pragmatic goals* in communication. There are many situations in which people have a tendency to hyper-articulate or enunciate their speech. For example, imagine two friends talking in a noisy bar with loud music and speech from other people around, or when a nurse explains the instructions for medicine taking to a senior patient with bad hearing, or when parents want to set clear limits to children who do not want to obey some routine, and repeat for the third time that the child is not to touch something. Or many other similar situations in which the context, the pragmatic goals of the speakers and other aspects favor speech that is hyper-articulated. You can compare these situations with similar contexts that normally result in hypo-articulated speech; for example when the bar is quiet with no music and no people, or when the patient has good hearing, or when the parents calmly tell the kid for the first time not to touch something. Hence, the continuum between hypo- and hyper-articulation is dependent on the situational context. Because speech in situations favoring hyper-articulation is typically louder, slower, has greater pitch range and more accented words than speech in the hypo-articulation contexts, the connected speech aspects are naturally also changing depending on this situational context.

---

<sup>1</sup> It might be, however, that context might affect the realization of primary and secondary stresses. The example I often use is the ‘Japanese’ realized as [ˌdʒæpəˈniːs] in ‘Do you speak Japanese?’ but as [ˈdʒæpəˌniːs] in ‘I have a Japanese friend.’

The discussion of the material in the book will thus help you with becoming aware of the native-like patterns but the actual practice and developing the native-like habits when you speak English is, unfortunately, left to you and nobody can do it instead of you.

In a more practical sense, the book assumes that you have the sound excerpts available and use Praat for exploring them before, during, and after reading the descriptions that the text contains. I also assume that you are familiar with Praat from the phonetics classes, but if this is not the case, there are multiple resources and online tutorials available so that every reader can familiarize herself/himself with the very user-friendly interface of this free computer software.

I provided a Praat TextGrid file for each excerpt discussed in the book and encourage you to open both the sound file and the textgrid together in Praat (highlight both objects in the Praat Objects window and click *Edit*) and listen and explore on your own. It is good to use headphones in a quiet environment when analyzing the files so that your perception of the speech patterns is not hampered by outside noise. Also, I suggest to select the spectrogram and pitch tracks while analyzing chunks (In Praat's Edit window *Show* → *View analyses...* click to *Show spectrogram* and *Show pitch*). Do not forget to adjust pitch range so that the blue contour representing f0 has minimal artifacts such as straight lines representing octave jumps. For female voices the range between 75Hz and 400Hz is usually fine. The pitch range can be adjusted in Praat's edit window in *Pitch* → *Pitch settings...* .

## 2. Crash course to basic suprasegmental aspects of fluent speech

This section presents a brief outline of the phenomena that will be discussed in the commentaries to the sound excerpts. We start with weak forms, then move to aspects of connected speech such as assimilations, elisions, and linking, and finally conclude with outlining the guidelines for identifying the basic prosodic structure in terms of boundaries and prominences.

### 2.1. Weak forms

There is a group of about 40 monosyllabic English function words that have recognizably at least two realizations: one with a full vowel and no consonantal adjustments, for example the conjunction ‘and’ as [ænd], and one with a shortened and neutralized vowel, commonly a schwa, and potentially also consonant deletions such as [ən]. The former has been termed the *strong form* and the latter the *weak form* of such words.

The list of those weak forms mentioned in Roach (2000) is summarized in Table 1 below. They are grouped into five parts of speech categories and one should be aware that verbs like ‘to be’ and ‘to have’ typically have weak forms only when serving as auxiliary, i.e. showing the grammatical relationship (e.g. ‘I *was* singing’, ‘she *has* written a letter’), and not in their full lexical meaning of ‘to exist’ and ‘to possess’ respectively. Also, ‘that’ when functioning as the demonstrative pronoun (e.g. ‘that ball is blue’) is typically realized in the strong form while the weak forms are possible for conjunctions and relative pronouns (e.g. ‘the ball that is blue’).

Table 1 Summary of weak and strong forms transcribed in IPA

Category	Spelling	Weak form	Strong form
<b>Determiners</b>	the, a, some	ðə(ɪ), ə(n), səm	ði:, eɪ, æn, sʌm
<b>Prepositions</b>	at, for, to, from, of	ət, fə(ɪ), tə(ʊ), fɪəm, əv	æt, fɔ:(ɪ), tu, frɒm, ʊv
<b>Conjunctions</b>	and, but, that	ən(d), bət, ðət	ænd, bʌt, ðæt
<b>Pronouns</b>	you, she, he, we your, him, her, them, us	jə(ʊ), ʃɪ, (h)ɪ, wɪ jə(ɪ), (h)ɪm, (h)ə(ɪ), ðəm, əs	ju:, ʃi:, hi:, wi: jɔ:(ɪ), hɪm, hɜ:(ɪ), ðem, ʌs
<b>Aux. verbs</b>	am, are, is was, were has, have does, do can, could, should would must	əm, ə(ɪ), ɪz wəz, wə(ɪ) (h)əz, (h)əv dəz, də kən, kəd, ʃəd, wəd, məs(t)	əm, ɑ:(ɪ), ɪz wɒz, wɜ:(ɪ) hæz, hæv dʌz, du: kæn, kʊd, ʃʊd, wʊd, məs(t)

As a rule of thumb, the weak forms are the default realization unless the target function word is preceding a major break or pause, or prosodically highlighted to signal pragmatic focus. Hence, for example, prepositions are in the strong form in questions in which the preposition is in the final position as in ‘Where are you from [frɒm]?’ or when they are highlighted to signal contrastive focus in ‘I am not looking for [fɔ:] you, but looking at [æt] you’.

Nevertheless, you should bear in mind two things. First, despite the apparent binary nature of strong vs. weak forms, there is in fact a continuum as to how a speaker might neutralize a strong form into a weak one, all depending on the context and speaker-dependent speaking habits. To stay with our example conjunction ‘and’, it is perfectly possible to come across forms like [ænd] or [n] or various forms of the nasal assimilating to the following consonant, e.g. [əm] when saying ‘and people’. This is illustrated in Figure 1 below. Hence, the binary division into the strong/weak form is essentially a concept for better illustration and grasping of this variability rather than the statement of reality.



Figure 1 Continuum of weakening and neutralization between strong (left) and weaker forms for conjunction ‘and’.

Second, although full lexical words like nouns, verbs or adjectives typically do not neutralize into an identifiable ‘weak’ form, there is also enormous variability in their pronunciation depending on the context. Hence, there is typically a clear difference between a noun that is accented and not accented, as will be shown in several examples, and these realization partly do use shortening and vowel neutralizations just like between the strong and weak forms of the function words. Yet, the degree of these changes is such that native speakers do not perceive different vowel qualities and thus no weak forms are identified. In other words, speaking about the continuum of reduction in the previous paragraph, this prosodic continuum is present for both functional and lexical words, but the neutralization side (right in Figure 1) of the continuum only expands to change the vowel perception in function words.

An approach to native-like pronunciation cannot be imagined without proficiency of L2 speakers in weak form production. The elusive yet so important suprasegmental aspects of speech such as rhythm and rapport in great part influence, and are influenced by, the patterns of weak form reductions. It is my firm belief that by increasing the awareness of speech habits that native speakers utilize in weak forms you will be better prepared to work on your own pronunciation. Your own rigorous training and exercise appended with the awareness and understanding afforded by the commentaries in the book will increase your proficiency of connected speech phenomena, which will essentially make your pronunciation more native-like.

## **2.2. Connected speech**

This terms is commonly used to cover the aspect of fluent speech taking place at the boundaries between individual words and they commonly include three basic phenomena: assimilation, elision, and linking. Before we discuss these phenomena in more detail, it is important to bear in mind that we are not talking about ‘rules’ of ‘correct’ pronunciation but rather tendencies that native speakers adjust to the social context of the communication as described above.

A useful approach to conceptualizing the processes of assimilation, elision, and linking is to first categorize the environments for the connected speech aspects in terms of the consonants and vowels appearing at the word boundaries. There are three basic categories:

1.  $C_f - C_i$  when the first word ends with a consonant and the second word starts with a consonant so that we have a word-final consonant ( $C_f$ ) followed by word-initial consonant ( $C_i$ ); for example ‘on paper’ or ‘most needed’,
2.  $V_f - V_i$  when the first word ends with a vowel and the second word starts with a vowel; for example ‘two orders’, ‘see objects’ or ‘four apples’ in non-rhotic dialects like Southern British English,
3.  $C_f - V_i / V_f - C_i$  when the first word ends with a consonant and the second word starts with a vowel or the first word ends with a vowel and the second word starts with a consonant; for example ‘on ice’ or ‘see people’

In the first case ( $C_f - C_i$ ), the most common generalization from the observed production data and thus the strategy that native speakers commonly use is simplification. The two adjacent consonants require the formation of two constrictions in the oral cavity and there are two basic ways for the simplification in suitable social context as described above. Either the adjustments to the consonant characteristics (e.g. the place, the manner of articulation or voicing) take place, resulting in the *assimilation* of some or more of these qualities to the adjacent consonant. Or the entire consonant could be apparently left out completely, which would result in *elision*. Importantly, however, some consonantal clusters are simplified more readily than others and thus not only the social context, pragmatic goals and the prosodic structure plays role but also the phonetic features of these consonantal clusters.

In the second case ( $V_f - V_i$ ) we have a hiatus of two adjacent vowels. The preferred strategy of speakers in this case is to actually insert a consonant to break this hiatus and link the two vowels. In English, most common linking consonants would be glides, rhotics and glottal stops [j, w, ɹ, ʔ]. The insertion of a consonant in this situation is a general process known also from other aspects of phonology within words. The reason is likely the general preference for syllables that have onset consonants. The situation like V.V as in this case, in which the two vowels serve of nuclei of two syllables, is always dispreferred compared to sequences like V.CV. In other words, V.CV (or a CV syllable) is a less marked structure than V.V (or an onsetless syllable V). This is known cross-linguistically by studying many languages since there are many languages that have both CV and V syllables, like Slovak or English, many that have mostly CV syllables, like Japanese or Hawaiian, but no language that would have V syllables and not the CV ones. This markedness principle maybe be also observed in speech development when babbling of infants, considered the onset of speech production, is predominantly in the CV forms.

In the third case ( $C_f - V_i, V_f - C_i$ ), since we have adjacent vowels and consonants, the simplest strategy is just to link them as we would do within the words and thus facilitate fluent speech production.

Since this section is not meant to provide the theory of the connected speech aspects, I will not discuss assimilation, elision, and linking in more detail here but the students are referred to available textbooks and other course materials. A good strategy, and the one we will adopt when discussing the individual sound clips, is to go over each word boundary, identify which one of the three types (C-C, V-V, or CV/VC) it is, and then analyze what is actually happening, or likely to happen.

In terms of practicing to improve your pronunciation, connected speech phenomena are somewhat different from weak forms. In weak forms, many non-native speakers of English have tendencies or habits not to reduce the words and use them in their full, spelled strong forms. This is partly due to the fact that in Slovak, these reductions are much less extreme and difficult to even perceive. On the other hand, assimilation, elision, and linking are processes more natural to speech in general and frequently occur also in Slovak or Hungarian. However, the devil (which in this case refers to the perception of non-native accent) is in the detail, and while some processes come to us more naturally like ‘on paper’ pronounced as [ɒmpɛɪpə] when the alveolar [n] changes to bilabial [m] in anticipation of the following bilabial [p], some other assimilation patterns are different in English and in Slovak. Consider for example the tendency to devoice pre-pausal voiced stops and fricatives. Native speakers of Slovak (like speakers of other languages such as Polish, German, or Catalan) tend to say [t] in ‘plod’ or [x] in ‘roh’ instead of the voiced [d] and [ɦ] respectively that we use in conjugated forms like ‘plody’ or ‘rohy’. However, this habit is different in native speakers of English for whom the voicing of the pre-pausal consonants tends to be at least partially preserved: people say [g] in ‘dog’ or [z] in ‘eyes’ which contrasts with ‘dock’ and ‘ice’ with final [k] and [s] respectively. Hence, with weak forms, developing native-like habits is not affected much by the habits in native speech while in connected speech aspects, some phenomena are effortless but other require small adjustments for which the mother tongue interferes. In my experience, the combination of minor adjustments away from native speech patterns is a more difficult one than forming a relatively novel habit that is not present in the native language.

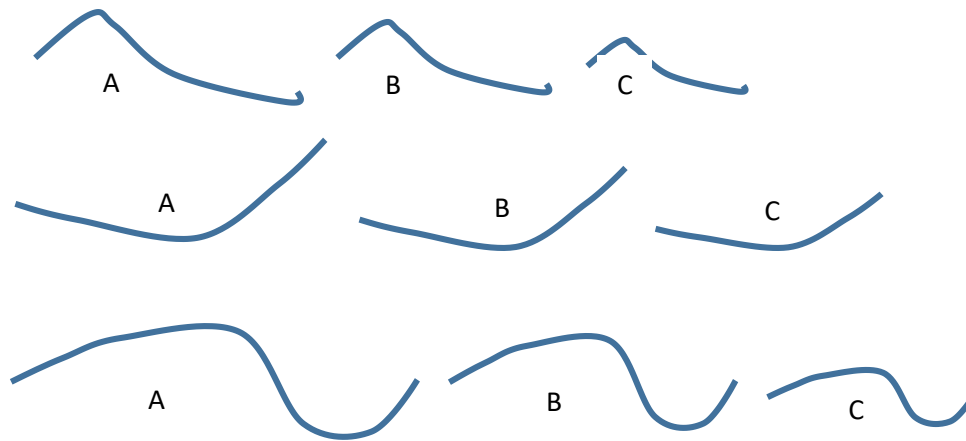
### 2.3. Prosody: Prominence and chunking

There are many aspects of speaking that affect the overall prosody; for example voice quality, speech rate, or the range of variability of our pitch. However, in this very brief overview we only discuss the most basic aspects of prosody linked to signaling which words are prominent (accented) and how words combine into larger prosodic units (chunks). Both of these aspects are intimately linked to the fluent speech aspects we discussed so far such as weak forms and connected speech phenomena. Additionally, accenting and chunking have a clear relationship to pragmatic meanings of utterances whereas voice quality or speech rate are more linked to the emotions and attitudes expressed through prosody and are thus more depending on individual speakers, which makes them more subjective and difficult to study systematically. But research on emotional prosody is advancing rapidly both as basic research found many scientific publications and as applied research for example in endowing automatic speech synthesis in robots, avatars, or computer games with speech with natural-like prosody and emotion.

A good strategy, and the one we will adopt when discussing the individual sound clips, is to decide first if a given sound excerpt forms a single chunk or multiple chunks. The chunks are commonly referred to as tone units in the British tradition of intonation description, and as intonational phrases in the American tradition. Typically, every *silent pause* means a boundary between chunks, but separate chunks can often be found also without an actual pause. The most typical markers of a boundary presence beside a pause is *pre-final lengthening* when the final syllable of a word preceding boundary is longer than usual, and *major pitch excursion* before the boundary when the melody undergoes significant changes.

Additionally, so called *pitch reset* might also be assessed but this requires a bit more expertise. Basically, speakers normally use the middle of their registers and have a tendency to declination (pitch lowering and register narrowing partly due to expelled air and lowering of subglottal pressure) in most cases, or alternatively a tendency to rising pitch and expanding register as in yes/no questions. A tendency to set the default pitch range anew at the beginning of each chunk thus can be used as corroborating evidence for the presence or absence of a prosodic boundary. Several cases of abstract stylized pitch contours showing resets are illustrated in Figure 2 below.





*Figure 2*      *Stylized examples of three utterances each with three chunks A, B, C and how pitch reset can be realized in boundaries between A and B, and between B and C.*

Although various labeling schemes recognize different number of boundary strengths, for our purposes, we will mostly use the major boundary (intonational phrase) marked with a double slash “//” and when appropriate also the minor boundary (intermediate phrase) marked with a single slash “/”. All four major boundary markers (pause, pre-boundary lengthening, pitch excursion, pitch reset) are continua and are expected to be more salient in the major boundaries and less salient in the minor boundaries.

It is also worth mentioning that the syntactic structure of speech might provide some guidance in determining chunks and boundaries. However, keep in mind that in spontaneous speech, as compared to read speech for example, the syntactic boundaries are not necessarily aligned with the prosodic boundaries. In other words, breaks between chunks might coincide with syntactic boundaries (either marked in text as commas or full stops) or inherently present (e.g. the boundary between a subject and a verb being stronger than between a verb and an object so that ‘John / bought a book’ is more likely than ‘John bought / a book’), but this is just a tendency and is often not respected in spontaneous speech.

So a good approach for chunking as the first step in the intonation analysis is to determine if a given stretch of speech can be divided into separate units (chunks) with a boundary searching for pausing, lengthening, and pitch movements. If such a boundary has been identified, then the same should be applied to the smaller chunks defined by that boundary recursively until each chunk is perceived as a single unit without breaks.

After this first step, we might proceed to the second step of this basic prosodic analysis which is the identification of prominences, or accents, in each chunk. That is, we need to determine the words that are prosodically highlighted. A convenient way to mark such accents in text is to use a star “\*”, capitalization, underlining, or other diacritics. In this book we will use the capitalization of the accented syllable. The only requirement is that each chunk has at least one accent.

In English, the last accent of each chunk is typically the most prominent and some descriptions of English intonation give it also a special name; e.g. the nuclear accent. The most pronounced pitch movement is typically associated with this last accent and adjacent boundary. In other description of English intonation, the melodic characterization of pitch accents is separated from the characterization of the boundary tones. The pitch movements associated with the last nuclear accent and the following prosodic boundary taken together constitute the perceived prominence of this location in general.

In my experience, it might be helpful to try to clap hands to identify the ‘rhythm’ of the chunk. Most likely, the individual claps correspond to the accents in the chunk. You may do it while listening to the excerpt first, and then imitate the pronunciation of the chunk the way native speaker says it and clap hands. Clapping hands helps us identify the rhythm and melody of the chunk since these two are intricately linked in English. There are several approaches to teaching English prosody that use similar approaches, for example C. Graham’s jazz chants (2001). For us it is essential to become aware of the differences in prosodic highlighting of individual words, how this participates in cuing the prosodic structure of utterances, and ultimately the signaling of pragmatic meanings given a situational context.

In this book we will not be concerned with characterizing pitch accents and boundary tones further. However, it is good to keep in mind that both accents and boundaries may be characterized a ‘contours’ like fall, rise, etc., or ‘targets’ like high or low. For a more detailed description of pitch accents and boundary tones, you are referred to other course material.

Given that the boundaries and accents are the most basic and natural ways of signaling the prosodic structure, and by extension pragmatic meanings in particular situational contexts, the realization of aspects of connected speech we discussed in sections 2.1 and 2.2 is in crucial ways dependent on this structure. For example, weak forms are extremely unlikely in the pre-boundary positions and most likely found in the chunk-internal positions. Similarly, aspects like assimilations and elision have a greater chance of occurring within the chunk than across a major boundary (albeit these are still possible). Furthermore, a greater chance of simplifications like assimilation in between two words exists when the two adjacent words are unaccented than

if one or both of them are accented. But again, these are just tendencies. On the other hand, an accented weak form is plainly impossible since all weak forms are unaccented.

### 3. Commentaries to individual excerpts

For each of the sound file excerpts, the CD and the dedicated website includes the sound file and Praat Textgrid file with the annotations. As mentioned before, the suggested approach is that you open the sound file in Praat, listen to it multiple times, attempt a basic analysis of prosody determining chunks and accents, and note aspects of connected speech (weak forms, assimilations, elisions, linking). Afterwards, compare your notes with the commentary provided in this book and work on understanding the differences. Note again that if you've analyzed something differently from me, or if you found some aspects not covered in my commentary, it does not mean you are wrong! It is perfectly possible that I may have overlooked things or that alternative and reasonable possibilities not mentioned in the commentary exist. Always try to answer the 'why' question; hence, try to understand what is going on given your understanding of the theoretical aspects and the evidence you see in Praat and can hear with your ears. Do not accept my descriptions unless you understand *why* I analyzed the file that way. Ideally, most of the differences between yours and mine analyses can be resolved by yourselves but I would be very grateful and open to discuss things you still do not grasp after you've made an honest effort at them using the book and the files.

#### 3.1. Weak forms

This section consist of commentaries to the first two sound clips; one from each speaker. We limit our discussion to the prosodic analysis and the realization of weak forms linked to the analyzed prosodic structure.

##### 3.1.1. s3\_m2\_120.6: what you do is you go down towards them and when you get

As explained in Section 2, we start with the analysis of the prosodic structure. Firstly we need to determine the chunks. It is obvious that this excerpt consist of multiple chunks. This is despite the fact that there is in fact only a single silent pause between 'down' and 'get'. Hence, let's make the first boundary splitting this utterance into two chunks at this boundary. Now we take the first chunk created this way ('what you do is you go down') and ask the same question recursively again: Can we split this chunk any further? Do we perceive that the speaker did not produce this as one unit but as more? The answer is, clearly in this case, yes. There is a definite break/boundary after 'do', and this is despite the fact that there is no actual silent pause between

‘do’ and ‘is’. The perception of a major boundary here comes from the extreme lengthening of ‘do’, and a very pronounced pitch movement associated with such boundary. The presence of this boundary also coincides with the syntactic boundary.

Using the recursive process that should be familiar by now, we check if the two newly created chunks ‘what you do’ and ‘is you go down’ can be further divided into even smaller units. For the first chunk the answer is relatively clear since the perception of a single unit is unproblematic. The second chunk (‘is you go down’) is much more challenging since the speaker definitely lengthens the word ‘go’. However, we see no pronounced pitch movement. Hence, since we are here interested only in a basic rough prosodic analysis, we treat the second chunk as one unit as well. Of course, it is interesting to note how the slowing down takes place in the second chunk, compared to the first chunk as well as the speech following this second chunk. We might speculate that this slowdown is related to cognitive demands of finding the best way for describing the instructions to the interlocutor.

Finally, we ask if the remaining chunk after the silent pause ‘towards them and when you get’ forms a single unit or could be split further. This is also not straightforward since one might perceive a disjuncture between ‘them’ and ‘and’, which would be supported by a syntactic boundary at this place and very minimal pitch lowering. Although there is some evidence for a minor boundary here, for our purposes of a rough analysis, we opt for the simpler analysis and consider this chunk as a single unit. So the chunking process resulted in analyzing this utterance as forming three units with a single possibility of a minor boundary after ‘them’:  
// what you do // is you go down // towards them (/) and when you get //.

Once we have the chunks, we proceed to the second step of the analysis determining which words are accented. Hence, for each chunk, we listen to it separately in Praat and try to identify if any of the words are highlighted more than the others. You may also try clapping with your hands while imitating the chunks, as explained in Section 0 above. It is also good to use clapping for considering alternative hypotheses. For example, if you hypothesize that in the first chunk ‘what you do’, all three words are accented, you might try to say it with a clap on each word and then listen to the chunk again to see if that is the closest rendition of the native speaker’s speech. Similarly to determining the chunks, some accents are easy to perceive and some are more challenging. And remember, we are not trying to determine the single ‘correct’ analysis; rather, we are trying to engage with the speech and understand what is going on.

In the first chunk, I perceive two accents, one on ‘what’ and the other on ‘do’. Note that the second accent is clearly more prominent than the first one, which is a typical situation already described as the nuclear accent. In my experience, many Slovak students have a

tendency to highlight (and in fact also perceive) the wh-word more than the nuclear accent. This is partly supported by some realizations of wh-word questions in Slovak, but partly probably also by simply linking directly the pragmatic meaning (what the question is asking about) and the prosodic realization (what word is the most accented).

The second chunk ‘is you go down’ is, similar to the boundary determination, much more challenging. The words ‘go’ and ‘down’ are significantly lengthened but show a very flat f0 contour. Hence, I would label an accent on ‘is’, and then also on these two words but I consider them weak accents.

The final chunk ‘towards them and when you get’ has, in my perception three accents: ‘towards’, ‘when’, and ‘get’.

Hence, our rough prosodic analysis gives us this prosodic structure for the clip with accented syllables marked with capitalization: // WHAT you DO // IS you GO DOWN // toWARDS them (/) and WHEN you GET //.

Now we are in position to consider weak forms. We know that weak forms might occur when function words are not accented or in pre-boundary positions. Let us start with the pronoun ‘you’ that appears three times in this clip. First, listen to the intervals with ‘you’ separately in Praat by clicking on the interval (it turns yellow), and then clicking the bar below or above this interval. What you hear is definitely *not* [ju:] with a long back rounded vowel. This is an important self-exploring exercise to realize that native speakers really do not say [ju:]; in these cases they do not even say [jʊ] but the realizations sound like [jə] or [jɪ].

All of them are also very short at around or less than 100ms. You may compare these ‘you’ syllables with other accented monosyllabic words in the clip like ‘what’, ‘do’, ‘is’ that are all more than 300ms long; hence three times as much as ‘you’ syllables.

The next potentially weak form is the pronoun ‘them’. Repeat the procedure and consider the quality of the vowel. Is it [e], as we would expect in the strong form, or more like [ə] that we would expect in the weak form? I would vote for the second option. One additional way of checking would be to compare ‘them’ with the realization of ‘then’ or ‘when’ where we would expect [e], especially in the accented ‘then’. This comparison also shows quite clearly the reduction of the vowel in ‘them’ into more centralized schwa-like vowel. Hence, we can conclude that ‘them’ was realized as a weak form.

This realization of ‘them’ also nicely shows the *continuity* of phonetic characteristics that I discussed in Section 2.1. Recall that we contemplated the presence of a boundary between ‘them’ and ‘and’ and finally settled on the possibility of a minor boundary. Now recall the general default pattern for the distribution of weak and strong forms: if in pre-boundary

position, strong forms are more likely. The word in question ‘them’ sounds like a weak form and occurs preceding a minor boundary. This shows how various requirements and tendencies are intertwined and influence each other since if the intention of the speaker was to make a more prominent boundary, the realization of ‘them’ before it would most likely be less neutralized. On the other hand, pronouns like ‘him, her, them’ occur happily also before a strong boundary formed by a silent pause. Imagine a call ‘Get them!’ with a perfectly reasonable realization of [ðəm] or [əm].

The next word we discuss in this clip with regard to weak forms is ‘and’. When characterizing the prosodic structure we described this word as not accented, but it does appear in the chunk-initial position, that is sometimes also prosodically prominent. Studies like Pierrehumbert and Talkin (1992), Fougeron and Keating (1997), or Cho (2006) among others found that chunk-initial positions, similarly to chunk-final positions, show various strengthening and hyper-articulation effects such as greater aspiration, longer contact durations (hold phases) for stops, or other articulatory kinematic effects. Listening to this ‘and’ separately, we can conclude that this is certainly not the strong form [ænd]. Even though at the first listening, the vowel might sound like [e], which many Slovak speakers have a tendency to pronounce in this word, the comparison with clear [e] realizations of ‘then’ or ‘get’ in this chunk clearly shows that the vowel in ‘and’ is neither [æ] nor [e] but sounds more like schwa. Furthermore, the final [d] is also extremely contracted, but this will be discussed with other connected speech phenomena.

Finally, consider the word ‘do’ in the first chunk. Its realization is an unquestionable [du:], and all structural prosodic characteristics supports this: it is accented and in pre-boundary position. But note that this ‘do’ does not fall into the group of function words potentially having weak forms. This is because here ‘do’ is a full verb and not an auxiliary one. In a slightly modified sentence ‘what do you do’, the first ‘do’ is auxiliary and the second lexical and thus only the first one is commonly realized as weak while the second one as strong.

### **3.1.2. s4\_m2\_60.8: just ignore the atomic bomb your road basically goes slightly to the left and then straight up**

As before, we start with the identification of prosodic units, i.e. chunks. Contrary to the first clip in which there was at least one silent pause, there is no clear silent pause in this one. Despite this, we cannot say that this utterance consists of only a single chunk. There is a relatively robust

boundary after the word ‘bomb’. Although pre-boundary lengthening is not very perceptible – ‘bomb’ is not much longer than it would normally be in the middle of a chunk – the strongest evidence for the presence of the boundary is the pitch movement. We can see and perceive a so called continuation rise, or the fall-rise melody in the British tradition of describing intonation during the last word of the chunk. Additionally, we can see some evidence of pitch reset.

Now we need to check if the two newly identified chunks can be divided further. The first one ‘just ignore the atomic bomb’ does not contain any indication of a prosodic break and the junctures between the words within this unit are smooth.

The second chunk ‘your road basically goes slightly to the left and then straight up’ is, however, much more challenging. My first impression was that this unit can be subdivided into three chunks ‘your road basically goes’, ‘slightly to the left’, and ‘and then straight up’. This would correspond to the flow of the speech. However, looking for the evidence for the first break after ‘goes’, we don’t see much. There is no perceivable lengthening and the pitch is rather flat. Playing just this chunk alone, I have a feeling of incompleteness as if the unit was not concluded. The situation with the second possible break after the word ‘left’ is similar. Here, we can perceive some lengthening and a slight fall in pitch but the visual inspection shows a rather flat  $f_0$  contour during this word. Hence, the chunking part of my prosodic analysis identifies one major and two minor boundaries giving this structure: // just ignore the atomic bomb // your road basically goes (/) slightly to the left (/) and then straight up //.

Next we move to identifying the accented syllables in the individual chunks. I think that with the first chunk there are two plausible hypotheses to consider. The first one is more minimalistic with accents on ‘ignore’ and ‘bomb’, and the second one assuming that all words except ‘the’ are accented; hence that there are four accents. In my rendition of clapping together with the speech of the native speakers, both sound reasonable to me. After listening to the chunk multiple times, I would vote for the first hypothesis but the four-accent solution is also plausible. My main reason is that ‘just’ is not sufficiently highlighted to have a clear full vowel [ʌ] and the pitch contour during ‘atomic’ just drifts between the high target of ignore and low target at the beginning of ‘bomb’ without a clear pitch excursion.

After I made this decision I looked into the original file, and the speaker mentions ‘atomic bomb’ for the third time in row (‘Where is the atomic bomb? Just ignore the atomic bomb I reckon’). And there are additional mentions of this object by the other speaker before. Hence, it is plausible that our speaker does not feel the need to accent ‘atomic’ very much since it is clearly established now which object the speakers are talking about and a single accent covering the object ‘atomic bomb’ is sufficient. Despite this analysis, labeling with four accents



is also reasonable and might be supported by a perceived pitch peak on ‘atomic’ and a relatively high f<sub>0</sub> during ‘just’.

Accenting in the second major chunk of this utterance is, in my mind, relatively straightforward. I perceive that the following words are accented: ‘road’, ‘basically’, ‘slightly’, ‘left’, ‘then’, and ‘up’. The only somewhat unclear cases might be the words ‘goes’ and ‘straight’ that are the only full lexical words that are not accented (there is one function word ‘then’ that is accented). Considering ‘goes’ as unaccented is supported by its short duration and subsequently reduced realization of the diphthong. This is appended by very robust accents on both adjacent words ‘basically’ and ‘slightly’ and for rhythmical reasons de-accenting ‘goes’ sounds natural. With regard to ‘straight’, clapping hands during the last minor phrase ‘and then straight up’ suggests that two-accent solution is closer to the native speaker rendition than the three-accent one.

To sum up our basic prosodic analysis of this chunk, we can label it in the following way: // just igNORE the atomic BOMB // your ROAD BASically goes (/) SLIGHTly to the LEFT / and THEN straight UP //.

With this in hand, we can proceed to discussing weak forms in this utterance. There are these candidates to consider: ‘the’, ‘your’, ‘to’, ‘the’, ‘and’. Recall that even ‘the’ can have a strong form and thus needs to be considered among the candidates. We can quickly dispense with two cases of ‘the’ since listening to them separately in Praat clearly shows they are produced with a schwa rather than with the strong form of [i:].

Now listen to ‘your’ as a separate word. Again, this is not even close to [jɔ:] what I suspect many Slovak speakers would say in this situation. This is a prototypical example of a very reduced weak form and listening to ‘your road’ together and trying imitating several times might be beneficial for those who have a tendency to use [jɔ:] in this context.

The next candidate ‘to’ is extremely interesting. I suggest you first listen to the minor phrase ‘slightly to the left’ several times. Try to determine how the native speaker produces ‘to’. After you have your best guess listen to this word alone by selecting the appropriate interval. Is this close to your guess? When you listen to it separately you notice that the vowel is not even present! Hence, the strong form of [tu] is weakened such that it contains not just a vowel reduced to [ʊ] or [ə] but it is shortened to such an extent that the aspiration from the release of [t] is all what is left from the vowel. You might compare this with the relatively long schwa of the following ‘the’.

Finally, consider the conjunction ‘and’. It is not accented and thus we might expect the weak form. Listening to this word we see that the realization is something like [ən], which, on

the continuum between strong and weak forms, is definitely closer to the weak form than the strong one.

### **3.2. Connected speech aspects**

In this section we strengthen the understanding of weak forms from the previous sections and expand by including also the discussion of connected speech aspects across the word boundaries. I strongly encourage you to go over the two subsections below, and then try going back to the two files of the preceding Section 3.1 and identify the connected speech aspects there as a very useful exercise.

#### **3.2.1. s4\_m2\_376.2: it crosses over it and goes**

We start with a short utterance ‘it crosses over it and goes’. Due to the relatively short duration, the prosodic analysis is rather straightforward. Despite the absence of any silent pause, I think this utterance can be divided into two chunks: ‘it crosses over it’ and ‘and goes’. The major phonetic signature of the boundary between ‘it’ and ‘and’ is neither lengthening of ‘it’ nor major pitch excursion before the boundary. In this clip it is the presence of glottalization. This is a term used to characterize the creaky quality and non-modality of voice and might be linked also to low pitch. For example, you probably achieve glottalization if you tried continuously saying [a:] with progressively lower and lower pitch. However, it is important to say that glottalization, or sometimes popularly also called vocal fry, might be a prevalent socio-linguistic feature in various social groups and contexts; for example with young females. In the spectrogram, you see glottalization as a series of vertical striations corresponding to irregular pulses emitted by the vocal cords. Although perceptually we hear low pitch in ‘it’, the pitch tracker in Praat does not show a blue contour precisely due to the irregular pulsing. Beware that in some other situations, Praat might show f0 tracks with various artefacts to be disregarded due to pitch halving or doubling, which is quite common in non-modal voicing quality like glottalization.

Since neither of the two chunks lends itself to further subdivision, we divide this utterance into two chunks. The first chunk is best characterized with two pitch accents on ‘crosses’ and ‘over’, which could be very well supported in clapping hands with the native speaker. The second chunk has just one accent on ‘goes’. If you listen strictly to the interval ‘and goes’ only, I see how some might consider also an accent on ‘and’. However, this becomes

impossible once we listen and clap with the entire utterance, and moreover, the length and the quality of the vowel in ‘and’ suggest that it is somewhat neutralized, which is highly unlikely with accented syllables. Hence, the prosodic analysis as a pre-requisite of further exploration of connected speech is: // it CROSSes Over it // and GOES //

With experience from the previous section 3.1, let us start with weak forms and then continue to connected speech aspects. The only candidate for a weak form here is the conjunction ‘and’. As we said above, the vowel is somewhat neutralized towards the schwa, and the final ‘d’ is missing. These two reasons are sufficient to describe this ‘and’ as a weak form.

Let’s continue now with the juncture between ‘and’ and the following word ‘goes’. We just said that the final ‘d’ is missing. In other words, there is certainly no release for ‘d’. This is one of the possible solutions in cases when the first word ends with a consonant and the second one starts with a consonant ( $C_f - C_i$ ). Moreover, we have a situation in which the juncture of two words created a cluster of three consonants (n, d, g) and the middle one is an alveolar stop ‘d’ or ‘t’, which is a prototypical environment for elision in fluent English.

One needs to be careful because the elision of a plosive consonant when followed by another plosive, as would be the case here when word-final ‘d’ is followed by a word-initial ‘g’, is not a ‘rule’. Other solutions might be available as we will shortly see when examining the first two words of this utterance.

But now back to the juncture ‘and goes’. Since we have determined that ‘d’ is omitted, we are left with another  $C_f - C_i$  situation with a word-final alveolar nasal ‘n’ and word-initial velar plosive ‘g’. In this situation, the very common solution is the *nasal place assimilation*. The alveolar nasals, and especially in a weak form of ‘and’ are produced with a rather weak tongue constrictions and the fluency of speech is greatly enhanced if the tongue does not form two constrictions (one for ‘n’ and one for ‘g’) but if only a single constriction is made.

Which of the two constrictions should survive and which should give in? There are two reasons to favor the survival of the velar constriction of ‘g’. First, the word-initial position is much more important, for example in the identification of the correct word, than the word-final position. Hence, the onset of the word-initial position enjoys a special status that their elision or modification are quite limited as this might interfere with correct decoding of speech. Second, the word-final ‘d’ ends the function word already shortened and neutralized into a weak form and thus further modifications of this word are not costly since ‘and’ is quite predictable here.

Hence, the word-final ‘n’ assimilates to the word-initial ‘g’ in terms of their place of articulation such that the alveolar quality changes to the velar place to agree with the place of the articulation of the following plosive ‘g’: [n] → [ŋ]. This is thus a case of *regressive assimilation*, sometimes also called anticipatory, since the word-initial consonant triggers the change in the preceding word-final consonant and thus the direction of assimilation is reversed to the natural flow of speech. In other words, when the speaker is going to say [əŋgəʊz], she anticipates the velar place of articulation of ‘g’ and changes the alveolar [n] into a velar [ŋ] to say [əŋgəʊz], which is more economical for speech production and does not hinder speech perception.

The example of juncture ‘and goes’ shows several important observations. First, the processes of connected speech in fluent productions might cluster together and it is perfectly plausible to observe more than one process at a single juncture. Here, for example, we described elision, assimilation, and the first word was a weak form. Second, for some multiple processes temporal ordering might be important. In this case, nasal place assimilation occurred only when elision took place and the former would not be very likely without the later. Third, some phonetic characteristics are easier to observe visually in the spectrogram or auditorily than others. For example, the absence of d-release is plainly visible in the spectrogram and it is also quite clear after listening to the juncture ‘and goes’. On the other hand, the evidence for nasal place assimilation, and the actual existence of [ŋ] rather than [n] is less robust. The two nasals are very similar, and if we do not want to go into details of formant transitions, which are beyond the scope of this book, the phonetic evidence is difficult to pinpoint. One hint might come if we listen to the interval starting with the beginning of ‘and’ and ending at the release of the velar ‘g’ of ‘goes’ at the temporal point of about 1.25s into the file. This interval sounds like [əŋk].

Let us now explore the second case of  $C_f - C_i$ , i.e. the situation when the first word ends with a consonant and the following word starts with a consonant: ‘it crosses’. We have adjacent voiceless plosive consonants here; one is alveolar and the other has the velar place of articulation. Many students have a tendency to analyze cases like this as elisions of the first consonant [ɪtkrɒsɪz] → [ɪkrɒsɪz]. Indeed, we do not hear the release of [t] when we just zoom to the interval of the juncture. However, imagine how some words with [ɪkr] sequence would sound, for example in ‘decry’, ‘decree’ or others. The duration of the closure for [k] would most likely be shorter than what we see and hear in our case of ‘it crosses’. With this juncture, you should hear and feel that temporarily, the time slot for [t] did not disappear, or in other words, that there is still something between [ɪ] and [k].

Given that the temporal organization suggests the presence of the word-final consonant, what is the natural articulatory strategy for producing two plosive consonants (requiring two complete obstructions) in a row without the release of the first closure? Note that the two plosives only differ in their place of articulation since both are voiceless. The strategy again involves some adjustments to the first of the consonants. In this contexts, two such adjustments might be considered.

The first alternative for the speaker would be to employ regressive assimilation of place, in which the [t] of 'it' changes to a velar plosive [k] to be identical in terms of the place of articulation with the following initial consonant of 'crosses'. This would produce the juncture [ɪkʰkrɒsɪz] with the symbol 'ʰ' denoting the absence of the plosive release. It would be similar to a 'long' consonant that are common in languages like Hungarian or Italian, and appear in Slovak in careful pronunciation of some words like 'panna', or 'mäkko'.

The second alternative, that I think might be somewhat more common especially in some dialects of English would be to glottalize the word-final [t] of 'it'. Hence, the closure for the alveolar plosive would not be created by the tongue blade touching the alveolar ridge but by closing the glottis; i.e. the space between the two vocal cords in the larynx. This would create a glottal stop which is a very common sound both in speech but also in non-speech sounds. For example, when you cough, you start with the closed glottis that is rapidly opened. In many dialects of British English, such as Cockney, speakers produce these glottal stops in places of 't' sounds also within words such as 'bottle'. Replacing the place of the obstruction from the supraglottal area above the larynx into the larynx itself thus frees the tongue to make just a single closure of [k] and thus simplify the production of two plosives in a row. Hence, the second option would be to say [ɪtkrɒsɪz] → [ɪʔkrɒsɪz] with the IPA symbol [ʔ] denoting the glottal stop.

Now, which of the two alternatives, both perfectly plausible, the speaker employs? My money would be on the first case of assimilation. This is because in the spectrogram I do not see any irregularities in the vocal pulses before the closure actually takes place. However, this type of reasoning is already quite advanced and the 'correct' answer is not as important as understanding the context and possible realizations of the juncture 'it crosses'.

The last juncture we will discuss in detail in this utterance is 'over it'. In non-rhotic dialects, this is an example  $V_f - V_i$  environment, i.e. the situation when the first word ends with a vowel and the following word starts with a vowel: [əʊvə ɪt]. From the brief outline in Section 2.2 we know that the natural tendency in these cases is to link the two adjacent vowels with a consonant. This is a prototypical environment for a 'linking r' since the first word ends with a

‘r’ in spelling and can thus naturally function as a linking consonant. Since our speaker is a native of Australia, and Australian English is commonly a non-rhotic dialect, the presence of ‘r’ in this juncture is due to the linking function of the consonant. The speakers of rhotic dialects like northern British or General American would most commonly pronounce [ɹ] irrespective of the environment in which the word ‘over’ occurs while the speakers of non-rhotic dialects would only use it to link with the following word starting with a vowel.

The only juncture of this utterance we have not discussed so far is ‘crosses over’. This is an example of  $C_f - V_i$ , and thus regular linking between a vowel and a consonant takes place.

We see that despite the short duration of this utterance; there were multiple processes characteristic of fluent pronunciation and all junctures except one represented either a prosodic boundary or some aspects of connected speech.

### **3.2.2. s3\_m2\_117.5: okay do you want me to go under A-gray e- grade eggs**

There are several reasons for including this clip for an analysis in this book, and one of them is the presence of a speech error and how the speaker deals with it. This of course impacts also the prosodic structure of the utterance.

The error concerns the production of the compound word ‘A-grade eggs’, which is one of the objects on the map that the two interlocutors need to navigate. The speaker said something like ‘grey’ instead of ‘grade’, hence she did not pronounce the final alveolar consonant. She caught herself at the beginning of the next word ‘eggs’, stopped, and produced the repaired ‘grade eggs’ after a short pause of about 120ms. An interesting question then opens regarding this silent pause: should we treat it as a boundary signal? You notice that there is no lengthening or pitch excursion on the syllables before the pause and there is also no pitch reset characterizing the boundary. In other words, all evidence shows that the speaker did not plan this to be a boundary and the silent pause only indicates the time required to put together an articulatory plan for the corrected ‘grade’.

With Praat, we can easily test this hypothesis by cutting the error out of the signal and listening to the clip. If you want to do this yourself, here is how. Read the sound (s3\_m2\_117.5.wav) and the associated Textgrid file (s3\_m2\_117.5.TextGrid) into Praat Objects window. Select both files and click *Edit*. Now identify the interval with the error and select this interval. It includes three intervals of the TextGrid: “grade e- #” and starts at around 1.468s and ends at around 1.93s. You can do this either manually by dragging the mouse over the spectrogram starting at approximately the beginning of ‘grade’ and ending at around the

end of “#”, or you may click *Select* → *Select...* in Praat editing window and enter the times 1.468 and 1.93 in the respective fields. Now make sure that the ‘Group’ field in the bottom right corner of the window is checked. Now go back to the Praat objects window and select only the sound and click *View & Edit*. You should see the sound with the interval between 1.468 and 1.93s pre-selected. Listen to the entire file by clicking ‘Total duration’ bar at the bottom of the window. Now go to *Edit* → *Cut* or use the shorthand *Ctrl-X*. This removes the selected interval from the clip. Listen again by pressing the bottom bar ‘Total duration’. You can go back one step with *Ctrl-Z* and restore the cut interval and listen again.

What I hope you will observe is that cutting the interval corresponding to the speech error before the repair results in an utterance perfectly complete intonationally and without any evidence of unnatural breaks, resets, or evidence for incongruity. Hence, we may conclude that, prosodically speaking, the speaker started her repair very close to where she started the original erroneous production of ‘grade’. And thus the silent pause characterizes the speech repair, and belongs to so called ‘edit signal’ or ‘disfluency interval’ (e.g. Nakatani and Hirschberg 1993) cueing the listener that an error has been made and a repair is about to start. A filled pause like ‘uh’ is also a common edit signal. This pause is not a typical prosodic boundary signal and we will not treat it as a boundary between two chunks in this discussion. However, longer such pauses, and possibly with some boundary signals present, are possible in speech.

So with the speech error and the silent pause associated with it now discussed, let us move to the rest of the prosodic analysis. Do you perceive any junctures that would constitute natural prosodic boundaries in this utterance? I perceive such breaks after ‘okay’ and after ‘go’. Neither of them contains a silent pause and both are relatively minor in my perception.

The first boundary separating ‘okay’ from the rest of the utterance is completely natural in terms of information planning, syntax, and pragmatics. This is because ‘okay’ here refers to the information previously provided by the other speaker and ‘okay’ acknowledges the reception of such information and signals to the interlocutor that with respect to this information they are on the same page. The material following ‘okay’ is the new information the current speaker puts forward. If you listen to just the ‘okay’ interval, there is a sense of completion as if this could have stood alone also in the conversation. We also see and perceive a slight fall in  $f_0$  albeit not much lengthening of the pre-boundary syllable.

The second prosodic boundary I perceive is the juncture between ‘go’ and ‘under’. The most salient phonetic signal for the boundary is here for me the lengthening of ‘go’. I invite you to listen a couple of times to the entire clip and say, or write in IPA, the pronunciation of the word ‘go’ without reading further. What do you hear? Probably most of you would hear

something like [gəʊ]. Now select just the interval for this word and listen to it several times. Write down what you hear again without reading any further. In my writing, the actual pronunciation of 'go' would probably be [gə.] or [gɜ:], hence I do not hear a diphthong but a single schwa-like vowel. Did you have a similar difference between the first listening to the entire utterance and the second listening to the separate word? Observations like these, and your curiosity in trying to understand what is going on, are one of the most important goals of this book.

If I listen now to the interval 'go under', I still hear [gə] rather than [gəʊ] but if I go back to listening to the entire clip, I again hear [gəʊ]. Where does the perception of [gəʊ] come from then?

This is rather puzzling since we said that one of the salient markers of this prosodic break after 'go' is pre-boundary lengthening and I do perceive 'go' as slightly lengthened. Given that there is relatively enough time to realize the full diphthong [əʊ], why would the speaker neutralize it to a schwa-like monophthong? Again, without reading further try to form your own hypothesis and list some arguments that might support your hypothesis.

I have to be honest here and say that I do not know the 'answer'. In other words, I don't know if my hypothesis is what actually happened and further scientific analysis would be needed. Your hypothesis might be just as good, or even better than mine, and I am here returning to the goals of the book and the course in Phonetics: it is less important to memorize correct answers than to become curious about how things work, form hypotheses for why they are like that, and then try to find arguments and evidence for or against such hypotheses.

So my hypothesis is that the next word 'under' starts with vowel [ʌ]. This is a rather central vowel in both British and Australian varieties of English (slightly more back vowel in American English). Hence, for the speaker to realize the full diphthong, she would have to move the articulators away from the central schwa-like position and then go back to this position. Our own perception of [gəʊ] when listening to the entire utterance serves as evidence that the speaker does not have to do all this for us to actually perceive [gəʊ]. Hence, the basic tenet of Lindblom's Hyper- and Hypo- articulation theory (1990), that the speaker exerts only as much articulatory effort as is needed for good perception under current contextual conditions, seems to provide a natural understanding of situations like these.

To sum up the basic chunking of this utterance for the purposes of our fluent speech analysis, I propose the following: // okay / do you want me to go / under A-grade e- (/) grade eggs //. As before, we finish with the identification of pitch accents. The first chunk is the single word 'okay' and since every chunk has to have at least one accent, it is straightforward to



assume that ‘okay’ will be accented. We can confirm this by listening to the sound and we should perceive a clear prosodic prominence associated with the second syllable of ‘okay’ that is loud, includes pitch excursion, and is also slightly lengthened.

The second chunk ‘do you want me to go’ is much less straightforward and quite difficult. I myself am not sure. I think that a pitch accent on ‘go’ is quite reasonable to label. However, if, and where, the second accent is, I am not sure. Possibly due to the rhythmical consideration, I think that an accent early in the unit might be possible. The logical candidate would be the word ‘want’ since neither ‘do’ nor ‘you’ have full vowels. But I am not really convinced and thus rather opt for not specifying the second accent.

The third chunk with the speech error also poses great challenges. I think the prominence and pitch accent on ‘under’ is rather clear. The intended compound word ‘A-grade eggs’ would normally receive the primary stress on ‘eggs’ and a secondary stress on ‘A’, which translates to pitch accent being on the syllable with the lexical primary stress and thus ‘eggs’. I think this was also the intention of the speaker since despite a minor prominence on ‘A’, pitch continues smoothly over ‘A’ and the first mispronounced ‘grade’ with a natural accent on the following ‘eggs’. However, due to the error, ‘eggs’ intended with the pitch accent was never realized, and in the reparandum, that is in the repaired section of the error, we hear clearly that the speaker puts prominence on ‘grade’ rather than on ‘eggs’. The natural question, again, is why. And as above, I encourage you to form your own answer or hypothesis before continuing reading, and try to support it with some reasons or evidence in the utterance.

My own hypothesis is that the accent on the second, corrected instantiation of ‘grade’ is there precisely to highlight to the listener that this realization is the correct one and needs to be contrasted with the error produced before. And this intention of guiding the listener to the corrected word took precedence in this concrete situation over the intention to signal the default pragmatic structure when the ‘A-grade eggs’ is the new information that should be accented and this accent should be realized in a default way on ‘eggs’.

One last observation worth mentioning for Slovak native speakers is the realization of the final word ‘eggs’. Due to the interference from our native language, Slovaks have a tendency to produce [eks] rather than [egz] for this word. This is because in Slovak we neutralize voicing distinctions before pauses since, for example, when we say ‘Kupil som chlieb’ or ‘Vyhrál som súd’, the word ‘chlieb’ is realized with a [p] and the word ‘súd’ with a [t].<sup>2</sup> Note how this

---

<sup>2</sup> This is not entirely true since Slovak likely has so called incomplete voicing neutralization as has been observed in other languages with similar patterns like German, Polish, or Catalan (e.g. Warner et al. 2004). Because the

speaker clearly produces [g] since we can identify modular vocal pulses during the velar closure for [g]. This can be seen if you select the interval ‘eggs’ and then zoom to selection in Praat by pressing *Ctrl-N*. The velar closure starts at about 2.36s and ends around 2.43s and is fully voiced. However, the final fricative of ‘eggs’ is somewhat devoiced, which is common for pre-pausal alveolar fricatives. But voicing during the closure and lengthening of [e] are clear indications that the speaker produces the voiced [g]. In the productions of many Slovak speakers of English, contexts like this tend to be produced with a clear [k], and a rather short [e] giving the realization [eks]. If you also have this tendency, you can use this clip to practice changing this habit bit by bit.

Hence, our basic prosodic analysis for identifying chunks and accents can be finalized as follows: // oKAY / do you want me to GO / UNder A-gray e- (/) GRADE eggs //.

We proceed with the analysis of fluent speech now combining weak forms and connected speech aspect since both have been covered before and they naturally intertwine. The second chunk ‘do you want me to go’ promises an interesting situation since we have already said that there is only a single accent and thus most of the function words are unaccented and prone to be realized as weak forms and other aspects of connected speech are likely to occur. It might be instructive, despite the fact that you have likely listened to this clip several times already, to form a hypothesis of how this chunk will be produced by the native speaker, and in fact, how every single word is produced. Try to say out loud the words separately as well as together.

Only after you have said it yourself, listen to the individual words in Praat one by one. What do you notice? With the first word ‘do’ we might say that the vowel is completely elided and thus the weak form is not [də] but rather only [d]. However, if you play just the interval consisting of ‘do you’, you will hear that the speaker does not say the alveolar [d] but rather a more palatalized version [ɟ], which is somewhat similar to the Slovak consonant at the beginning of words like ‘dedo’ or ‘ďalej’. So what happened? Due to the elision of the vowel in ‘do’, the original  $V_f - C_i$  situation with regular linking has changed to  $C_f - C_i$  since [d] of ‘do’ and [j] of ‘you’ are now adjacent. We know that [d] is alveolar and [j] is palatal and the most natural strategy for native speakers is the regressive assimilation of the place of articulation in which the palatal quality of [j] is triggering the change of alveolar [d] into more palatal [ɟ]. The tempo is so fast that the two adjacent consonants might coalesce into a single [ɟ] or sometimes [dʒ].

---

observed incompleteness, the [p] in ‘chlieb’ is not exactly like [p] in ‘lep’ but depending on pragmatic context and requirements speakers might preserve some of the underlying voicing of ‘b’.

The vowel in ‘you’ is again extremely neutralized and realized with a short schwa that smoothly links to the following initial [w] of ‘want’. The juncture between ‘want’ and ‘me’ also promises some connected speech phenomena since we get a situation  $CC_f - C_i$  with three consonants in row and the alveolar [t] in the middle. We said before that this is the prototypical environment for consonantal elision. By listening to the ‘want me’ juncture and looking at the spectrogram it is clear that the ‘t’ of ‘want’ is not present. However, I do not think that this is a case of complete elision because the end of ‘want’ shows some vertical striations in the spectrogram and irregularity in the sound wave. Hence, as with this speaker before, the ‘t’ of ‘want’ is glottalized but even this simplification is not fully realized since the complete closure did not take place (there is no interval corresponding to silence during the closure). Rather, the glottal closure is only attempted but not fully realized.

Since there is no oral constriction for the ‘t’ of ‘want’, the original  $CC_f - C_i$  environment has changed to  $C_f - C_i$  with ‘n’ of ‘want’ and ‘m’ of ‘me’ being now adjacent. When listening to the juncture ‘want me’ we tend to perceive the disappearance of ‘n’ and the realization roughly corresponding to [wɒmɪ]. Although this is a plausible hypothesis, we have said already when discussing previous clips that a complete elision of the first consonant in  $C_f - C_i$  contexts is not a preferred strategy and at least temporally, the word-final consonant tends to be preserved and the place of articulation assimilated. Hence, with the assumed regressive assimilation of the alveolar ‘n’ of ‘want’ to become like the following bilabial ‘m’ of ‘me’, such a realization would be transcribed as [wɒmmɪ]. Do you think the spectrogram offers support for this hypothesis?

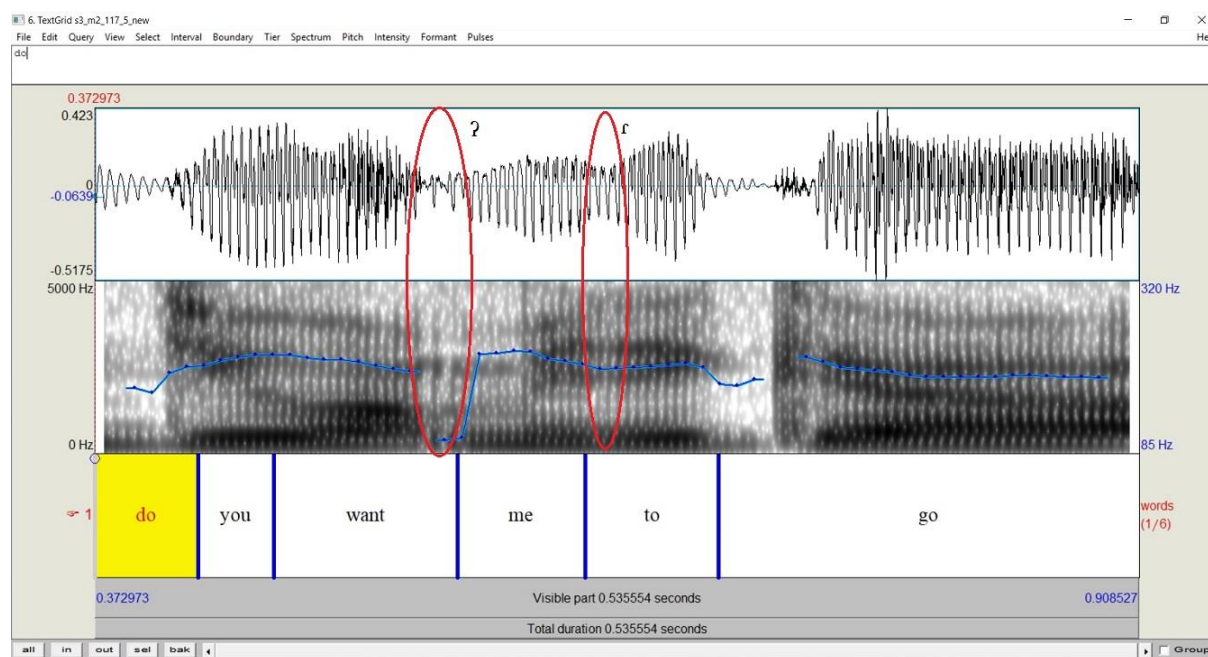
My opinion is that in this case, the nasal consonant [m] is not lengthened. Moreover, we clearly see in the signal the glottalization at the end of ‘want’. Hence, my analysis of this juncture would be [wɒʔmɪ].

Of course, ‘me’ is realized in its weak form with short and centralized [ɪ]. It is clear from our listening to the clip so far that the following ‘to’ is also produced as a weak form but you might be surprised to find out what the speaker actually said. When you listen to this word alone you perceive that the vowel is not really a schwa but something closer to [ɪ]. The question you might have predicted from the reading above is again why. What might be the reason for the speaker to say [tɪ] instead of the default [tə] in this context? As before, try to form a hypothesis and support for it before continuing reading further.

My answer to this puzzle is linked, again, to the context in which this word occurs. Note that what precedes ‘to’ is ‘me’ and that both words are weak forms and quite short. Hence, the tongue would need to move from the higher-front position of [ɪ] to the central position of [ə]

during the closure for [t]. Note how the closure for [t] is extremely brief, not even complete, and we might characterize this [t] as flapped since the vocal cords continue vibrating during this closure as can be seen by clear periodic waves during the entire closure. Therefore, a rather big mass of the tongue body needs to move extremely fast to be ready for [ə] when the flapped [ɾ] is released and in this fast tempo and very brief apical contact on the alveolar ridge the tongue body is lagging slightly and did not make it. This is a very nice evidence for the effect of V-to-V coarticulation for which we have evidence since Öhman's seminal study (1966) in which he described how vowels influence each other even across intervening consonants.

We might thus conclude the discussion of the connected speech aspects in this chunk by providing the narrow transcription of what I think happened in the mouth of the speaker when saying this utterance: [ɹəwəʔmɪrɪgə]. You might appreciate how this is very different not only from the dictionary canonical version with the strong forms [du: ju: wɒnt mi: tu: gəʊ] but also from the version using the weak forms [dʊ jə wɒnt mi tə gəʊ]. The screenshot from Praat with the ovals marking the two realizations of the alveolar t-closure as glottalization [ʔ] in 'want' and flapping in 'to' are in Figure 3 below.



*Figure 3* Sound wave, spectrogram and the textgrid in the three panels starting from the top as presented in Praat for the chunk 'do you want me to go'. The ovals show the realizations of the alveolar t-closure as glottalization [ʔ] in 'want' and flapping [ɾ] in 'to'.

The junctures in the rest of the utterance are straightforward. We might only note the absence of any consonantal intrusion in the  $V_f - V_i$  juncture of ‘the A’ and the smooth  $C_f - V_i$  transition in ‘grade eggs’.

### 3.3. Exercises with sample answers

In this last section I will discuss two additional utterances and here I encourage you to perform the entire analysis on your own first. You should have enough understanding of the concepts and approach after reading the previous sections of the book.

If you accept the challenge, it is essential that you actually write down your analysis. You might pretend this is a test. In my own experience, it is very different if I just try verbally, in my head only, to produce some analysis, explanations, or arguments, and if I actually sit down, and spend time writing it down. It helps putting the exercise in perspective, seeing connections between individual concepts, and most importantly, illuminates questions, gaps, and problems in the current understanding that you might raise with the instructor during the classes. The comparison with sample analyses also gives you some feedback of where you stand.

Hence, what you should do is the same as the approach discussing utterances so far in the book. Start with the prosodic analysis and try to identify the chunks, or units of speech that are prosodically defined. That is, there is some evidence for the prosodic coherence and fluency *within* the chunk and the opposite of prosodic coherence and fluency *across* the boundaries between the chunks. We have identified several characteristics to consider: pausing, lengthening,  $f_0$  movements, pitch reset, etc. Always try to articulate reasons for chunking along these or similar lines.

Once you have prosodic boundaries between the chunks, play each chunk separately and try to identify the words that are prosodically prominent and highlighted. You again rely on pitch movement, loudness, lengthening, and the realization of the vowel quality in the stressed syllable. You might check your proposed analysis with clapping hands on the accented syllables and see if that resembles the production of the chunk by the native speaker. There has to be at least one pitch accent per chunk.

With the understanding of the prosodic structure, consider the realization of function words as weak or strong forms, of course in connection to their accenting and chunk position, and try to go through every single juncture and discuss how the two words are linked together.

Do not forget to listen to the words separately as well as in the groups of 2-3 to get the feel for how they are linked.

Finally, try to imitate the individual chunks on your own getting as close to the native speaker production as possible, and then join the chunks into the entire utterance. Analyzing the clip first in the way described above should bring the awareness of the prosodic phenomena and ways leading to the flow and fluency of the native speakers that should ultimately help you adjust your own habits in speaking English to resemble those of native speakers some more. Naturally, since we are talking about ‘habits’, exercise like this needs to be repeated with different sound clips and utterances continuously to actually improve your rapport and flow of spontaneous fluent English.

### 3.3.1. s3\_m2\_73.8: and then it goes in between the two speech pens

In this utterance, the speaker explains to her interlocutor that the road on the map that they need to reconstruct, goes between two speech pens. ‘Speech pens’ is one of the objects on the map and the line in Figure 4 below depicts the road on the map of one of the subjects that needs to be drawn into the maps of the other subject without the road.



*Figure 4 The image of the speech pens object and the road going in between them that the interlocutors needed to reconstruct*

Have you written down your own analysis of this utterance? If so, below you will find one possible approach to analyzing the aspects of prosody and connected speech in this clip.

In terms of dividing this utterance into smaller chunks, I would analyze it as consisting of two chunks: ‘and then it goes’ and ‘in between the two speech pens’. The reasons and evidence for the presence of this boundary include the following aspects. First, the word ‘goes’ is perceivably somewhat lengthened although this is not major lengthening.

Second, the word ‘in’ starts with a brief interval of almost complete silence. This is not a silent pause since the duration is way too short, only about 25ms and we normally do not perceive silent pauses below 100ms. Moreover, there is evidence of glottalization just before the silent interval. These observations suggest that the speaker started the word ‘in’ with a glottal stop, which is one of the strong signals of disjuncture between two words. In other words, the evidence that the transition between the two words is not smooth and prosodically coherent.

Third, we also may see some evidence for pitch reset. In the first chunk pitch lowers to about 180Hz and on average moves around 200Hz. The second chunk, however, starts much higher and on average moves about 250Hz for ‘in between’. Nevertheless, please be aware that Praat’s pitch tracker in some settings might be misleading, since the start of the word ‘in’ is shown with a sharp fall in  $f_0$  but listening to it does not correspond to such pitch movement. This would be a case of pitch doubling or pitch halving (depending on the pitch range setting you have in Praat).

Fourth, listening to just the first chunk ‘and then it goes’ gives the sense of completion and coherence and one can imagine situations and contexts in which this chunk would be fine to stand alone without any further speech from the speaker.

Subsequently I try to determine if the two individual chunks can be further divided in a meaningful way. The answer is a clear ‘no’ for the first chunk. With regard to the second chunk, one might consider dividing the chunk after ‘between’. I do perceive some lengthening of the second syllable of ‘between’ but that would be the only evidence. In the absence of other evidence, the lengthening is attributed to the lengthening associated with pitch accent.

Hence, for all these reasons, my analysis of chunking in this utterance is: // and then it goes // in between the two speech pens //. Moving now to identifying prominent words, in the first chunk ‘goes’ is clearly accented. I am not sure if the pre-nuclear accent on ‘then’ can be supported. On the one hand, there is a very small hump in  $f_0$  and the rhythmical structure with clapping my hands seems fine with two accents. On the other hand, the vowel of ‘then’ is not really fully realized. I leave this open, I think both options (accent on ‘then’ or no accent) can be somewhat supported. This again is an example of the approach of the book, rather than looking for one single ‘correct’ answer, the much more important, and interesting, aspect is trying to weigh potential analyses against the evidence that we see and hear in the sound clip.

In the second chunk, both ‘in’ and ‘between’ are clearly prominent and receive a pitch accent. The challenge arises with ‘two speech pens’. Here I see as many as three possible analyses and all have some merit. The first one is to say that all these three lexical words are de-accented and thus not receiving any pitch accent. It is true that ‘two speech pens’ are less

loud and generally falling in pitch in the declarative sentence without any major departures from this declination. This analysis is also supported pragmatically by cheating slightly and looking at the preceding context. The interlocutor described that the speech pens are leaning to the right (as shown in Figure 4) and that the line goes through them. The current speaker then clarifies and seeks acknowledgment that the line goes ‘in between’ the speech pens rather than crosses them. The ‘two speech pens’ is the object both of the speakers are aware of, it is given in the current discourse, and thus it does not need to be accented to highlight its novelty, although many given elements are actually accented. The crucial new information that the current speaker wants to clarify is that the line is going ‘in between’ them rather than crossing them.

The second possible analysis would be to assign a pitch accent on each of the three words ‘two speech pens’. When you listen to just these three words, they all seem to have roughly equal prominence. Also, clapping three times during the production of ‘the three speech pens’ resembles reasonably well the production of the native speaker.

The last possibility would be to assign the pitch accent only on ‘speech’. This is the solution that sounds plausible when I listen to the entire chunk ‘in between the two speech pens’. In addition to this perception, the accent on ‘speech’ is reasonable for phonological considerations. The object ‘speech pens’ is a compound word. As with most compounds composed of two nouns, the primary lexical stress in this compound falls on the first noun ‘speech’. Hence, if ‘speech pens’ are produced as a compound and are pitch accented, the accent should fall on the lexically stressed syllable, thus on ‘speech’.

This is another example of how determining the prosodic structure is not an exact science with a single ‘correct’ analysis. It usually is a continuum and many times subjective considerations enter into the process of analysis. Some things are relatively clear, such as the presence of a boundary after ‘goes’, or the accent on ‘between’. I expect you to be able to identify these clear cases. However, some aspects are less clear and more challenging, such as accenting on ‘two speech pens’, and in these cases the discussion of possible alternatives with reasonable arguments for and against them is entirely sufficient since it shows your understanding of the aspects influencing the realization of the prosodic structure.

Regarding then the final prosodic analysis, I propose the third alternative above with the accent on ‘speech’: // and then it GOES // IN betWEEN the two SPEECH pens //.

Continuing now with the connected speech aspect, in the first chunk we have three monosyllabic function words in rapid succession following by the accented ‘goes’. The conjunction ‘and’ is clearly produced in the weak form since neither the vowel is [æ] nor the



final consonant [d] is fully realized. Listening to just this word alone, the closest IPA transcription to the native speaker realization is [ən]. Therefore, the final [d] of ‘and’ might be considered as elision since the most common environment for elision – alveolar ‘t’ or ‘d’ in between two other consonants – is met in the original ‘and then’.

After the elision takes place, the juncture between ‘and’ and ‘then’ is a case of  $C_f - C_i$  with [n] of ‘and’ preceding [ð] of ‘then’. This is again the environment prototypical for nasal place assimilation. The conjunction ‘and’ is not accented and the alveolar [n] is followed by a dental [ð]. The most natural strategy for joining these two words by the speaker is to employ regressive assimilation such that the alveolar [n] changes to the dental nasal [ɲ] to assimilate its place to agree with the following dental consonant.

It is true that listening to the juncture ‘and then’ does not give a clear perception of this nasal place assimilation. This is because the difference between the alveolar and the dental nasal is difficult to perceive and the two words are produced quite fast. For some speakers of English – mostly non-natives but also some native dialects – who produce ‘then’ with an alveolar stop [d] rather than dental fricative [ð], no assimilation would be needed since both sounds would be alveolar. However, this realization is not what most native speakers would do and thus if you want to get your pronunciation closer to those of native speakers, the dental realization of ‘then’ together with nasal place assimilation is the habit to try to acquire.

Since the juncture ‘then it’ puts together adjacent consonant and vowel, regular linking takes place. The next juncture, however, again presents an interesting situation. We have a  $C_f - C_i$  juncture involving the word-final plosive ‘t’ and the word-initial plosive ‘g’. With two adjacent plosives, my hypothesis even without listening to the file would be that the first plosive would not be released, and possibly assimilated in place to the second one, which is the default strategy employed by native speakers in junctures without a significant prosodic boundary. Recall that we said that the elision of the first consonant usually does not take place since the temporal ‘slot’ for the word-final consonant is preserved and the closure for the plosives is extended to give the perception of the preservation of the first plosive.

Now, listening to the juncture in question in this chunk and visually inspecting the spectrogram, I would slightly adjust my original hypothesis. I still think the first plosive is not elided, that it is unreleased, and possibly assimilated in place to the following plosive, but in addition to this, I also see the evidence of glottalization in the form of a single vertical striation just before the closure for ‘t’ around 0.315s timestamp.

Also note how the disagreement in voicing is resolved. In Slovak, our natural tendency would be to voice the word-final consonant to make it agree with the voicing of the word-initial

one. Imagine the juncture ‘kryt gaštanu’ to resemble the consonants in the English juncture ‘it goes’. In Slovak, we would assimilate the word-final ‘t’ of ‘kryt’ to sound like [d] so that it would agree in voicing with the voiced word-initial ‘g’ of ‘gaštanu’. You should try this and observe and introspect your own realization of ‘kryt gaštanu’. In English, however, this is not what native speakers usually do. Our clip also shows that there is no voicing in the word-final consonant of ‘it’, and, importantly, the word-initial ‘g’ of ‘goes’ is rather devoiced since there is no voicing preceding the release. The perception of ‘g’ comes from the absence of aspiration but phonetically, this word-initial ‘g’ is realized as [k]. This is yet another perfect example of the importance of minor phonetic differences and how they participate in non-native Slovak-like accent when we speak English.

It is true that Praat shows some energy in the middle of the closure around 0.35s and one might assume that this is actually the release of the first consonant. However, there is really low energy and without any burst typical of plosive release phase and for this reason listening to this we do not perceive any release. Hence, I would disregard this in my analysis.

To summarize the first chunk then, my narrow transcription of connected speech would be: [əŋðenɪʔkʰkəʊz]. As usual, this is radically different from what we might expect based on the dictionary form transcriptions.

The second chunk is not as rich in connected speech phenomena as the first one. This is partly due to the makeup of consonants and vowels at word boundaries but more importantly, to greater prominence and the number of lexical words. Recall that in our discussion of accenting, we were sure about the accents on ‘in’ and ‘between’ and contemplated accenting on the remaining three lexical words.

Let’s start with the first juncture ‘in between’. Without listening we might form a hypothesis that nasal place assimilation will take place. This would be the most natural realization since we have a  $C_f - C_i$  situation with word-final alveolar ‘n’ and word-initial bilabial ‘b’: [ɪmbɪtʰwi:n]. However, when listening to the juncture in Praat I clearly perceive the word-final [n] rather than [m] and I must, as you should as well, ask myself why. The only plausible hypothesis is that this hyper-articulation (and the absence of hypo-articulation involving nasal place assimilation) is related to the prosodic prominence on both words. Recall our previous discussion of the pragmatic aspects and preceding speech in which we said that ‘in between’ is meant to ascertain that the line does not cross through the speech pens but goes in between them. Due to this pragmatic focus, realized as prosodic prominence and hyper-articulation, the otherwise natural nasal place assimilation does not take place.

The potential case of another nasal place assimilation in the following juncture ‘between the’ is much less clear. The definite article is clearly realized in its weak form since it is rather short and the vowel is schwa. As with the discussion of the ‘and then’ juncture in the previous chunk, I have to admit that I am not sure if nasal place assimilation took place. I can only speculate what and why happened, and this is precisely what I encourage, and expect, you to do. If the assimilation took place and the juncture is realized as [bit’wi:ɲðə], the support would be the presence of the weak form and the naturalness and ubiquity of nasal place assimilation in this context.

However, an alternative hypothesis is the realization [bit’wi:nnə]. In support for this analysis, we do not see any major changes during the closure for the consonants roughly between 1.374 and 1.447 seconds in the file neither in the sound wave nor in the spectrogram. Hence, it looks more like a single sound than a transition between a nasal and a fricative. Moreover, note that the following word ‘two’ starts with an alveolar ‘t’ and thus it is quite economical and requires less effort to keep the tongue in the alveolar place of articulation for the ‘between the’ juncture since it will need to be at this place after an extremely short vowel of the weak form ‘the’. For these two reasons I would analyze this transition as [bit’wi:nnə].

The following two junctures involve adjacent consonants and vowels hence the linking is unproblematic. The final juncture ‘speech pens’ is a  $C_f - C_i$  juncture but it involves an affricate [tʃ]. Both fricatives and affricates are sounds quite resistant to elisions and assimilations since they are perceptually prominent and changes to their articulation greatly affect their acoustic qualities. Moreover, there is no bilabial affricate in English and so regressive place assimilation targeting [tʃ] triggered by [p] would be extremely weird.

This concludes the analysis of the clip s3\_m2\_73.8.wav and the utterance it contains. How does it compare to your analysis?

### **3.3.2. s4\_m2\_200.2: they’re quite close together so where you were in the middle of the fantasy god you just go straight down**

In this clip, the speaker responds to the question from her interlocutor about how to get from one object on the map (the fantasy god) to the next object (the football pitch), and more specifically if she should just go straight down. The speaker instructs that “they’re quite close together so where you were in the middle of the fantasy god you just go straight down” and by ‘they’ she means the fantasy god and the football pitch.



*Figure 5 The section of the map with the images of the fantasy God and football pitch and the road that the interlocutors needed to reconstruct*

Please, as before, listen carefully to the clip and visually analyze it in Praat. Then write down your own analysis of this utterance covering all the basics of chunking, accenting, and fluent speech phenomena such as weak forms and juncture linking. Only when you have your analysis ready and written up, continue reading for one possible approach to analyzing the aspects of prosody and connected speech in this clip.

Regarding chunking, this is a relatively long utterance of a little over 4 seconds, and the longest of the clips analyzed so far in the book. Despite this, there is no silent pause that would facilitate the first step in chunking, which previews that the prosodic analysis of chunking and accenting will be rather complex and probably the most challenging of the clips analyzed so far.

Although the clip does not contain silent pauses, my perception is that it can be further divided into three smaller chunks. The first boundary is relatively clear to me and I would place it before ‘so’. The second boundary is less clear and I feel it is around the word ‘yeah’, but I am not sure if it is before or after this word. Let me discuss these two boundaries in turn.

Regarding the first boundary, it is actually quite surprising that there is no silent pause since I feel there should be one and I feel I hear it there. Do you have a similar perception? This is further evidence of the intricacies and interesting aspects in the relationship between production and perception of speech. But let’s get back to the analysis and by now the familiar question WHY: Why do I perceive such a strong break? Or in other words, which phonetic features are responsible for this perception?

The first aspect that cannot be disregarded is the syntactic structure and its related discourse considerations. It is plainly clear that there is a major syntactic boundary before ‘so’ and in the textual transcript of this utterance, either a full stop or a comma could very well appear in this position. Hence, ‘so’ functions in this context as a discourse marker separating, and linking at the same time, the material before it with the subsequent thought that follows it. The first reason for perceiving the break thus arises from the syntactic and discourse structures of the utterance.

The strongest phonetic cue of this boundary is, in my mind, related to the fundamental frequency while temporal cues such as pre-boundary lengthening are not prominent in this boundary realization. The pitch range in the interval ‘they’re quite close to-’ is rather steady at around 200Hz and then falls on the last two syllables ‘-gether’ to about 170Hz. In the next interval ‘so where you’ it is up again to over 200Hz. This f0 fall before ‘so’ associated with the pitch reset after ‘so’ are the phonetic strategies the speaker employed in this context to signal the prosodic structure, which, in this case, is in agreement with, and complements, the syntactic structure.

The second boundary around the word ‘yeah’ also prosodically supports the discourse structure of this utterance. It is obvious that ‘yeah’ is another discourse marker similar to ‘so’. However, the function of ‘yeah’ is a little more complex here because it not only links the information provided preceding ‘yeah’ with the concluding definite instruction to go straight down, but it also links this final instruction to the original question from her interlocutor. Recall from the description above in Figure 5 that the interlocutor actually asked specifically if she should just go straight down. And thus the final instruction ‘you just go straight down’ is linked to this original question with the affirmative function of ‘yeah’. Note that this ‘yeah’ helps the orientation of the interlocutors in the discourse and the utterance would be syntactically perfectly formed without it.

Now back to the question of whether ‘yeah’ belongs to the second or third chunk of this utterance. My primary reason for favoring the latter analysis is that there is slight, yet perceivable, lengthening of the word ‘God’. This signals that a prosodic boundary is aligned with the right edge of this word. Moreover, try listening to the second chunk as both including and excluding the word ‘yeah’, and consider which option gives a better impression of completeness and prosodic coherence. For me it is the option with ‘yeah’ excluded. Finally, connected speech aspects might help as well. If the boundary between ‘god’ and ‘yeah’ was a weak one within a chunk, we might expect modifications of the consonantal realization. In this situation we would have [d] in the word-final position and [j] in the word-initial position with

very common and natural realization as [dʒ]. However, no such modification to the realizations of the consonants can be observed and we hear a clear [d] and [j] produced.

Unfortunately, the fundamental frequency contour does not reveal any other marking of this break so we have to rely on lengthening and connected speech aspects only. Note how this is different from the first break of this utterance with which we identified f0 marking and not lengthening. This shows further the complexity, and optionality in employing these cues for boundary marking in spontaneous speech on the one hand, but on the other hand serves as the evidence for the robustness of the cognitive system underlying the relationship between the pragmatic meanings and prosodic structure shared by the interlocutors.

So to move to the step of identifying prosodic prominence, we start with the chunks determined above: // they're quite close together // so where you were in the middle of the fantasy god / you just go straight down //. In the first chunk 'they're quite close together', I feel that the pitch accents on 'quite' and 'together' are unproblematic and both words are produced as prominent. The tricky decision involves the word 'close'. I think that one might consider a pitch accent there but I prefer here the analysis with just two pitch accents mostly for greater simplicity and greater reliability of analyses in which only the salient accents are marked.

The second chunk 'so where you were in the middle of the fantasy god' is quite difficult to label for pitch accents. The f0 contour is mostly quite flat apart from a peak during the words 'were you'. I again start with the clearest accents and then move to less clear cases. I am reasonably sure that the words 'where' and 'God' are accented. There is a pronounced f0 rising on the first word and somewhat less clear but still perceivable f0 movement starting on the last syllable of 'fantasy' and continuing to 'god'. The content words in between these two accents are much less clear. My analysis would include pitch accents also on words 'were' and 'fantasy'. However, I cannot offer a concise argumentation as to the reasons supporting this perception that I have. I might speculate that in the absence of salient f0 movements, I also rely partly on loudness. I do perceive a pitch target on 'were' that is a step down from the one on 'where', and in listening to 'fantasy god' I cannot clearly determine which of them is more prominent. For all these arguably very subjective reasons, my suggestion would be: 'so WHERE you WERE in the middle of the FANtasy GOD'. The accent on 'middle' that some might also consider is not included for the same reasons as the accent on 'close' was not included in the previous chunk.

The final chunk 'yeah you just go straight down' presents similar challenges as the previous two chunks. I perceive that the accents on 'yeah' and 'down' are quite salient and unproblematic. I was also considering the accent on 'straight', which would fit in terms of the

rhythm and my hand-clapping help yields relatively similar results when ‘straight’ is accented and not accented. Because of my uncertainty, I employ the simplicity/reliability rule as in the previous two chunks and opt for not assigning a pitch accent on this word.

The final result of the prosodic analysis of this utterance is thus as follows: // they’re QUITE close toGEther // so WHERE you WERE in the middle of the FANtasy GOD / YEAH you just go straight DOWN //.

Now we can proceed to the examination of the connected speech aspects such as weak forms and the realization of junctures between words. In the first chunk, the only candidate for a weak form is the contracted ‘they’re’ as we see already from the text transcription that I employed. Listening to this word separately we hear that not only the original two syllables of ‘they are’ are shortened to a single syllable, but even this syllable is extremely neutralized so that it could even be transcribed as [ðe]. The diphthong [eə] that we could expect here is thus shortened and simplified.

The link of this contracted structure to the following word ‘quite’ is uneventful since we have a situation of a vowel and a consonant at the juncture that are easily linked. However, the next juncture ‘quite close’ presents an interesting situation. We have a  $C_f - C_iC$  situation involving an alveolar stop in the  $C_f$  position. Even without listening to the file we can already speculate that some modifications will take place. This is because we have the least salient of the stops, the alveolar one, in the prosodically weak  $C_f$  position (recall that the coda position is much weaker than the onset position for word comprehension). I did not keep the count but the majority of the elision and assimilation changes affecting consonants in the utterances we have analyzed so far involved a ‘t’ or a ‘d’ in the word-final position. Nevertheless, we have to be aware that the word ‘quite’ is accented and thus the possibility of hyper-articulation and the full realization cannot be discarded.

It is quite useful to think of possible alternatives for the resolution of this juncture even before listening to it. This is similar to the advice I usually give before trying to solve a problem or puzzle: if you have enough time, try to have an educated guess about the answer before you start solving the problem. This applies to, for example, listening comprehension exercises in tests like TOEFL or Cambridge First Certificate or regular math problems involving numbers. It is easier to have guesses and alternatives to select from than search for the answer in the dark. Of course, with this approach you have to be prepared to reject your guesses or alternatives if the evidence does not support it.

So the alternatives we should be aware of include:

- a) the full realization [kwat kləʊs]
- b) unreleasing of 't' [kwat̚ kləʊs]
- c) glottalization of 't' [kwaiʔ kləʊs]
- d) place assimilation, most likely with unreleasing and/or glottalization [kwaik̚ kləʊs] or [kwaiʔk̚ kləʊs]
- e) complete elision of word-final 't' [kwai kləʊs]

So what strategy does the speaker employ given these alternatives? I would select the option c). The spectrogram and the wave forms show clear evidence for glottalization in the form of irregularities in pulses and vertical striations similar to those illustrated in Figure 3 at the end of Section 3.2.2. I do not have any grounds for deciding if in addition to the glottal stop the speaker also uses the tongue to produce either alveolar or velar constriction for the final 't' of 'quite'. Hence, again, in the absence of salient evidence, either visual or auditory, we opt for the simpler explanation.

The final juncture of the first chunk 'close together' involves a  $C_f - C_i$  context. One might expect that some alteration from the canonical pronunciation of the consonants at the juncture might take place. However, note that we have an 's' in the  $C_f$  position and we have already said that fricatives, and especially sibilants are extremely resistant to place or manner changes. Moreover, both  $C_f$  and  $C_i$  consonants share the same place of articulation, so the tongue is required to move very slightly raising the middle of the tongue blade from the narrowing between the tongue and the alveolar ridge in 's' to the complete closure for 't'.

In the second chunk, there are several candidates for weak forms: 'you', 'were', 'the', and 'of'. The verb 'were', however, does not function here as an auxiliary but as a full lexical verb and thus it would be surprising if it was neutralized to a schwa of a weak form. In fact, listening to 'were' separately it is not extremely short and it is roughly twice as long as the weak form 'you' and thus I would transcribe this as [jə wɜː]. It is also clear from listening to the individual words that both instances of the definite article are fully neutralized into weak forms [ðə]. Finally, the preposition 'of' can also be considered as weak form since it is not accented, although I do perceive that the vowel is not a 'clear' schwa and some traces of the [ɒ] vowel could be heard.

Most of the junctures in the second chunk 'so where you were in the middle of the fantasy god' involve a  $V_f - C_i$  or  $C_f - V_i$  context and thus they are smoothly joined via regular linking common for transitions between consonants and vowels. Three junctures deserve further scrutiny and discussion: 'were in', 'in the', and 'of the'. The first one involves two adjacent



vowels while the remaining two involve adjacent consonants across the word boundary. The juncture ‘were in’ presents a prototypical environment for joining the two words with the help of ‘linking r’. This consonant is present in spelling but in non-rhotic dialects of English not pronounced in the syllable coda position unless it is in the word-final position, the following word starts with a vowel, and, importantly, the strength of the boundary between the two words is weak. All of these requirements are present in the juncture ‘were in’. The auditory check of the production of this juncture supports the presence of linking r.

The second juncture ‘in the’ presents the alternative of nasal place assimilation of which we have already see several examples in the utterances analyzed above. But let us consider the plausible alternatives before we actually listen closely to the sound file:

- a) the full realization [ɪn ðə]
- b) regressive nasal place assimilation with the dental n [ɪn̥ðə]
- c) regressive nasal place assimilation and progressive manner of articulation assimilation [ɪnn̥ə]
- d) alveolar realization [ɪnə] or [ɪnnə]

With these alternatives in mind we can go to the sound file in Praat and listen to the juncture several times. I think the first alternative a) can be dismissed right away since I hear no fricative sound and in the spectrogram there is no change during the interval of the closure (approximately 1.83s – 1.88s) suggesting that neither the place nor the manner of articulation is changing. This consideration also yields alternative b) [ɪn̥ðə] as highly unlikely since the manner of articulation would have to change from the nasal stop to the oral fricative even though both share the same place of articulation. Hence, we are left with two decisions to make: first, is the place of articulation of the nasal alveolar or dental, and second, can we treat this constriction as a single consonant and thus a case of elision or are the temporal slots for two consonants preserved. Both of these decisions are difficult and belong among not clear cases with no ‘correct’ solution. I do not hear clearly the place of articulation and neither the duration but my subjective impression would prefer slightly the realization c) [ɪnn̥ə]. I know from other realizations of ‘the’ from this speaker that she uses [ð] and not the sub-standard alveolar realization [d]. Hence, I have no reason to suspect that the dental place of articulation would not be preserved. Regarding the temporal aspects, these two function words are extremely short and despite this the closure is relatively clear with sufficient length and thus I opted for the preservation of both consonantal slots.

The narrow transcription for the second chunk ‘so where you were in the middle of the fantasy god’ would thus be [səʊweɪəwɜɪɪnn̥əmɪd|əvðəfænəsɪɡɒd]. Note also the simplification

of the 'nt' cluster to a single 'n' in the word 'fantasy', which is a rather common pattern occurring also in words like 'dentist' or 'interesting'.

After finishing the discussion of the second chunk we may switch to the final chunk of the utterance 'yeah you just go straight down'. We have a clear weak form of 'you' produced as extremely shortened [jə]. Regarding junctures, there are two cases of  $C_f - C_i$  and both will be discussed in turn.

First, there is the 'just go' juncture. This presents a case of a three-consonantal cluster [stg] across a word boundary with the weak alveolar 't' in the weak coda position of the first word. As we mentioned several times already, this is favorable environment for t-deletion and thus the case of elision. Both the auditory check and the visual inspection in Praat support this analysis. Note also that the word-initial 'g' is again realized as completely devoiced, possibly due to the voice assimilation from the preceding 's' and thus the absence of aspiration is the major signature for the voiced character of this consonant. The narrow transcription of this juncture would thus be [dʒʌskə].

Second, there is the final 'straight down' juncture. My analysis includes the glottalization of the word-final 't' accompanied by not releasing the closure: [streɪʔ]. Notice again that the word-initial 'd' is at least partially devoiced.

This concludes the discussion of the analysis of prosody and connected aspects of the file s4\_m2\_200.2 as well as the entire Section 3 in which sample analyses of individual clips were presented.

## 4. Conclusion

First, let me congratulate those who finished reading and engaged in exploring the sound clips, finding patterns and increasing their awareness and understanding of these patterns in spontaneous speech. You have a valuable, and I also hope interesting, experience in which you developed your analytical skills that are an integral part of the curriculum in the Phonetics course.

I also hope that the material, the approach to the analyses, and step-by-step descriptions accompanied with your hands-on exploration of both auditory and visual characteristics enabled in Praat was useful not only for the course in Phonetics but also for continually improving your non-native English pronunciation. Moreover, I firmly believe that the work with data, identifying patterns, potential alternatives, reasoning about them, discovering the relationships between seemingly separate phenomena or actions, all of which is included in this approach to analyzing speech, are skills that higher education should promote and that the job market appreciates.

One aspect that was mentioned but I would like to stress it again here is the interdependency of all the aspects of spontaneous speaking that were discussed and analyzed here. It may seem that when we analyzed chunking, accenting, weak forms, and junctures separately that they form a separate system of patterns. However, just like context is essential and crucial in the realization of individual sounds, these aspects of spontaneous speaking are firmly couched in the situational context both in the broad sense in understanding the interlocutors, their roles, intentions, and so on, as well as in the narrow sense understanding how chunking affects juncture realization, or accenting the weak forms.

Finally, I would really value and appreciate your comments, questions, or concerns that you might have related to the commentaries in this book. Learning is a two-way process and just as I have learned from sitting down and putting these commentaries together, and just as I hope this was the effort that will help you learn and gain skills and understanding, feedback from you and subsequent discussion of that feedback has the potential to move both me and you even further.

## References

- Beňuš, Š., Bírová, J., Chebenová V., Molnárová, A. (2010). *Výučba výslovnosti cudzích jazykov pomocou porovnávacej akustickej analýzy*. Nitra: UKF.
- Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in English. In: Goldstein, L., Whalen, D. H., Best, C. T. (eds.) *Laboratory Phonology 8: Varieties of Phonological Competence*. Walter De Gruyter, New York, pp. 519-548.
- Fougeron, C., Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728-3740.
- Graham, C. (2001). *Jazz chants old and new*. New York, London: OUP.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech Production and Perception*. Kluwer Academic Publishers, Dordrecht, pp. 403-439.
- Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39: 151-168.
- Nakatani, Ch., Hirschberg, J. (1993). A speech-first model for repair detection and correction. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pp. 46–53.
- Pierrehumbert, J., Talkin, D. (1992). Lenition of /h/ and glottal stop. In Docherty, G., Ladd, D. R. (eds.) *Papers in Laboratory Phonology 2*, Cambridge University Press, Cambridge UK, pp. 90-117.
- Roach, P. (2000). *English Phonetic and Phonology: A Practical Course* (3<sup>rd</sup> ed.). Cambridge University Press, Cambridge UK.
- Warner, N., Jongman, A., Sereno, J., Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32: 251-276.
- Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics* 38(3): 329-336.

Štefan Beňuš  
Practice material for English Phonetics 1:  
Basic prosody and connected speech  
Nitra: Univerzita Konštantína Filozofa v Nitre  
2018