

**Practice material for English Phonetics 2:
Comprehensive analysis of spontaneous speech**

Štefan Beňuš

Contents

1. Introduction.....	3
2. Review of Slovak/(Hungarian) native speech interference in their non-native (L2) spoken English.....	6
2.1. Differences arising from segmental inventories	6
2.1.1. Consonants	6
2.1.2. Vowels.....	8
2.2. Differences arising from allophonic variation.....	10
2.3. Prosody	13
3. Overview of simplified ToBI for annotating prosody	16
3.1. What aspect of intonation and why ToBI is useful	16
3.2. Crash course to simplified ToBI.....	18
3.3. Sample of compositional intonational meanings.....	22
4. Commentaries and analyses of a dialogue	24
4.1. Dialogue description.....	24
4.1.1. Text Transcript	25
4.1.2. IPA transcription	27
4.2. Turn-by-turn commentary	28
5. Conclusion	61
References	62

1. Introduction

In this material we build on the approach and aspects covered in ‘*Practice material for English Phonetics 1: Basic prosody and connected speech*’ and assume that you have engaged with the material analyzed in Part 1 and gained fair understanding of the aspect of spontaneous speech covered there. The hallmark of Part 1 was to increase awareness of basic prosody, weak forms, and connected speech aspects via engaging with real speech excerpts in Praat, and analyze speech patterns with the help of the guidance in sample commentaries. There are several aspects of Part 2 that will be kept from Part 1, and several that will be expanded or different.

Let us start with the ones that stay. The basic approach remains identical. There will be sound material provided in the form of the sound files and Praat Textgrid files and you will be expected to engage with the material through Praat and attempt to analyze what speakers are doing and why they are doing it. I will also provide guidance in the form of commentaries but the ultimate aim is for you to understand the speaking habits of the native speakers as the first step toward making your own pronunciation more native-like.

We will also discuss basic prosody, weak forms, and connected speech aspects, which is the material covered in Part 1. This additional analysis should strengthen the understanding gained in Part 1. You will see that new contexts and material opens up both patterns already discussed in Part 1 but also habits that were not covered there. These aspects might also serve as a test and feedback for your proficiency developed in Part 1. If these aspect of fluent speaking go smooth for you also in this new material, you know you are on the right track and the effort and time spent on Part 1 was well spent.

The last thing mentioned here as staying identical from Part 1 is my firm belief that the type of learning promoted in this approach to analyzing speech offers the development of skills and competences valuable in essentially any job or position and those expected from university educated people in the 21st century. Hence, while the content of the work covers English Phonetics, the skills are applicable very broadly and generally. I am pre-emptively answering here the (rhetorical) question of many students and the one I expect also from many of you: Why do I need this? How am I going to use this in my future life? These vital skills that you can gain by engaging in the type of analysis promoted in this book include, but are not limited to, finding patterns in data, applying understanding to novel situations, forming hypotheses and potential alternatives and checking them with data (evidence), using computer software for supporting your conclusions, discovering the relationships between seemingly separate phenomena or actions and hopefully many others.

Now, let us preview those aspects of this Part 2 material that are different from Part 1. First, the discussion of the segmental aspects will be included primarily with the goal of increasing awareness of the habits arising from the mother tongue interference when Slovaks or Hungarians speak English in both segmental and suprasegmental domains. Hence, some aspects such as aspiration might be highlighted in the segmental domain, and these were not typically covered in Part 1. Moreover, I will spend more time of pointing out potential challenges for non-native speakers of English. It is hoped that engaging in trying to imitate these challenging patterns, those that are somewhat different from the native patterns of Slovak or Hungarian, you will get further opportunities for improving your English pronunciation.

The second major expansion will concern the coverage of sentence prosody. In Part 1 we concentrated on identifying basic prosodic structure in the sense of detecting prosodic boundaries, and prominent, pitch accented words. In the current material we will expand on this basic analysis by discussing also the type, or identity, of the tonal targets associated with both boundaries and pitch accents. In other words, after identifying the presence of boundaries and accents, we will also discuss whether the boundary is low, high, or can be characterized as a so called continuation rise, and similarly if the accent are essentially low, high, or combines these two targets.

We will rely on the wide-spread system for annotating prosody entitled Tones and Break Indices (ToBI) that is based on seminal work by J. Pierrehumbert (1980), elaborated in terms of communicative meanings of these tonal targets in Pierrehumbert & Hirschberg (1990). The descriptions of this framework is available in several review papers, but most importantly as a self-study material developed by A. Brugos, N. Veilleux and S. Shattuck-Hufnagel for a course at MIT “Transcribing Prosodic Structure of Spoken Utterances with ToBI” available through MIT’s open courseware initiative (<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm>). Section 3 will review the basic typology of Tones that we will use as well as their possible communicative meanings.

The main reason for this elaboration of the basic prosodic analysis covered in Part one is that the identity of tonal targets can signal various structural and grammatical functions of intonation such as contrastive focus or questioning. Hence, all three major prosodic distinctions – presence vs. absence of a prosodic boundary, presence vs. absence of a pitch accent, and the type of the tonal target – that cue structural meanings, will be covered. Of course, intonation plays many other functions, especially in terms of cuing emotions and attitudes, and the phonetic features participating in these domains include fundamental frequency but also include loudness, speech tempo, or voice quality that are not covered in ToBI.

The third major difference of Part 2 in comparison with Part 1 will be that only a single dialogue will be analyzed in this part. Hence, rather than exploring several non-continuous clips, albeit from a single dialogue as we did in Part 1, we will now investigate turn-by-turn speech of two interlocutors sequentially. This opens new possibilities for discovering the interactional and turn-management aspects of fluent spoken dialogues such as potential interruptions, or ways of signaling turn-yielding. Moreover, the extension compared with Part 1 is also in the mode of the dialogue. While in Part 1 the two speakers were engaged in solving a particular task collaboratively, which limited their domain of discussion in some sense, the dialogue in Part 2 is less limited since the interlocutors are not solving a task but chatting on a radio show about a particular topic.

Finally, I would like to urge you here to read again Section 1.3 of Part 1, in which I discuss the approach of this series of books. I emphasize there that all the patterns discussed in the commentaries, and appearing isolated from the rest of the speech for presentational reasons, should be understood as being intimately and necessarily linked and intertwined with all the other aspects of fluent speech. Furthermore, I stress that the approach is descriptive rather than prescriptive in that we are mainly interested in describing the patterns in the speaker's mouth and her mind, and through studying these patterns in context, we hope to understand better the underlying mechanisms responsible for the pattern we observe in speech. The book is not concerned very much with providing 'correct' solutions or ready-made answers for you to memorize. Rather, it should be a guideline for your own exploration of speech. And finally, I urge you to always bear in mind the social context, the pragmatic goals of the speakers, and the relevant aspects of the situation. The *WAY* we say things (in addition to *WHAT* we say) is greatly affected by these contextual considerations both in the narrow sense of what was said before and planned to say next, and also in the broad sense of forming conceptual pacts (e.g. Clark and Brennan 1996) or negotiating social identities and bonds (Giles et al. 1991) through communication.

The book proceeds as follows. In section 2 we will review the major differences between the inventory of Slovak/Hungarian vowels and consonants with the aim to identifying potential areas responsible for the non-native accent in your English, as well as reviewing the major allophonic patterns and other relevant suprasegmental phenomena. In section 3, we will discuss ToBI and the major communicative meanings of some tonal targets. Section 4 will be the core of the book and present the commentary of the major section of the dialogue.

2. Review of Slovak/(Hungarian) native speech interference in their non-native (L2) spoken English

In this section I will review both segmental and suprasegmental differences between Slovak and English that serve as most common sources for mother tongue (L1) interference into speaking L2 English. My notes will concern Slovak mostly but some observations will be made also regarding Hungarian from my experience of native Hungarians in Slovakia. In no way does this brief review provide a complete picture of all aspects of this interference and you are encouraged to consult additional sources such as Král'ová (2011) or Beňuš (2010). The review is meant to point your attention to most common problems Slovaks face when they try to minimize the L2 interference and the need for working on changing your habits if these are problematic to you as well. We will not discuss strategies or exercises here for developing and strengthening the native-like habits.

2.1. Differences arising from segmental inventories

2.1.1. Consonants

There are five major challenges for Slovaks/(Hungarians) speaking English arising from the differences in the consonantal inventories and differences in the actual realization: dental fricatives [ð,θ], velar nasal [ŋ], labiodental approximant [w], glottal fricative [h], and alveolar consonants. We will discuss them briefly in turn.

First, dental fricatives [ð,θ] realizing the grapheme structure 'th' in English are notoriously known problem of Slovak and Hungarian speakers since these sounds do not exist in their inventories. They are commonly replaced with consonants sharing only a subset of features with the target dental fricatives. Either they are fricatives that are not dental but labiodental [f, v] or alveolar [s, z], or stops with the dental-alveolar place of articulation [t, d]. Quite commonly, the realization is a mixture of these influences. The replacement of [ð] by a (weak) [d] is the least obtrusive strategy, employed even by native speakers in some dialects of English. However, native speakers are sensitive to these realizations and notice the non-native accent easily.

In Praat, the fricative vs. plosive realization can be sometimes identified since the plosives typically have a complete closure followed by a burst of energy. The complete closure is in the waveform recognized as almost a flat line (or low-energy periodicity with voiced plosives) and in the spectrogram as a very light interval. This contrast with the following burst that usually has much higher energy in higher frequencies. Unfortunately, the difference between the dental and alveolar

fricatives is very difficult to see by inspecting waveforms and spectrograms, only that the alveolars tend to have the energy of the friction in higher frequencies than the dentals.

Second, the velar nasal [ŋ] poses problems in the word-final positions of the grapheme structure ‘ng’ since neither Slovak nor Hungarian have this sound in this position. Word-medially, the problems are less extreme since in most cases the pronunciation is [ŋg] perfectly resembling Slovak words like ‘tango’ or ‘hangár’ and only some morphological boundaries between stem and affixes like ‘sing-er’ are realized with [ŋ] and not [ŋg]. In word-final positions, the non-native accent in realizing words like ‘sing’ as [sɪŋg] are exacerbated by the voicing neutralization in Slovak in which voiced obstruents (stops and fricatives) are devoiced when they precede a pause or another voiceless consonant. Hence, many Slovaks say [sɪŋk] rather than [sɪŋ] when they mean ‘sing’. Alternatively, the velar nasal is replaced with the alveolar nasal [n] in these positions. This is less salient and noticeable than [ŋk], and in some dialects native speakers may use this strategy. Nevertheless, similarly to the plosive realization [d] in place of dental fricatives [ð] above, [n] replacing [ŋ] is the least obtrusive, yet noticeably non-standard strategy.

In Praat, the realization [ŋk] instead of [ŋ] can be seen in the period of almost flat lie in the sound wave and associated very light region in the spectrogram followed by comparably higher energy of the closure release while the realization [n] instead of [ŋ] is virtually indistinguishable at our level of expertise.

Third, a common problem for Slovak speaking English is labio-velar approximant [w] in the syllable onset position that might be replaced by labio-dental [v], but more commonly vice versa when [v] is replaced by [w]. At first sight, this is a paradoxical problem since the Slovak inventory does not include [w] in the onset position of native vocabulary and in foreign words spelled with ‘w’, this is normally replaced by [v]. Hence, native language interference might provide some grounds for replacing [w] with [v], for example saying ‘well’ as [vel], but replacing [v] with [w] as in [werɪ] for ‘very’ cannot be solely attributed to the differences in inventories. The paradox and apparent non-sensicality of producing a spelled ‘v’ as [w] with Slovaks speaking English is further strengthened by the fact that this never happens with native speakers of neither English nor Slovak.

My own speculation, not supported in any rigorous research, is that this [v] → [w] replacement is a case of over-application and over-generalization error. When starting with English, the new learners must acquire the understanding that in addition to [v] that they know from Slovak, English also has [w] in the syllable onset position. This pattern is then generalized and overapplies also in cases in which [v] is required and present in spelling. Maybe because [w] sounds more ‘English’ than [v] and new learner think that saying [w] makes them sound more ‘English’.

A related problem is the tendency of Slovaks to replace word-final and coda position [v] with [w] as is quite widespread in words like ‘have’ realized as [hew]. This habit can be attributed to native language interference since in Slovak we do realize coda ‘v’ consonants with [w] as in ‘stavba’ and ‘hláv’ standardly realized as [stawba] and [hla:w] respectively. However, this is more linked to the differences between Slovak and English in contextually dependent allophonic variation as will be discussed in the next section.

Unfortunately, the difference between [v] and [w] is not easy to recognize in the spectrogram since both are voiced and the presence of friction with [v] is typically low. Hence, other than the relative difference in energy ([w] being louder than [v]), the visual differentiation between these two sounds is rather challenging.

Fourth, the glottal fricative [h] in English is by default voiceless whereas in Slovak it is by default voiced, for which the IPA has the symbol [ɦ]. Hence, especially in the word-initial positions, words like ‘head’ or ‘hill’ are produced with a voiceless fricative whereas Slovak words like ‘Hana’ or ‘hustý’ with the voiced one. Interestingly, a distinct Russian accent is achieved if this voicelessness is overdone and sounds like Slovak ‘ch’ marked in IPA as [x]. In word-medial positions, and especially between two vowels or two voiced sounds, English ‘h’ is also typically voiced. Therefore, words like ‘behind’ or compounds like ‘greyhound’ have a voiced [ɦ]. Similarly to some of the previous patterns, this difference between Slovak and English is a case of contextually dependent allophonic variation and will also be discussed in the next section.

Fifth, and the least noticeable difference, is the place of articulation for alveolar consonants with tongue-palate contact [t,d,n,l]. In English, most speaker produce the contact with the tongue blade touching against mostly the alveolar ridge whereas in Slovak, the area of the contact includes both teeth and the area of the alveolar ridge right behind the teeth and both the tip and the blade of the tongue. Although quite minor and not crucial, this difference in the place of articulation has, in my mind, effect on the aspiration of [t]; see section 2.2 below.

2.1.2. Vowels

The Slovak/Hungarian interference on vowels poses equally, if not greater problems than consonants. Not only because the English vowel inventory is notoriously complex, but also because the positions of articulators, and especially the tongue, are difficult to introspect since the contact between the tongue and the rest of the mouth is typically minimal or non-existent. Hence, the learners must rely on their ears to detect the difference and also rely primarily on ears in forming the new habit reflecting the native vowel realization better than their native language habits.

Here we discuss only the vowels in the stressed syllables and leave the prosodic aspects related to word stress to section 2.3. Also, the patterns discussed below are the ones I identified as most common but the description below is not meant to be exhaustive.

One of the main phonological contrasts in vowels in Slovak is their length, sometimes also referred to as quantity. All five basic vowel qualities ‘i, e, a, o, u’ have short and long counterparts with minimal qualitative differences between the members of each pair; only long ‘á’ is sometimes perceivably more open than short ‘a’. In other words, short vowels like ‘i, u’ in words like ‘syrá’ or ‘sud’ sound almost identical to their long counterparts ‘í, ú’ in words like ‘síra’ or ‘súd’ apart from the difference in duration. In English high vowels, the difference between the short [ɪ, ʊ] as in ‘fit’ and ‘foot’ and their long counterparts [i:, u:] as in ‘feet’ and ‘food’ is realized with start and readily perceivable difference also in the quality of the vowels. The long [i:, u:] are characterized as ‘tense’ so that the articulators exert greater muscle expansion: the tongue is high and more peripheral (advanced in [i:] and retracted in [u:]) and the lips are spread ([i:]) or rounded ([u:]). Compared to these realizations, the short vowels, characterized as ‘lax’, are less peripheral: in [ɪ] the tongue is lower and less advanced and the lips are less spread than in [i:], and similarly in [ʊ] the tongue is lower and less retracted and the lips are less rounded than in [u:].

Since Hungarian vowel inventory also includes a similar contrast between short and long high vowels as in Slovak, the tense/lax distinctions tends to be problematic also the native Hungarians when they speak English.

The other major interference is now almost complete merger of the difference between ‘e’ [e] and ‘ä’ [æ] vowels in Slovak, which means that both are now pronounced as [e]. Hence, only in certain dialects and older speakers you might still hear words like ‘pät’ or ‘mäso’ pronounced as [pæc] or [mæso] while the majority of speakers would say [pec] or [meso]. In English, the difference between [e] and [æ] is preserved and thus the words ‘pet’ and ‘pat’ are typically pronounced with a salient contrast between the two vowels: [pet] vs. [pæt]. Because of the lack of [æ] in the inventory of most Slovaks, and the merger between [e] and [æ], a common interference when Slovaks speak English is the absence of this contrast: ‘pat’ is pronounced identically to ‘pet’; both as [pet].

Whereas Hungarians tend to share the problem of tense/lax distinctions with Slovaks, the [e]-[æ] contrast poses slightly different problems for them. While Slovaks tend to realize English [æ] as [e], some Hungarians tend to realize English [e] as [æ]. In other words, they would say [pæt] for both ‘pet’ and ‘pat’. This stems from the different vowel inventory of Hungarian in which short ‘e’ is realized much lower, almost as [æ] in Hungarian, when compared than the long ‘é’.

We can also mention other common problems such as the difference in quality between ‘short’ [ʌ], as in ‘cup’ and ‘long’ [ɑ:] as in ‘car’ for both Slovaks and Hungarians, or tendency for some Hungarians to round the long central vowel [ɜ:] in words like ‘turn’. Both of these stem from the different inventories between Slovak/Hungarian and English.

The English diphthongs, in my experience, pose less severe problems arising from the native language interference. A rather common tendency, however, is to pronounce centering diphthongs [ɪə, eə, uə] as long initial vowels [i:, e:, u:] and thus words like ‘pier’ ‘pear’ and ‘poor’ tends to be realized as [pi:(ɪ)], [pe:(ɪ)] [pu:(ɪ)] rather than more native-like [pɪə(ɪ)], [peə(ɪ)], and [puə(ɪ)]; the (ɪ) marks the possibility of rhotic realization as in American English and non-rhotic as in Southern British English.

2.2. Differences arising from allophonic variation

The problems outlined in the previous section can be characterized as ‘general’ since they refer to the realization of the sounds in almost any context and in any word. Hence, the difficulty of that type of interference lies in acquiring a habit that is not common in the native language. The type of interference outlined in this section poses a slightly different, and in my mind greater, challenge: the application of the different or novel realization that is, however, limited to certain context only, and this limitation is sometimes starkly different from the similar pattern in the native language. In other words, it is not only ‘mechanical’ drill of learning how to move the muscles to create an English-like sound, but on top of it acquiring the ‘rules’ for when to deploy the different realizations of a sound.

To use an example from Slovak, every native speaker ‘knows’ that ‘v’ might be realized as one of at least three possibilities: [f, v, w]. In ‘včela’ we say [ftʃɛla] and would recognize immediately a non-native Slovak speaker if s/he said [vtʃɛla]. But the ‘rule’ that we subconsciously know cannot be as simple as saying: if ‘v’ precedes a voiceless consonant, then say it as [f]. This is because in ‘ovčí’ in which ‘v’ precedes ‘č’ as well, we do not say [f] but [w] and if somebody says [oftʃi:], we would think s/he is Czech (or from the west part of Slovakia called ‘Záhorie’).

Hence, the realization of ‘v’ is predictable for native speakers and we can easily say, even for words we never said or heard ourselves, whether ‘v’ should be realized as [f], [v], or [w]. You can already imagine the troubles of some foreign nationals trying to acquire this single system of how to say ‘v’ in Slovak. By the way, can you think of the ‘rule’ that would explain these foreigners learning Slovak how to pronounce ‘v’? Importantly though, even if these foreigners themselves learn how to say perfectly sounding Slovak-like [f], [v], and [w], they have to develop a system of deploying

them in precisely specified environments, or less commonly in some exceptions. If you want another challenge, try to think how you would explain foreigners why we pronounce palatal [j] and [ç] in ‘deti’ but the alveolar [d] and [t] in ‘deratizácia’.

Every language has many such mini-systems of predictable variation that native speakers effortlessly command but non-native speakers must acquire if they want to sound more like native speakers. And English is, of course, not an exception. The goal of this section is not to list all such mini-systems, but to draw your attention to this phenomena and outline just several examples whose improved acquisition significantly affects how much non-native you sound to native English speakers.

We start with the aspiration of voiceless stops [p, t, k]. It is a perfect example of a combination of the two types of interference we outlined above: it is a pattern we do not have in Slovak and have to create ‘muscle memory’ by repeated deployment of naïve-like patterns in various environment, and its deployment is constrained and predictable to certain contexts and we have to create the cognitive mini-system to know when to do it.

As I mentioned before, the goal of this material is not to provide guidance for practicing native-like patterns, neither the articulatory patterns nor the contextual ‘rules’; we do that during the seminars. Here I only say that in terms of articulation, aspiration involves a specific coordination between releasing the complete obstruction (e.g. opening the lips for [p]) and setting the vocal cords inside the larynx such that they begin their vibration and thus start voicing. Whereas in Slovak we tend to do these two actions almost simultaneously, in aspirated English consonants [p, t, k] the opening of the obstruction precedes the beginning of voicing by approximately 50-100ms, which creates the perceived effect that we call aspiration and in IPA transcribe with an upper script ‘h’: [p^h, t^h, k^h].

Regarding the environment in which these aspirated allophones of voiceless stops occur in English, it can be described as the initial onset consonant in a lexically stressed syllable. The ‘onset consonant’ means that it starts a stressed syllable. Hence, these three consonants are aspirated in ‘pick’, ‘depict’, or ‘prick’. Note that the beginning of a stressed syllable does not necessarily mean the beginning of a word, as is shown with ‘depict’, in which ‘p’ is clearly aspirated, but in words like ‘parade’ or ‘perceive’, the aspiration is much weaker and more variable since the first syllable is not stressed. Importantly, if an ‘s’ precedes these consonants in the onset of stressed syllables, aspiration also does not take place: ‘speak’ is [spi:k] and not *[sp^hi:k]. When aspiration affects stops followed by voiced sonorant consonants – [w, r, l, j] are the set of consonants that can occur in this position in English – then these consonants are partly or fully devoiced. This is typically marked with a different convention in IPA, but the mini cognitive system is the same: ‘play’ is

transcribed as [p^hleɪ] and ‘quite’ as [k^hwaɪt] which are accurate descriptions of real speech than [p^hleɪ] or [k^hwaɪt].

When these consonants are in the coda position of a syllable, a completely different mini-system takes place, particularly for ‘t’, but we will not discuss it here since it does not involve aspiration as the phenomenon we are describing primarily here.

Another difference characterized as allophonic variation involved the systematic variability in the duration of vowels. This pattern is easy in terms of articulatory strategy. In both Slovak and Hungarian the length of vowels varies a lot due to various factors and thus it should be no problem to adjust their duration. It is not the case that we need to learn a novel articulatory mechanism, as is the case with aspiration. However, the mini system guiding this variability is different, or more precisely it produces more salient differences, in English than in the other two languages. The ‘rule’ can be described as follows: shorten the vowel in those stressed syllables, in which the coda consonants are voiceless. Hence, ‘cap’ and ‘cab’ are realized with noticeable different vowel duration such that [æ] in ‘cap’ is clearly shorter than in ‘cab’. In IPA the diacritic for shortening is the little ‘hacek’ over the vowel: [kæ̘p] vs [kæb]. The longer the vowel, the more noticeable the difference is and affects both monophthongs and diphthongs in stressed syllables.

Importantly, however, the voiceless consonant must be in the same syllable as the vowel since if it is in the following syllable, the shortening does not take place. For example, [æ] is approximately equally long in ‘cabinet’ and ‘capital’ despite the fact that it is directly followed by either voiced or voiceless consonants respectively. Crucial from ‘cap’ vs. ‘cab’ pair, however, is that the bilabial consonants are already in the second syllable in ‘cabinet’ and ‘capital’ while they are from the first syllable together with the affected vowel in ‘cap’ and ‘cab’.

The final pattern we mention here is more challenging for Hungarian speakers than for Slovaks, and involves the realization of the lateral approximant ‘l’. The allophonic variation between the so called ‘dark l’ and ‘clear l’ is quite well known and extensively studied. Despite more recent experimental findings of great variability of patterns and differences in various dialects of British English (e.g. Turton 2016), it is generally described that the light variant occurs in syllable onsets, or when followed by a vowel (e.g. ‘lip’) while dark variant in syllable codas, or when followed by a consonant or end of the word (e.g. ‘pill’ or ‘cold’).

The articulatory strategy for signaling this variability involves again the coordination of two actions in the vocal tract. The first one is the actual touching of the tongue blade or the tongue tip against the alveolar ridge. This activity is easy to introspect since all of us can feel the contact of these two

articulators. The second one is the retraction of the tongue body, or tongue dorsum, which is in part required for the tongue tip to move upward.

In ‘clear l’, these two actions are performed more or less simultaneously. But in ‘dark l’, the tongue body retraction precedes significantly the tongue tip/blade contact against the alveolar ridge. Hence, the dark ‘l’ is also commonly called ‘velarized’ since the tongue dorsum approaches the velum area. This is much more difficult to introspect since no contact of the articulators is involved.

While Slovaks command the variation in the quality of ‘l’ that is not very different from English, possibly also due to greater phonological variability in terms of short and long ‘l’, Hungarian, similar to other languages like German, for example, tend to use only the clear variant. Hence, in words like ‘pill’ or ‘cold’, native speakers distinctly recognize the non-native accent.

To sum up, in all three patterns covered (aspiration, vowel shortening, and clear/dark ‘l’), the changes affect individual sounds only in some limited set of environments, and typically present variation that does not affect the understanding of words, hence it is not phonological, but clearly signals the absence of native-like mini-systems of knowing which variant of a sound should be employed in which environment and context. Hence, the native language interference arises both from the ‘mouth’ in lacking the articulatory habits, and from the mind in lacking the system of ‘rules’ guiding the distribution of these variants in connected speech.

2.3. Prosody

In this final part of the section devoted to the native language interference when Slovaks (Hungarians) speak English we briefly mention the aspects connected more directly to prosody. Note, however, that all three patterns in the previous section involved at least the syllable, or a stressed syllable, and thus they can also be construed as suprasegmental patterns rather than just involving individual segments.

The core challenge, in my experience of working with the pronunciation of non-native speakers for decades, is the realization of stressed vs. unstressed syllables. In Slovak and Hungarian, the lexical stress system is very similar in that the leftmost syllable of the word receives the primary stress but the phonetic realization of this is rather weak, and secondary stress is very elusive and difficult to characterize. Hence, the phonetic difference between the stressed and unstressed syllables are relatively weak.

On the other hand, lexical stress in English is radically different. The position of stress is not fixed and virtually any syllable of a word can be stressed. Phonetically, the contrast between the stressed and unstressed syllables is very robust and salient. It involves not only greater difference in terms of loudness, duration of pitch, but also sometimes radical differences in the quality of vowels. Most unstressed syllables are realized as schwa or schwa-like centralized vowels.

The challenge of Slovak and Hungarian speakers is thus two-fold. First, they must acquire the cognitive system, the knowledge, of which syllable is the stressed one for each word. This is partly predictable in English, but we will not devote space here for listing the tendencies for English stress. Of course, the interference is that many Slovak/Hungarian speakers have a tendency to stress the initial syllable.

While approximating the native speakers in terms of lexical stress and knowing which syllable to stress is demanding, and requires great commitment especially for students arriving with bad habits in word stress from their previous schooling, the second challenge is, in my mind, equally difficult and tricky. It involves acquiring the habits and phonetic mechanisms for making the clear contrast between the stressed and unstressed syllables. It requires working on increasing the robustness of the stressed syllables, and as well, and possibly more importantly, neutralizing and weakening the unstressed syllables.

The more we move away from segmental aspects of native speaker interference to the prosodic ones, the more difficult the description is due to great complexity of the problem. I am assuming you have covered the aspects of fluent speaking regarding weak forms, and ways of linking words that are covered and practiced in Part 1 of this teaching material. That engagement with the sound files and the commentaries provide a better picture than what I would be able to describe theoretically in the short space here.

Similarly, the variability in intonation, and more specifically in the pitch and melody, are so contextually determined and complex that I prefer discussing them with concrete sound files through hands on engagement and experience rather than reviewing the theoretical concept. I mention here only fragmentary aspects of interference arising from the differences between Slovak and English for you to keep in mind when I review the system for describing intonation in the following section, and more importantly, when you read my commentaries on intonation in the sound material analyzed in this book.

One of the major differences between Slovak and English is in signaling prominence; or, in other words, highlighting the words to which the listener should pay attention. In Slovak, one of the systems we use for this is variable word order and place the important words toward the end of the

utterances. Consider for example a simple question my wife asked me this morning: ‘Donesiem ti kávu?’ ‘Shall I bring you the coffee?’. We feel that she really asked about the coffee. However, if she said ‘Kávu ti doniesiem?’, coffee is assumed, or given, for example because both of us know we drink coffee every morning, and the question is focusing on whether she should bring the coffee or I will get it myself. Although we put prosodic emphasis on ‘kávu’ in the first case and ‘donesiem’ in the second one, prosody plays a supporting role together with word order in signaling this information.

English, however, does not have this system at its disposal for signaling what is important (apart from more complex syntactic structures as pre-posing or clefting). In English, the word order is much more fixed and thus prosody plays a more important role. In both situations highlighted above, the same English utterance ‘Shall I bring you the coffee?’ would most likely be used, but the focus of the question would be signaled only prosodically: ‘Shall I bring you the COFFEE?’ vs. ‘Shall I BRING you the coffee?’.

While the above example presented a general difference between the linguistic systems of Slovak and English potentially leading to several interference mechanisms, there are also differences in the prosody-meaning systems of the two languages that are specific for cuing certain functions. For example in both languages the wh-questions are realized with emphasis on the wh-word and generally falling pitch. However, in English there is a greater tendency to emphasize also the last prominent word of the question by producing a clear f₀ peak on this word and creating so called hat pattern whereas in Slovak the melody tends to fall gradually over the entire question.

The important aspect for the non-native speakers are not just differences potentially affecting the meaning of utterances, but especially for prosody, also continuous adjustment participating in signaling emotions, attitudes and other personal states. For example, it is long known (e.g. Mennen et al. 2007) that non-native speaker utilize lower ranges of pitch and loudness and thus their speech is less varied and sounds more flat. However, these features are commonly associated with boredom, shyness, or other aspect in native speakers and thus increasing variability and ranges of prosodic features is one of the greatest challenges for both Slovaks and Hungarians when they speak English.

3. Overview of simplified ToBI for annotating prosody

Just like the International Phonetic Alphabet (IPA) represents a useful tool for describing the segmental aspects of speaking, some tool is also required for discussing the intonational aspects. Unfortunately, the research on intonation, its complex, language-specific, and sometimes subjective nature, results in less unity in how prosody should be labeled when compared to individual sounds. In this material we will be using a simplified system for annotating prosody called Tones and Break Indices (ToBI). It should be noted, however, that ToBI should not be considered as an IPA-like system for prosody primarily because the fact that the relationship between the prosodic form, and the pragmatic meanings signaled with this form varies among languages.

There are several reasons for using ToBI including its greater simplicity compared to other approaches, its widespread use and ‘localization’ for characterizing various other languages, and the availability of manuals and online tutorials with examples and sound clips for anybody to practice. The brief description of the system can be found at <https://www.ling.ohio-state.edu/~tobi/>, complete guidelines are at http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf, compact description for example in Beckman et al. (2011) and importantly also self-study material mentioned already in the Introduction developed by A. Brugos, N. Veilleux and S. Shattuck-Hufnagel for a course at MIT “Transcribing Prosodic Structure of Spoken Utterances with ToBI” available through MIT’s open courseware initiative (<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm>).

In what follows I first describe the types of phonological contrast that we wish to describe and basics of ToBI as the approach to this type of intonation variation. Then I describe the subset of primitives from this framework that I will use when discussing intonation in this book, and conclude with several examples of pragmatic meanings signaled through the intonational variation.

3.1. What aspect of intonation and why ToBI is useful

There are three fundamental aspects of prosody that are responsible for cuing most of the categorical and discrete aspects of pragmatic meaning. What I mean by discrete and categorical is similar to our discussion in section 0 in that replacing [θ] with [s], not aspirating in ‘pat’, or failing to signal the contrast between the first and second syllables of ‘event’ are all discretely perceivable aspects of non-native speech. On the other hand, the fact that someone may speak slower or use a narrow pitch range does not create a meaning contrast albeit of course conveying various emotional

and attitudinal aspects. The three such aspects of intonational phonology are: 1) which words are prominent, or accented, 2) what are the tonal targets, and 3) how are utterances chunked into smaller units. We will discuss them briefly in turn below.

The choice of which words are accented has already been exemplified in the previous section with the question ‘Shall I bring you the coffee?’. We mentioned that there is a clear contrast between the two realizations: ‘Shall I BRING you the COFFEE?’, which is neutral and signals ‘broad focus’ in the sense that no specific previous context is assumed, whereas ‘Shall I BRING you the coffee?’ signals a narrow focus on the action of ‘bringing’ and is most natural in the context in which ‘coffee’ is given such that both the speaker and listener already know about the coffee either directly from context (e.g. if we talked before about what type of coffee we want this morning), or might be inferable from situation (e.g. if we both hear and smell the coffee is brewing). In this sense, the two utterance signal different meanings, placing the first one in the context of the second one, or vice versa, would sound strange, and thus signal the categorical and discrete aspects of prosody-meaning system.

The second aspect mentions the identity of the tonal targets. Here we also mention one example to illustrate the meaning contrast. First, imagine the positive answer ‘Yes’ to a question ‘Do you like pizza?’. There are literally tens of ways how this ‘Yes’ could be said. It could be said with a fall (from high to low) in melody, conveying positive answer and a bit of finality. But it could also be said with f₀ rise (from low to high) signaling apprehension, expectation that something will follow. None of these meanings is signaled with the first falling ‘Yes’. Hence, another case of categorical and discrete contrast signaled by different tonal targets.

A continuous contrast here might be signaled with varying degrees of slope in the falling version of ‘Yes’: the speaker might be mildly enthusiastic about the option to have pizza for dinner, or s/he might love it and be excited. Both of these would be signaled with a fall but the extent of this fall would be the continuous aspect of prosody not covered in ToBI.

The last, third, aspect is the presence vs. absence of prosodic boundaries, or breaks because their distribution defines how longer stretches of speech are divided into smaller chunks or units. The example we use here relates to so called structural ambiguities of English. The sentence ‘This fare is restricted to retired politicians and civil servants’ is ambiguous regarding the civil servants: is the fare restricted only for retired civil servants or any civil servant? The presence of a prosodic boundary might suggest one of the meanings and thus provide disambiguation. If we divide the utterance with a break before ‘and’ (‘This fare is restricted to retired politicians // and civil servants’), we favor the meaning that any civil servant might use the fare. Without such break, the other meaning is favored.

In this approach to identifying phonological (i.e. contrastive discrete) meanings conveyed by variation in intonation (targets and boundaries), ToBI makes the natural choice as the framework for annotating intonation. It tries to capture only meaningful functional contrasts, not continuous aspects like rate or loudness. This was motivated by the effort to include prosody in computationally processing speech and not just rely on text, and the success of adding some syntactic labels into text corpora. Also, ToBI has been extensively tested on the agreement of labelers, which other approaches did not stress. Moreover, ToBI labeling is integrally linked to visual inspecting the speech signal, particularly f0, and thus adds a degree of objectivity. Nevertheless, the reliance on ears and native speaker perception of the intonation is always given priority. Lastly, ToBI is ‘living’ adjustable framework rather than a set of rules and thus each researcher might add or omit those aspects that s/he feels are relevant for the purpose of the labeling. And thus we will also use a simplified version of ToBI, outlined in the following section, that suits our needs better than the full framework.

3.2. Crash course to simplified ToBI

The two major aspects of ToBI are tonal targets (Tones), and the strength of disjuncture between words (Break Indices). There are two major categories of tonal targets: those associated with prominent words, these are called pitch accents, and use the diacritic ‘*’, and those associated with the edges of prosodic boundaries, and use the diacritics ‘%’ for major boundaries of intonational phrases or ‘-’ for minor boundaries of intermediate phrases.

ToBI describes intonational contours as arising through interpolation between successive tonal targets. And these targets essentially can only be high, labeled with ‘H’, and low, labeled with ‘L’.

One of the strict, and thus easy to implement, conventions in ToBI is that a major boundary always subsumes the presence of a minor one and thus a ‘%’ symbol is always used together with a ‘-’ symbol, for example L-H%. A minor boundary can occur on its own (e.g. L- or H-) but a major boundary always combines the intermediate phrase accent and the intonational boundary tone.¹ As it is clear from the difference between ‘major’ and ‘minor’ boundaries, the degree of disjuncture is perceived as not as strong in the minor boundaries compared to the major ones.

We discussed the phonetic cues for marking the boundaries and their strength in Part 1 of this material; hence you are referred to that discussion or to a well-presented outline of ToBI with

¹ This will apply to our simplified system. In a complete ToBI framework, an H% can marginally be used to label high beginnings of some intonational phrases.

examples in the MIT courseware. Briefly, the phonetic signatures include the degree of pre-final lengthening, the number of pitch targets (e.g. two in L-H% but only one in L- or L-L%), the degree of f0 excursion, and the evidence for pitch reset following the boundary.

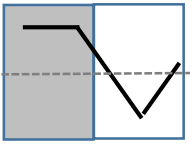
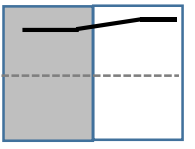
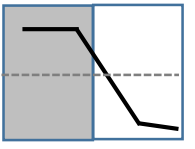
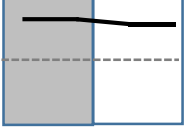
There are five levels of boundary strength in ToBI and this index is associated with the right edge of every word. The lower levels 0-2 are not associated with tonal targets and in our simplified system we will only make the distinction between break index 0 that indicates connected speech phenomena such as assimilation, elisions described in Part 1 of this book, and break index 1 indicating regular smooth transition between two words.² Break indices 3 and 4 are associated with the phrase accents (e.g. L- or H-) and full boundary tones (e.g. L-L% or L-H%) respectively.

The other basic rule simple to remember is that every phrase (either minor intermediate or major intonational) has to have at least one pitch accent. Hence, we now have the basic inventory of primitives used in our simplified ToBI framework:

- pitch accents: H*, L*
- phrase accents: H-, L-
- boundary tones: H-H%, L-L%, H-L%, and L-H%

Let's take now a simple one-word utterance 'Really' as a response any proposition from the previous speaker. The table below illustrates schematically and describes the four possible contours when the intonational boundaries combine with the H* pitch accent. The dotted line indicates the approximate midline of the speaker's pitch range. The two boxes refer to the two syllables of the word, the shaded box marks the first stressed syllable.

² Break index 2 is most commonly used to label hesitations, speech errors, and mismatches between the phonetic cues for the index 1 and index 3

	H* L-H%	The stressed syllable is associated with a high target. F0 then dips to below the speaker's midline and before the end rises to approximately middle of the range. This is a prototypical example of so-called continuation rise expressing that the speaker is not finished and will continue.
	H* H-H%	The contour starts with a high pitch target and then does not lower and stays either high or get little bit higher. It might be used for example expressing extreme joy and expecting further confirmation.
	H* L-L%	This is a typical falling contour that starts high on the stressed syllable and ends low by the end of the word. It expresses finality, signaling that the speaker is finished and does not plan to say more.
	H* H-L%	This is a typical plateau flat contour despite the L target at the end that here refers to no rising, rather than falling

In a similar fashion we can combine the low pitch accent L* with all four boundary tones, and also label edges of minor intermediate phrases with the four logical combinations H* H-, H* L-, L* H-, and L* L-.

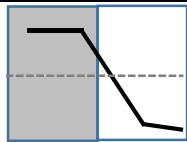
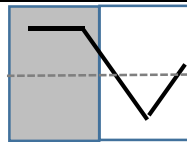
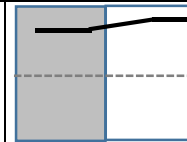
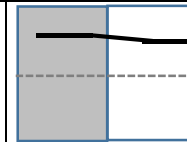
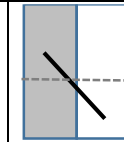
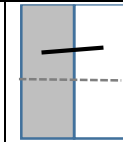
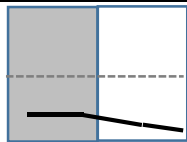
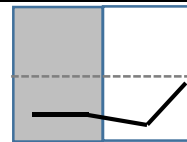
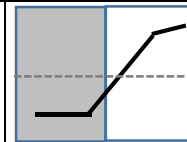
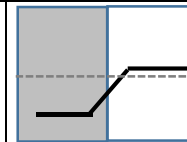
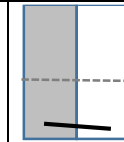
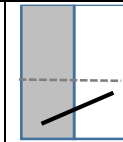
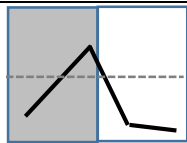
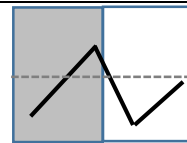
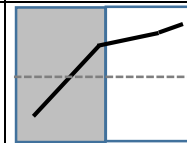
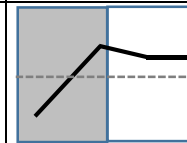
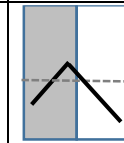
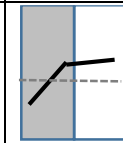
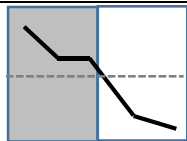
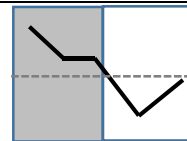
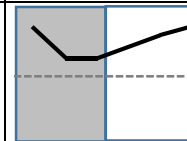
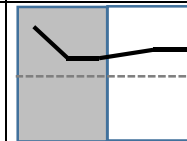
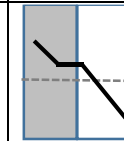
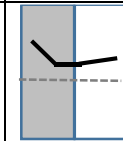
There are two elements of the full ToBI system that we will add to our simplified system as well. Both of them involve the possibility to have bi-tonal accents. Hence, in addition to simple H* or L*, the two most common bi-tonal pitch accents are rising L+H*, and a 'stepped-down' accent H+!H*. As you can see, we are adding one more symbol, the plus sign, to our inventory of ToBI primitives. This sign marks that the two tones are associated to a single stressed syllable, they form a single unit, and the tone followed by the star is the more prominent of the two both in terms of perception (L+H* is perceived more as an H-target than an L-target) and the alignment of the individual sounds and tones (most of the rhyme of the stressed syllable is rising, or already high).

The first bi-tonal accent, L+H*, typically shows, and is perceived as, a 'scooped' accent in that pitch falls slightly, typically before the stressed syllable, and then rises during the major portion of the stressed syllable. The wide-spread use of this tone is to mark extra emphasis, or in some cases contrast with what was said or assumed before. For example, L+H* L-L% with 'Really' expresses extreme surprise when what was just said contrasts with what the speaker thought of believed.

The second bi-tonal accent, H+!H*, includes the notion of ‘downstep’, marked with an exclamation mark, and expressing the only syntagmatic relationship in ToBI. That is, !H tone is dependent on a preceding H-target as it can only follow such an H-target, either in H* or in L+H*. All other tones are paradigmatic; that is, they are not restricted in their occurrence by preceding or surrounding tones.

!H means that the accented syllable is realized by a perceptually lower f0 than that of an immediately preceding High tone: the tone has ‘stepped down’ from the preceding High. This downstepped high target can stand alone as !H*, for example when listing successive elements, the first one is typically H* and the subsequent ones fall down as if ‘down from the stairs’ and are labeled as !H*. Downstep can also form a bi-tonal accent H+!H* in which the unstressed material before the accented syllable constitutes a high target but the stressed syllable itself is lower than this initial H-target.

Table with schematic illustrations of boundary tones and accents is below.

	L-L%	L-H%	H-H%	H-L%	L-	H-
H*						
L*						
L+H*						
H+!H*						

3.3. Sample of compositional intonational meanings

The brief notes collected in this section arise primarily from the work of Jannet Pierrehumbert and Julia Hirschberg (1990). The idea is to show that compositionality, as the hallmark of any linguistic system, is also present in the ToBI-inspired system of intonational meanings. Compositionality refers to the idea that in a system of individual primitives, these could be combined to form novel structures and that the meaning of these structures can be (partially) inferred from the meaning of the individual primitives. For example, combining a verb (e.g. ‘buy’ with a noun (e.g. ‘book’) in the sentence ‘She bought a book’ creates the notion of ‘book-buying’ which is novel and interpretable from the meanings of ‘buy’ and ‘book’ separately. Pierrehumbert and Hirschberg show that intonation works in a similar way.

Additionally, my notes offer the notion of intonational meaning that is complementary to the more traditional meanings of intonation categorized as attitudinal, grammatical, accentual, and discourse as for example in Roach (2000).

The basic assumption developed in Pierrehumbert and Hirschberg’s proposal is that *tunes, i.e. intonational contours decomposable to pitch accents, phrase accents, and boundary tones, convey the relationship between the hearer’s beliefs, the contribution of the current utterance with respect to these beliefs, and the relationship to the upcoming utterances*. For example, if a speaker says an utterance with certain proposition (e.g. ‘I like pizza’) s/he considers what the hearer beliefs are before saying this (e.g. whether the hearer knows this or does not know), how this utterance is going to change these beliefs, and how this utterance is going to be linked to the future utterances (e.g. if any related proposition is to be made or if the intention of the speaker has been completed with this utterance and any subsequent utterance expresses a novel intention not dependent on the previous utterance). Intonational meanings construed this way are less dependent of specific contexts and situations than meanings capturing attitudes, emotions, politeness, irony, or similar meanings, that are cued by intonation as well, but due to these dependencies more difficult to capture in a formal system.

One of the general examples used by Pierrehumbert and Hirschberg is the utterance in (1).

(1) The train leaves at seven

There are three discourse-relevant elements (‘train’, ‘leaving’, ‘seven’) and the placement of pitch accents signals the speaker’s assessment of the salience of these items in discourse and to the hearer. Hence, we can accent one, two, or all three items to indicate which items should the hearer include in her beliefs. Additionally, the speaker signals what s/he thinks about the beliefs of his/her

hearer. This is because if any of these elements is not accented, it is likely that the speaker thinks that the hearer already has this in his/her list of active elements.

If the boundary tone is high, we expect that something is to follow and that we should wait to hear that follow up and only then make the interpretation of the entire utterance. Note that this is what we do if something related follows. For example, if the utterance in (1) ends in a high boundary tone and is then followed by ‘or nine thirty’, we naturally link these two utterances as forming a single thought. However, the same happens also if some apparent non-sensical continuation like ‘There’s a full moon tonight’ follows utterance (1) after a high boundary tone. We have a natural tendency to search for the connection between the train leaving at seven and today’s full moon. This *forward looking* meaning of high boundary tone is in contrast with low boundary tones, that suggests the two utterance divided by this tone are separate.

Regarding pitch accents, H* signals that the referents should be treated as new in the discourse, added to the mutual belief, and conveying some new information to the speaker.

In comparison, L* accents tend to signal that the item is already part of the mutual beliefs system, or, as in yes/no question, the speaker does not add anything new but asks the hearer to make the addition. This applies to situations expressing genuine shared beliefs (as if responding to a question about wishes for birthday gifts the speaker says I want a Lego,... with L* on ‘Lego’ conveying something like ‘as we both know...’), but also when the speaker might reprimand, or mock the hearer essentially saying ‘you should have known that...’.

The L+H* is the accent that invokes a scale and more specifically, suggesting that the accented item rather than some alternative, should be placed into the mutual beliefs system. Hence, this accent most commonly marks contrasts or corrections.

The full list of meanings discussed by Pierrehumbert and Hirschberg is beyond the scope of this book but they propose similar types of meanings for downstepping contours, and show how the meaning of individual tonal targets (both pitch accents and boundary tones) is compositional in analyzing the meaning of various combinations of these primitives.

4. Commentaries and analyses of a dialogue

The current section is the core of the book. It continues with the approach of Part 1 in the basic approach to the material. First, it provides my commentary to about a minute of a continuous dialogue. It expands on the aspects covered in Part 1 (basic prosody, weak forms, and connected speech aspects) by including also the analyses of interference in the segmental domain, and a slightly enriched analysis of intonation with ToBI.

Second, in keeping with the approach of Part 1, I again start with discussing the prosody of the turn, then reviewing the aspects of connected speech, and finishing with discussing the segmental patterns deserving attention of non-native speakers of English with Slovak or Hungarian mother tongue.

Third, you are again expected to use Praat, and inspect both the auditory and the visual characteristics of speech offered by opening the sound file and the accompanying textgrid files in Praat. You are also expected to ‘get your hands dirty’ with the material, meaning to explore on your own, listening and imitating the portions, and eventually the entire turns of the native speakers.

Finally, as in Part 1, the analysis is incomplete in that only the first half of the dialogue is discussed and the second half is left for you to engage in, and analyze on your own, and possibly discuss during the seminar.

The next subsection provides the basic description of the analyzed material by providing the situational context and transcripts, both text and IPA, of the entire dialogue. The rest provides the commentary for each utterance turn-by-turn.

4.1. Dialogue description

We are going to analyze the first half of the initial part (0:17 – 2:55) of the dialogue between Linda Holmes and Stephen Thompson that was aired on National Public Radio (NPR) as part of the program Pop Culture Happy Hour Small Batch entitled Remembering Prince, and was aired on April 21, 2016, 5:43 PM ET. This file is used with the permission of the NPR. The website includes the description, and also contains a downloadable sound file from the entire program.

<http://www.npr.org/sections/monkeysee/2016/04/21/475142333/pop-culture-happy-hour-small-batch-remembering-prince>

The current topic of the regular program is the death of Prince, the popular musician, that was announced that week. The seasoned collaborators on this radio show as interlocutors are in the studio and the genre is a radio discussion for broad audience. This situational context has to be always kept in the back of your minds when analyzing the dialogue. Speech would be different if this was a telephone conversation without the visual contact or if these were just two friends talking without the audience, or if the roles of the speakers were different; for example if one was an anchor and the other a guest on only this single show. Their goal, as stated on the website, was to share first thoughts on Prince after his death was announced and express their sadness and frustration on hearing the news.

4.1.1. Text Transcript

The speakers are represented with their initials: Linda Holmes as LH and Stephen Thompson as ST. Squared brackets indicate overlapped speech from both interlocutors. The transcript was done by me and I decided to include only end of sentence punctuation, such as full stops or question marks, for well-formed sentences syntactically and leave out the placement of commas and other punctuation. The regular black color indicates the material that is analyzed in the book and the gray color following the dashed line the material left undiscussed.

LH: Steven this sucks.

ST: This is not a good year.

LH: This is not. I'm tired of doing this with you, uh. Steven and I are in the studio today uh having just received the news of the death of Prince who was fifty seven. What is this year man?

ST: Agrh. I mean looking over a list of musicians who've died this year alone we didn't even talked about Merle Haggard on the [show].

LH: [No] we [didn't].

ST: [Uh], Vanity uh was one that hit me [really] hard.

LH: [yeah]

ST: She was a collaborator of Prince's who died [earlier this year].

LH: [My I'm glad you] mention that.

ST: Also at fifty seven. And that's the first thing I wanna get into with Prince is just how frequently and extensively he worked with and celebrated and provided songs for this incredible array of woman musicians. Uh he had women in his band throughout his career you know looking over a list of of his collaborators and not only women he performed with or produced but women he wrote songs for. He wrote "I feel for you" by Shaka Khan he wrote "Nothing compares to you" which Sinead O'Connor covered Uh

LH: He wrote uh "When you were mine" which Cindy Lauper covered [I know].

ST: [He] he wrote uh for Shena Easton and Sheila E and Apollonia and and this incredible array of women whose music he celebrated and then you get into his own career his own body of just this unbelievable collection of songs.

LH: yeah

ST: This towering canon of funk and R&B and pop songs. Part of what always struck me about Prince I always bought Prince's albums when I was a kid 'cause they were not only popular they were not only on the radio but they were just so cool.

LH: They were so cool.

ST: They were just so good.

LH: They were so cool and the other thing is you know I am much less closely connected to music than you are in an ongoing way but for me when I t- I talk sometimes about the fact that like look my generation when people talk about the music of like the eighties when I [was a] teenager

ST: [hm]

LH: you take a lot of beatings for some of the stuff that was popular [then. And and and]

ST: [yeah, no absolutely]

LH: people make fun of you whatever you know the people who are my proximate generation darn it they were able to recognize that "Purple Rain" was like a great great record and that that record was incredibly popular.

ST: [yeah]

LH: In [addition] to being recognized as great it was incredibly popular. It was playing everywhere so when I think about the fictitious line between popularity [and]

ST: [right]

LH: and quality

ST: yeah

LH: that is literary the record that I am usually going to. The other thing is like you said it's just the volume

ST: [yeah]

LH: [because] as we were coming in here I was talking about the song "Seven"

ST: yeah

LH: and you just have to hear the beginning it is like an electric shock.

4.1.2. IPA transcription

This is a transcription the 'text' of the first half presented above. It includes strong forms and thus represents a 'dictionary' pronunciation of single words produced separately. The conventions that were used for this transcription include the use of American English (AE) as both speakers speak this variety. Also, long vowels were represented with a colon [:] but vowel [ɑ] is short despite its somewhat longer realization in AE. Since AE is a rhotic dialect, [ɹ] was used in syllable codas. Finally, full stops in the transcript are marked with double slash as major prosodic boundaries [//]. The phonetic and prosodic choices made by the speaker are discussed in detail when analyzing each turn.

LH: 'sti:vən ðɪs sɪks //

ST: ðɪs ɪz nat ə gɒd jɪə //

LH: ðɪs ɪz nat aɪm 'taɪəɪd əv 'duɪŋ ðɪs wɪð ju ə 'stɪvən ænd aɪ ɑ:ɪ m ðə 'stu:di, ʊt tə'deɪ / ə 'hævɪŋ dʒʌst rə'si:vɪd ðə nu:z əv ðə deθ əv prɪnz hu wɪz 'fɪfti 'sevən // wɪt ɪz ðɪs jɪə mən //

ST: x: aɪ mi:n 'lɒkɪŋ 'ʊvəɪ ə lɪst əv mju'zɪfənz hu:v daɪd ðɪs jɪə ə'loʊn // wɪ 'dɪdənt 'i:vən tɔ:kt ə'baʊt mɜ:l 'hæɡəɪd ən ðə [fʊt] //

LH: [noʊ] wɪ: ['dɪdənt] //

ST: [ɜ:] 'vænəti ə wɪz wʌn ðæt hɪt mi: ['rɪəli] hɑ:ɪd //

LH: [jeə] //

ST: ʃi: wəz ə kə'ləbə,reitəɪ əv 'prɪnsɪz hu: daɪd ['ɜ:ɪliəɪ ðɪs ʤɪəɪ] //

LH: [maɪ aɪm glæd ju:] 'menʃən ðæt //

ST: 'ɔ:lsoʊ æt 'fɪfti 'sevən ænd ðæts ðə fɜ:st θɪŋ aɪ 'wʌnə get 'ɪntu wɪð prɪns ɪz dʒʌst hæʊ
'frɪkwəntli ænd ɪk'stensɪvli hi wɜ:kt wɪð ænd 'selə,breitəd ænd prə'vaɪdəd sɔ:ŋz fəɪ ðɪs
ɪn'kredəbəl ə'reɪ əv 'wʊmən mju:'zɪʃənz // ə hi: hæd 'wɪmən ɪn hɪz bænd θru'aʊt hɪz kə'riəɪ
ju noʊ 'ləkɪŋ 'oʊnəɪ ə lɪst əv əv hɪz kə'ləbə,reitəɪz ænd nat 'oʊnli 'wɪmən hi: pəɪ'fɔ:ɪmɪd
wɪð ɔ:ɪ prə'dʌst bʌt 'wɪmən hi: ruʊt sɔ:ŋz fɔ:ɪ // hi: ruʊt aɪ fi:l fɔ:ɪ ju: baɪ 'ʃʌkə kən hi ruʊt
'nʌθɪŋ kəm'peəɪz tu ju: wɪʃ ʃɪ'neɪd oʊ'kʌnəɪ 'kʌvəɪd ə //

LH: hi: ruʊt ə wen ju: wɜ:ɪ maɪn wɪʃ 'sɪndɪ 'lɔ:pəɪ 'kʌvəɪd [aɪ noʊ] //

ST: [hi] hi ruʊt ə fɔ:ɪ 'ʃɪ:nə 'i:stən ænd
'ʃɪ:lə i: ænd ,æpə'lounɪə ænd ænd ðɪs ɪn'kredɪbəl ə'reɪ əv 'wɪmən hu:z 'mju:zɪk hi:
'selə,breitəd ænd ju: get 'ɪntu: hɪz oʊn kə'riəɪ hɪz oʊn 'bædi əv dʒʌst ðɪs ,ʌnbə'li:vəbəl
kə'leɪʃən əv sɔ:ŋz //

4.2. Turn-by-turn commentary

The text transcription of individual turns are identical with those in section 4.1.1 but the IPA transcripts are, sometimes radically, different from the dictionary form of transcription used in section 4.1.2. The IPA transcripts below try to reflect what and how was actually said by the speakers.

(2) LH: Steven this sucks.
'sti:vən // ðɪs sʌks //

This is the opening statement of the dialogue. The utterance is clearly divided into two major chunks or units or intonational phrases. Hence, both boundaries are major boundaries requiring boundary tones and the phrases have to have at least one pitch accent. The first phrase 'Steven' has a typical high rising pitch contour (H* H-H% in the ToBI notation). Note how this rising contour is not a question but is used to grab the attention and signals with H* that this, and what would follow, is new information. The high boundary also signals that phrase 'Steven' should be interpreted together with what follows this phrase.

The second phrase 'this sucks' has a single pitch accent on 'sucks' and a falling contour. The type of the pitch accent is the donwstepped (H+!H*) in which the main pitch accent is on 'sucks' but the high target on 'this' corresponds to 'H+'. The low target at the end is compounded by utterance final

creaky voice, shown as visible vertical lines in the spectrogram, which is a common expression of low pitch target both physiologically as well as socio-linguistically. These notions were already discussed in Part 1. Note that Praat is not interpreting this creakiness correctly and the pitch contour is absent but our hearing clearly indicates a low target.

Regarding connected speech, the phrase ‘this sucks’ ends and starts with a fricative [s], which makes the frication a bit prolonged than as if only a single [s] was present. Hence, this is not a case of elision, as some of you might assume as this falls under a general pattern of preserving the temporal slot of the word-final consonant rather than elision in situations when a consonant in a final position of a word is followed by an identical consonant in the initial position of a following word.

(3) ST: This is not a good year.
ðɪs ʔɪz nɑt eɪ gʊd jɪəɪ //

Prosodically, this sentence is pronounced in markedly slow speech rate as each syllable is longer than 300ms. You see this when you compare this with the syllable durations in some faster chunks of speaking later in the excerpt. Almost each word is prosodically prominent, which amplifies the slow speech and adds emphasis and gravity to the statement expressed in this utterance. Regarding chunking, we have two options. First, treat this as a single unit, or treat each word as a separate chunk. Despite some discontinuity among the words and not smooth transitions, I would opt for the second option primarily due to the absence of pitch reset between the words. In other words, the declination of in the entire turn is a strong indication that these are not separate chunks and thus I labeled the breaks with #1 break index.³

The first pitch accent is a very nice prototypical example of the rising accent (L+H*) since the entire rise occurs during the stressed syllable. All the following accents are successively lower in the ever decreasing pitch range of the speaker. Hence, I would label them as downstep high (!H*) in despite the fact that especially last two words are really low in the speaker’s register. Yet, the end of the word ‘year’ is produced with a marked creaky voice, which suggests that the low target of the final L-L% is even lower than the pitch accent on ‘year’ and we can still perceive a falling contour on this word; hence, !H* L-L%.

In terms of connected speech, the slow and emphasized way of pronouncing each word leads to the realization of a strong form even in the determiner ‘a’ [eɪ], and also an initial glottal stop in the realization of ‘is’. Hence, not only accenting but word boundaries as well participate in signaling the speaker’s intention of emphasis on individual words.

Segmentally, [ʊ] of ‘good’ is a very nice example of this vowel and a nice model for practicing the quality of this English vowel that is very different from short Slovak/Hungarian [u]. As discussed

³ In full ToBI, this would be a good candidate for #2 breaks.

in section 2.1.2 you may compare with the Slovak pronunciation of words like ‘sud’ or ‘guma’. In English, which vowel is much more central than in Slovak, and if the lips are rounded at all, the rounding is extremely minimal whereas in Slovak we tend to round the lips more.

Another useful word here is ‘not’ that shows a typical elongated [ɑ] vowel of American English compared to a more rounded [ɒ] typical for British English.

Finally, the accenting of ‘this’ renders the initial dental fricative [ð] as very salient and shows clearly that this is a fricative not a plosive, as many non-native speakers might produce it. If this was a stop [d], the spectrogram would show a sudden burst and increase of energy. However, it shows gradual increase with no burst and elements of friction noise in the waveform.

- (4) LH: This is not I’m tired of doing this with you uh Steven and I are in the studio today uh having just received the news of the death of Prince who was fifty seven. What is this year man?

ðɪsɪznɒt // ʌm 'taɪəɪd ə'duɪŋðɪs wɪð ju // ə 'stɪvən ənaɪ ɑ:ɪ ɪŋŋə 'stju:ri,ou tə'deɪ //
 ə 'hævɪŋ/ dʒʌst rə'si:vðə nu:z // əv ðə deθ əf prɪns // hu wʌz 'fɪftɪ 'sevən // wʌt ɪz
 ðɪs jɪəɪ mæn

This is a very complex and varied turn prosodically since the speaker produces syntactically unfinished utterances, changes in speech rate, or several hesitation. Once we discuss dividing this turn into individual prosodic units, or chunks, we will take each chunk in turn and discuss its intonation, connected speech aspects and segmental aspects.

As the first cut, we will make use of the end-of-sentence punctuation included in the transcript. There is one full stop and we may assume that it will be signaled with a major boundary, but we will have to always check if that was the case. The full stop after ‘seven’ is associated with a long silent pause and thus the placement of a major prosodic boundary coincides with the syntactic one. The unit following this pause ‘what is this year man’ is similar to ST’s slowed down and emphasized turn ‘This is not a good year’ and despite some clearly perceived disjunctures, I will treat it as a single intonational phrase with a break after ‘what’ as a very plausible alternative hypothesis.

Moving now to the initial part preceding the full stop, the sound file again provides some help. There are salient pauses after ‘news’ and ‘Prince’ clearly indicating major prosodic boundaries, and a less salient and shorter, yet still clearly perceivable pause after ‘today’. Hence, with just silent pauses, we have now more manageable chunks that we inspect to determine if further divisions can be made. There are two such further divisions. First, I would place a major break after ‘not’. It is

supported by a very minor silence, which however might also be attributed to the word-final 't', but more importantly by a clear pitch reset and prolongation of pre-boundary 'not'. Hence, almost all signatures of a prosodic break, listed in Part 1 of the book, are present here. The second division can be made following the word 'you'. There is a clear glottalization as well as lengthening of the pre-boundary 'you' and pitch reset following it. In summary, we have divided this turn into seven chunks as is indicated in the IPA transcription in (4) above. We will now discuss these chunks in turn.

In the first unit 'this is not', there is a clear downstepped pitch accent on 'not' and additionally a higher target on the first word 'this'. Hence, we have two options: either there is one H+!H accent or two accents (H* on 'this' and !H* on 'not'). I would vote for the second option, as clapping with the speaker two times seems to fit better with the speaker's utterance than clapping just once, but both alternatives are very similar. The boundary tone is unmistakably a high plateau for the boundary (H-L%). The syntactic incompleteness of this utterance is characterized by the prosodic high plateau, also signaling incompleteness in this context.

In terms of other aspects in these chunks, the realization of 'is' is interesting. If you play just this word separately, you will notice almost schwa-like quality of the vowel clearly different from the 'i' type vowel we have in Slovak or Hungarian and thus the laxness of [ɪ] is further neutralized by the unstressed nature of this word. It is also interesting that word-final 't' of 'not' is unreleased despite the presence of a major boundary.

The second part 'I'm tired of doing this with you' is both syntactically and prosodically well formed and completed. It receives two main pitch accents ('tired', 'doing'), both of them downstepped: the first one (H+!H*) downstepping from 'I'm' to the first syllable of 'tired', and the second one a plain downstep high (!H*). The boundary tone is low (L-L%). The prosodic affiliation of the filled pause 'uh' is questionable and it is not clear if it should be included in this unit, in the following one, or if it forms a unit on its own. Due to the glottalization after 'you' I opted for including it in the following chunk.

Regarding the aspects of connected speech, there are two weak forms: [aɪm] is simplified to [ʌm] and [əv] to [əv̥] with almost imperceptible [v]. The preposition 'with' is also extremely shortened. On the other hand, the pronoun 'you' is realized in the strong, unshortened form due to its position at the end of an intonational phrase. The liaison between 'doing' and 'this' is most like a case of nasal place assimilation in which the velar [ŋ] changes to dental [ɲ] to assimilate to the following dental fricative [ð].

Segmentally, for this speaker, the interdental fricative at the end of 'with' tends to be produced almost as a stop, which is visible in the spectrogram as a release burst (vertical lines) after the 'white' region. Note also a significant aspiration of the initial [t] of 'tired'. This aspiration is a

secondary phonetic cue for the importance and emphasis of this word cued primarily through the pitch accent and accompanying lengthening of the accented syllable.

The third chunk of this turn, 'uh Steven and I are in a studio today' also presents noticeable variability in the speech rate. The first part, 'Steven and I are' is fast, 'in a' is slowed, and 'studio today' is back to relatively fast rate. All content words receive a pitch accent (Steven, I, studio, today), possibly also function words. The boundary tone is a plateau as no perceived lowering is present and thus H-L% despite the fact that pitch is rather low. However, the low levels of pitch may be attributed to the end of the phrase in which the pitch range of the speaker is severely lowered and thus even high targets are realized with a rather low pitch.

In terms of connected speech aspects, 'and' is significantly reduced to its weak form [ən] neutralizing the vowel and eliding the final [d]. The linking of 'in' and 'the' also includes a prolonged [n], which is most likely the result of progressive assimilation of manner in which the dental fricative changes to dental nasal [ɲ].

As mentioned in the introduction to Part 1, one of the basic forms of increasing awareness regarding the aspects of English pronunciation for non-native speakers of English is the comparison of producing a single sound in various environments. A nice example of this in this section is the production of consonant [d] in words 'studio' and 'today'. In the first case, the consonant is produced with only a minimal (or non-existent) contact of the tongue with the alveolar ridge, which results in so called 'flap' typical for American English. This can be observed in the spectrogram where the tongue approaches the alveolar ridge around 8.088 seconds but the spectrogram shows that the complete contact is not reached (the formants are continuously present). In the second case of 'today', the contact between the tongue and the alveolar ridge is significant and lasts at least 30 milliseconds (between 8.395s and 8.424s). The prosodic contexts in which these two realizations are present determines their distribution: a full-contact [d] is commonly present at the beginning of a stressed syllable while a flap is common in between vowels in cases when [d] starts an unstressed syllable. This is an example of a generalization (or a rule if you want) that speakers of American English acquired and unconsciously operate during speaking.

Finally, segmentally there are two more areas of interest. First, the pronunciation of /t/ in the word 'studio' is somewhat unexpected. The speaker is an American and we would thus expect very little palatalization of alveolar sounds preceding [u:]. Whereas for British speakers we expect pronunciation similar to [j] in the transition between an alveolar consonant and the following [u:] as in [nju:z] or [dju:rɪŋ] for 'news' or 'during', American speakers tend to produce this transition with very minimal or non-existent [j]: [nu:z], [du:rɪŋ]. Interestingly, the word 'news' is produced in the very next phrase and you may thus compare this with 'studio'. Despite the fact that LH is an American, we can hear [j] to some extent in her production of 'studio'.

Second, the vowel of the first syllable of ‘today’, normally a schwa, is completely de-voiced and we only hear a voiceless release of the initial [t] and no modal vibration of the vocal cords. This situation might be described as an elision of the unstressed schwa, but de-voicing is a more apt characterization since there is ‘something’ between [t] and [d] of ‘today’, and it lasts about 80 milliseconds according to the spectrogram, but this ‘something’ is not voiced and we see (and hear) only noise.

The next prosodic phrase is ‘uh having just received the news’. Before discussing accents, I hear a minor prosodic break after ‘having’, primarily signaled by pre-boundary lengthening but also a high phrasal tone (H-). The entire chunk is again produced in a somewhat slowed down speed, which gives the impression that every word is accented with an f0 plateau at the end signaling pragmatically that the message the speaker wanted to convey is not completed. We called this forward looking reference of non-low boundary tone. The last three words present a nice example of downstepping since their production is melodically as if descending down the stairs: ‘just’ is the highest (mean pitch for the vowels of the stressed syllable is 251 Hz), ‘received’ lower (196 Hz), and ‘news’ even lower (160 Hz).

There are two potential sites for elision of t/d that might happen in connected speech when a word-final t/d appears in between two another consonants. This environment is satisfied in ‘just received’ and ‘received the’: [dʒʌst rə'si:vɪd] and [rə'si:vɪ ðə ðə]. While in the first case, no elision takes place and we clearly hear (and can see in the spectrogram the whitish region of closure followed by the burst of energy) the [t] of ‘just’, the [d] of ‘received’ seems to be gone and we might characterize it as elided. This is yet another example of optionality in the aspects of connected speech in naturally producing speech.

Noticeable segmental aspects include the production of ‘having’ and ‘news’. First, notice how [h] of ‘having’ is voiceless, which contrasts with the production of this sound in Slovak and was also mentioned in section 2.1.1. The clear accented production of this word and its pre-boundary position also allows for a nice observation of the final [ŋ] that shows no [g] closure whatsoever that many Slovak speakers of English tend to produce in –ing forms.

Just like the initial and final sounds of ‘having’, the word ‘news’ is also interesting at the edges. The initial [n], as we have already mentioned when discussing the previous intonational phrase, is produced with a minimal perception of [j]. The final sound here is also interesting and instructive. Our ‘rule’ of the pronunciation of plural and past tense suffixes tells us that the voicing; i.e. if we say [s] or [z] in the plural, or [t] or [d] in the past tense, is determined by the voicing of the word-final sound. Hence, in the word ‘news’, the plural should be pronounced as [z] since a voiced vowel precedes it. Yet, the voicing of these word-final fricatives is so individual and context dependent that the ‘rule’ is often not corresponding with actual pronunciations of native speakers. In the case of ‘news’, we see that the final fricative starts as a voiced [z] at about 10.927s but switches to a

voiceless [s] at about 10.986s. But even this short period of voicing is sufficient to convey the perception of voicing in the final consonant, or more precisely, the perception of difference if the word ‘news’ was compared to a word like ‘noose’ with word-final [s].

The following prosodic unit is ‘of the death of Prince’. It has two clear high pitch accents (‘death’ and ‘Prince’) with a final low boundary tone L-L%, i.e. a prototypical falling contour. In the British tradition, this tone unit would be characterized with a typical ‘hat pattern’ in which the syllable receiving the first pitch accent, the head, is high, then slight declination of the non-accented material and f0 is high again on the last pitch accented syllable of the unit. Notice also that this low boundary signals completion despite the fact that the full stop is not present and ‘who was fifty seven’ is added. This low boundary signals that we do not really need the information that prince was 57 for adding the information of his death to our mutual beliefs. The continuation rise here might be confusing since it might signal that there are several people called ‘Prince’ and we need his age to identify which one is meant.

In connected speech, we can observe the weak forms of both realizations of ‘of’ with a schwa, but the difference arises in the final fricative consonant. While the first ‘of’ ends clearly in the voiced [v], which is expected in the canonical, or dictionary, production of this word, the second ‘of’ has a more voiceless frication and resembles [f] rather than [v]. The reason behind this difference can again be found in the context. While the first ‘of’ is followed by a voiced consonant, the second ‘of’ is followed by a voiceless [p]. Hence, the noticeable de-voicing of [v] to [f] in the second ‘of’ can be characterized as a case of regressive assimilation of voicing where the voicelessness of the initial [p] causes de-voicing of the preceding word-final, normally voiced, [v].

In terms of the segmental aspects, there are two interesting points. The first one is a rather non-standard production of the dental (or inter-dental) frication starting ‘the’. Normally, we would expect frication throughout this consonant, but in this realization, the consonant seems to be produced with a closure between the tongue tip and the teeth, which can be verified by observing the whitish column during this closure in the spectrogram (the interval 11.732-11.761 seconds) followed by a small, but observable release burst. Both the duration of this closure and the intensity of the release burst are significantly lower than the production of [d] in the following word ‘death’, but these spectral characteristics clearly show that the speaker did not produce a clear fricative [ð] but a [d] with a short and light tongue contact. This realization of [ð] is quite common in various dialects of the US or Britain, for example in Brooklyn. However, this is also very common among the Slovak speakers of English since [ð] that is not present in the phonemic inventory of Slovak is replaced by the most similar sound [d]. Slovak speakers of English who tend to produce [d] in these words should still aim for the fricative [ð]. This is because their [d] in these cases tends to be the same as [d] in words like ‘death’, which we see is not how native speakers talk, and their non-native accent is thus very obvious.

The second segmental aspect worth mentioning is the devoicing of [r] in the word ‘Prince’. A significant portion of [r], around 100ms between 12.322 and 12.425) is voiceless compounded with breathy noise. As we outlined in section 2.2, this type of de-voicing is linked to aspiration since it affects the sonorant consonants /l, r, j, w/ when they appear after syllable-initial /p, t, k/ in a stressed syllable (aspiration affects the vowels following /p, t, k/). Note that this kind of de-voicing is different from assimilatory de-voicing characterized for [v] of ‘of’ preceding ‘prince’. This is because the assimilatory de-voicing can only affect obstruents (plosives and fricatives) not sonorants (nasals, liquids, semivowels) while aspiratory de-voicing affects liquids and semivowels. This word is a useful model example for practicing this type of de-voicing for Slovak speakers.

The next unit of this turn is ‘who was fifty seven’. Melodically, there are two pitch accents (‘fifty’, ‘seven’). The accent on ‘fifty’ is clearly an H* but the second syllable included pitch raising and pitch fall again on ‘seven’. We have to somehow account for this high target and if we just labels with two H* accents, this peak would not be explained. Hence, the second accent H+!H* is a better choice explaining the peak on the second syllable of ‘fifty’. Surprisingly, we observe a plateau at the end of the phrase signaling the prosodic incompleteness of the thought despite the syntactic completion and the full stop. This shows nicely the variability of spontaneous speech and if this was a different genre, for example read broadcast news, a final fall would most likely be used.

In connected speech, we may say that ‘was’ is produced a bit weakened but not with a clear schwa, and the final [z] is somewhat devoiced toward the end due to the following voiceless [f]. Hence, another example of partial regressive voicing assimilation resulting in a partially de-voiced [z]. The long vowel of ‘who’ is somewhat shortened.

Segmentally, I should mention the production of the second syllable of ‘seven’. We see in the spectrogram that the schwa is not reduced and despite sounding very nasal, the vowel is definitely there. Hence, a possible production of this word with a deletion of schwa and the syllabic [sevŋ] does not take place. This is likely due to the fact that this word is prosodically accented and also in the lengthened pre-pausal position, and thus simplification and shortening has no reason to be used. This case exemplifies very nicely how difficult it is to tease apart segmental and suprasegmental aspects and how closely they are linked.

Finally, the exclamation ‘what is this year man’ concludes this turn. The speech rate is slowed down, which adds gravity and incredulity to the statement, and I already mentioned a possibility of a minor prosodic break after ‘what’.⁴ Additionally, all words but ‘this’ receive a pitch accent, and we again see a down-stepped pattern with the highest pitch on ‘what’ and then successively lower accents on the following words. Since all accents are clearly rising, we label L+H* for the

⁴ If we label with a minor phrase boundary after ‘what’, it would be an L-, and thus the accent on ‘is’ would be H* and not L+!H* since the low target would be explained by L- and we do not need to cover it in the pitch accent, and !H cannot start a new phrase.

first one and L+!H* for the successive ones. The end of the phrase is not clearly a low target, and thus the typical fall for wh-questions, statements, and exclamations, and we could characterize it as a low plateau.

Due to the prosodic emphasis on each word, words are realized somewhat separately and thus, no aspects of connected speech operating at the boundaries of the words are saliently produced. A nice example of this is the link between the first two words. The speaker produces a significant glottal stop after the release of word-final [t], which clearly separates the two words (and may serve as additional support for the boundary following ‘what’). If the two words were linked, we would expect a typical American flap instead of the produced [t].

The segmental realization is interesting for non-native speakers of English in the production of ‘what’ and ‘man’. The first sound of ‘what’ is normally [w], but in some dialects, especially the South and with older speakers, people can still produce a voiceless variant of this sound sounding like [xw]. The extra emphasis on this word produced by this speaker seems to lead to the realization somewhat close to this [xw], and a closer examination might correspond even to [uhu] realization of what normally is [w]. Also note that the final [t] is not released with a burst, which would be common in Slovak, but with a very light release of the tongue from the alveolar ridge, possibly due to the presence of the glottal stop.

The second interesting word is the final one: ‘man’. We can observe a typical diphthongal production, which is nicely observable due to extra lengthening resulting from the pitch accent and the pre-pausal position. Additionally, the vowel is significantly nasalized, which you can nicely hear if you play just the [ma] portion of ‘man’. This is thus a nice example of the allophonic process of vowel nasalization that affects stressed vowels that are followed by a nasal coda consonant in the same syllable.

(5) ST: Agrh I mean looking over a list of musicians who’ve died this year alone. We didn’t even talked about Merle Haggard on the show.

x: aɪ mi:n / 'lʊkɪŋ 'oʊvəɪ ə lɪst əv mju'zɪʃənz həv daɪd ðɪs jɪə ə'loʊn // wɪ 'dɪdŋ
'i:vən tɔ:kt ə'baʔ mɜ:l 'hæɡəɪd ɑːŋə ʃoʊ //

The first grunt, or loud communicatively meaningful in-breath, of this turn is interesting in that its transcription into words, and even into IPA is very problematic. Yet, its pragmatic meaning is very clear and it expresses displeasure, even disgust. This example shows the richness of communicative behavior and challenges that the link between the phonetic form and pragmatic meaning creates for researches. We will include it with the following material prosodically despite the fact that there is some evidence of disjuncture.

The division of this turn into prosodic units is relatively unproblematic. There is a clear silent pause signaling a major boundary after ‘mean’ and no clear indication of further divisions until the next prolonged silence following ‘alone’. The material following this silence is also fluently connected into a single unit without further subdivisions.

The first prosodic chunk is thus ‘I mean’. Its flat intonational contour suggests that it is used as a filler, or a hedge, most likely to plan the next utterance. We label with a prototypical plateau H* H-L%.

We observe that the canonical diphthong of ‘I’ [aɪ] is simplified in the process called monophthongization. Actually, we hear it if we listen to ‘I’ separately but if we played ‘I mean’ together, the diphthong is there. In some dialects, especially the Southern US and African-American Vernacular English, this monophthongization is a salient dialectal feature. In this context, however, the diphthong is partially monophthongized not because of the dialectal speech of ST but for prosodic reasons as the pronoun is realized in its typical weak form. The final [n] of ‘mean’ is also significantly lengthened and quite uncharacteristically released into a clear schwa-like vocoid, which could be thus transcribed as [mi:nə]. This might plausibly be a beginning of a filled pause ‘uh’ that is, however, truncated right at its onset.

The next intonational phrase is ‘looking over a list of musicians who’ve died this year alone’. It presents a nice example of the rhythmical prosodic aspects of English. The phrase receives five major pitch accents on the lexically stressed syllables of ‘looking’, ‘list’, ‘musicians’, ‘died’, ‘alone’, again with a rising first accent and then successively down stepped ones in the following words finished with a low plateau observed already before. If you try to clap/tap on these syllables as you produce this chunk, you notice how you need to shorten the remaining unstressed syllables to fit them within the rhythmical pattern provided by the stressed syllables. Particularly challenging is not accenting ‘year’ that many non-native speakers reading this turn might accent. The type of the accents is also a challenging decision since the ones toward the end are rather low in the ST’s pitch range, but compare word ‘alone’ with the word ‘year’ in his previous turn, and you clearly see that the former is still higher than the latter. Hence, I opted for essential H-targets albeit in the decreased pitch range of the speaker the pitch is actually quite low.

The aspects of connected speech are not numerous here. Preposition ‘of’ is produced in the weak form with a schwa but note that it is not significantly shortened. On the other hand, the contraction ‘who’ve’ is extremely shortened and the vowel of ‘who’ is not [u] but a schwa. An interesting case is the link between ‘died’ and ‘this’. Word-final [d] is an alveolar plosive while word-initial [ð] is a dental fricative. Both of these sounds are produced with the front part of the tongue as the major articulator. Given that the spectrogram shows both the interval of complete closure and an interval of light frication, it is plausible that the speaker produced a complete closure at the dental area and then continued with the fricative. He thus realized the transition between these two words through

a case of regressive assimilation of the place of articulation: word-final alveolar changed to a dental plosive sound to have the same place of articulation as the following word-initial dental fricative [ð]. Finally, note that the speaker did not produce place assimilation in the following transition between ‘this’ and ‘year’. Many speakers would produce this phrase with a slightly retracted [s] to assimilate with the following palatal [j], which would result in [ðɪʃ jɪəɪ]. But this speaker in this context did not produce such assimilation.

Segmentally, the very first word is a good model example for practicing the production of [ʊ] and [ɪ], which are both problematic for many Slovak speakers. Similarly, [ɪ] in ‘list’. As mentioned already in 2.1.2, the lax vowels [ʊ] and [ɪ] are difficult for Slovak speakers since most of them formed a habit of producing them like short Slovak [i] and [u] for example as found in Slovak words like ‘luk’ and ‘list’. The second case is extremely instructive since the same word, ‘list’, has a markedly different vowel when produced as a Slovak word and when produced as an English word. This transition, or separation of the two phonemic inventories when speaking these two language lies at the core of acquiring English pronunciation with minimal accent due to the mother language. Finally, we notice again the significantly de-voiced production of [d] that starts a stressed syllable in ‘died’. This is quite a salient difference between the native speakers and Slovaks speaking English.

The final intonational phrase in this turn is ‘We didn’t even talked about Merle Haggard on the show’. The intonational contour is characterized by a very high pitch on ‘we’ and somewhat limited pitch range and flat melody in the rest of the phrase. The same thing discussed in the previous unit regarding rhythm can be said about this unit as well. There are four major pitch accents falling on ‘we/didn’t’, ‘talked’, ‘Haggard’, and ‘show’, and rhythmical regularity is achieved by fitting the unstressed syllables into rather evenly-spaced prominences of these four pitch-accented syllables. While the category of the accent is clear (L+H*), the area of uncertainty is whether ‘we’ or ‘didn’t’ receives the first pitch accent. Both have high f0 peaks and ‘didn’t’ would seem natural for prosodic highlighting expressing the contrast with the assumed belief that Prince’s death is a lone event. However, I felt that the accent on ‘we’ fits better with the speaker’s realization. I would speculate that the speaker’s beliefs were that some other people have talked about it, and this other death preceding Prince’s death is general knowledge, but they have not mentioned it yet. In this interpretation, L+H* expressing contrast on ‘we’ makes sense. Also, the low plateau of the previous phrases is present here as well indicating that the speaker wanted to continue.

The rhythmical aspects are achieved by highlighting the stressed syllables as well as making the unstressed syllables prosodically weak (short, low volume, non-high pitch). This falls under the domain of connected speech. Note how four syllables of ‘didn’t even’ are almost as long as the single syllable of ‘talked’. This short interval that the rhythm allows for the production of these two words affects not only the duration of the syllable nuclei but also how the words are linked

together. For example, the spectrogram does not show any evidence for the final [t] of ‘didn’t’ and thus supports the elision of this consonant despite the fact that the following word starts with a vowel, and thus presents an environment where elision usually does not take place. Also, the schwa of the weak syllable is not present and the nasal becomes syllabic: [ˈdɪdn̩].

Next, if you play just the word ‘about’ you notice that the initial schwa is almost missing, the final [t] cannot be heard either, and the diphthong [aʊ] is simplified to just the first part. Hence, we could transcribe the actual production of this word in this context as [ba] with final glottalization replacing [t]. Compare this now with selecting ‘talked about’ together where you probably ‘hear’ the schwa. This is the strong expectation we have to actually ‘hear’ the schwa but the speaker does not really produce it.

Finally, the linking of ‘on’ and ‘the’ is produced [ɑn̩n̩ə] where the two [n̩] sounds are dental. Hence, what happened is that first, regressive assimilation of place changes the alveolar [n] to the dental one due to the following dental fricative, and then progressive assimilation of manner takes place in which the dental fricative [ð] changes to dental nasal [n̩]. Importantly, as with many cases of linking two words with identical segments, this is not the case of simply eliding [ð].

Segmentally, the schwa of the second syllable of ‘even’ is still present despite the high demand on time and the possibility of producing this word as [i:v̩n̩] with a syllabic [n̩] sound. But the same does not apply to ‘didn’t’ where the schwa is elided to [dɪdn̩]. For Slovak speakers, the aspiration of [t] in ‘talked’ is very salient with the duration of about 83 milliseconds. This is mostly due to the prosodic prominence of this word discussed above. Aspiration, together with lengthening, high pitch and increased volume, also participates in prosodic highlighting of words. Finally, note also the voicelessness of the initial [h] of ‘Haggard’, which is another difference in the phonetic realization between English and Slovak in that English [h] is normally voiceless and becomes voiced only between vowels while Slovak [h] is never voiceless.

Since this is another longer turn, it would be good to practice the three chunks separately and then combine them into the entire turn aiming to achieve the prosodic flow of the native speaker particularly by accenting the prominent words, neutralizing and shortening the non-prominent ones and at the same time paying attention to the segmental aspects.

(6) LH: no we din’t.
[noʊ] wɪ ˈdɪ[dənt] //

The first thing to notice is that this turn starts before the preceding turn ended. The overlap is short, around 120 milliseconds. Several studies have argued that in dialogues, people strive to achieve smooth transitions in which there is neither a very long silence nor prolonged overlap between the

turns (originally in Sacks et al. 1974). The production data in large corpora show that the variability in turn transitions is great and interactions with prolonged silences or long overlaps do take place; nevertheless, smooth transitions, i.e. transitions with either short silences or short overlaps, are by far the most common pattern. It is a subject of extensive ongoing research to better understand the processes that allow interlocutors predicting when the current turn will end. Notice that not only in the case of overlaps, but also in the case of short pauses, the next speaker has to start planning his/her utterance well before the current speaker finished his/her turn. This is because we know from neurolinguistic and psycholinguistic research that planning the semantic content, syntactic form of a message, and its encoding into the phonological and ultimately phonetic forms takes at least 700 milliseconds (Levinson 2016). Hence, for most common turn transitions, speakers have to have a very good prediction of when the current speaker will end his/her turn to avoid long pauses or overlaps at turn transitions. It is thus instructive to observe also potential cues that might facilitate these predictions and notice how effortlessly we are able to navigate these in our own spoken interactions both in the native language and when using non-native English. We observe that in this short turn, both transitions (ST to LH, and LH back to ST) are realized with a short overlap. In this genre, a rather informal dialogue between two people on a relatively similar levels, overlaps are rather common and you will see several examples of this in this excerpt. In some other excerpts in which the genre might be more formal, or the roles of interlocutors are clearly on a different level (e.g. an interviewer and interviewee), overlaps might be less frequent.

Let us now turn back to the prosodic and segmental aspects of this turn. First, notice how LH lowered her intensity of voice. This is one of the cues to signal that she only wants to support, or acknowledge, previously produced information and does not really want to add much new content into the interaction. LH managed to produce this acknowledgment during the breath taken by ST between his utterances. Intonationally, weak pitch accents are produced on ‘no’ and the first syllable of ‘didn’t’. It is not clear what pitch target LH planned to use at the end since the next speech from ST results in cross-talk and it is impossible to assign clearly this target. Standard ToBI has the option of using X in these cases: X-X%. I opted for L* L-L% in this case. The speaker clearly used the weak form of ‘we’.

- (7) ST: Uh Vanity uh was one that hit me really hard.
 [a] 'vænəri / ɜ: wʌz wʌŋðə? hɪ? [mi: 'ri:li] hɑ:ɪd //

The most straightforward way of dividing this turn in chunks is to posit a boundary after ‘Vanity’ due to tonal and temporal markers of the boundary. The first unit in this turn thus includes the first two words: ‘uh Vanity’. The filled pause, or sometimes also called (conversational) filler is unaccented and it might be there to stall for time since ST was not expecting LH to add something and was ready to continue right away but used the filler when he noticed LH added the

acknowledgment. The second word receives the major pitch accent. This word presents a nice example of a rising pitch accent followed by a so-called continuation rise to mark intonationally the end of this unit. In the ToBI conventions, this would be L+H* for the pitch accent followed by L-H% for the boundary tone.

There are no aspects of connected speech since the only transition between the words is a transition between a vowel and a consonant and all such transitions are commonly produced with a simple linking of the two sounds.

Segmentally, we discuss here two things. The first one is the quality of the vowel in the filler ‘uh’. Commonly, this filler is realized with the most central and neutral vowel schwa, as also shown in the canonical IPA transcription above. However, this is not what speaker ST produces here. The vowel is more similar to a low central [a] rather than to a schwa [ə]. Second, we mention the example of American flap as the consonant of the final syllable (recall the pattern for this sound: t/d in between vowels starting an unstressed syllable), and the quality and prolongation of the final vowel that is clearly [i] and not [ɪ].

The other intonational phrase of this turn is ‘uh was one that hit me really hard’. The most interesting melodic aspect of this phrase is the presence of a major pitch accent on ‘me’. In a neutral context, the default accenting would fall on ‘hit’ and not on a rather predictable personal pronoun ‘me’. Therefore, this is a clear example of the general feature of English (and other languages as well) that virtually any word might receive a pitch accent or be prosodically highlighted otherwise, if the pragmatic context requires it. So let us consider why this word, that is normally prosodically weak and might be commonly produced in its short weak form [mɪ], is accented in this context. What is the speaker’s intention signaled by this prosodic means? Plausibly, the speaker wants to contrast him against other people who were not particularly hit by the death of Vanity. He thus wants to convey the idea that Vanity’s death in connection with Prince’s death, is crucial for what he wants to express, in contrast to other people who might not assign too much importance to Vanity’s death in this context. Therefore, despite the overlap form LH and the resulting unreliability in pitch tracking, we may safely assume that the pitch accent on ‘me’ is L+H*.

Note how a similar pragmatic meaning may have been signaled also by putting a strong pitch accent on ‘really’ and not on ‘me’, but this utterance would convey a slightly different meaning in that only the mental state of current speaker would be described as being REALLY hit by the death of Vanity whereas the accent on ‘me’ provides this, and on top of it also the contrast to other people and the implication that they were not hit hard as they probably should have been.

The processes of connected speech are observable in the phrase ‘one that hit me’. All three transitions involve a consonant on both sides of the word boundaries. In the first case, we see an already discussed process of regressive nasal place assimilation in which the [n] of ‘one’ changes

its originally alveolar place to the dental one to be similar to the initial consonant of ‘that’. This next word ‘that’ is produced in its weak form since it functions here as a conjunction (or a complementizer syntactically) rather than the demonstrative pronoun that is commonly realized in its full form. The transition between ‘that’ and ‘hit’ is realized in significantly neutralized final [t] in which the contact between the tongue and the alveolar ridge does not take place and the speaker produces slight glottalization (a constriction in between the glottal folds rather than in the mouth). This glottalization is much more strongly produced in the transition between ‘hit’ and ‘me’ and we can also see it in the spectrogram with vertical striations visible between 25.738 and 25.785 seconds.

In terms of the segmental realization of this unit, the quality of the initial filled pause is interesting, particularly in comparison to the quality of the filler in the preceding prosodic unit. We have said above that the first filler of this turn has quality of [a] but the second one is clearly a centralized long schwa [ɜ:]. You can very nicely observe the difference by clicking on the intervals with these fillers and listening to them one after another. Also the voice is worth mentioning here. While the turn-initial filler is produced with mostly modal voice, i.e. regular vibration of the vocal cords, the second filler has a much creakier, or pressed, voice quality. Although we do not spend much time discussing voice quality features in this book as they commonly convey individual, emotional, or attitudinal aspects of the message rather than meaningful pragmatic differences, it is good to be aware also of this enormously rich area of expressing personal stances or states through spoken language.

Another interesting difference can be found by comparing the initial sounds of ‘hit’ and ‘hard’. While the first one is completely voiceless, as discussed several times above, the second one can be characterized as somewhat voiced. This is probably due to the segmental context in which these two [h] sounds appear. While in the first case, [h] is preceded by a canonically voiceless sound [t], realized with a brief voiceless glottal stop, the [h] of ‘hard’ is preceded by a voiced vowel and it is thus between two vowels in a smooth juncture, which is the environment where otherwise voiceless [h] might be voiced. Finally, words like ‘hard’ are instructive for those who want to explore the differences between British and American pronunciation since the coda [ɹ] sound in British English is typically not produced whereas in American English it significantly affects the quality of the preceding vowel, through so called r-coloring, and might also be fully, or partially realized as a consonant.

(8) LH: yeah
[ja:]

Regarding the realization of ‘yeah’ by LH, its timing and the quality of the vowels will be discussed here. This pragmatic agreement very likely expresses the LH’s idea that the contrast between ST

and other people who were not hit hard by Vanity's death, discussed above, does not include her, and she, like ST, also thinks that Vanity's death is relevant in the context of Prince's death and that she was also hit hard by her death. In this sense, H* adds this new information because she was assuming it was not part of the mutual belief, or that ST was not aware of this. The slightly higher level of intensity and longer duration in this turn support this interpretation as she also wants to contrast her position with the other people, and this contrast is expressed through voice intensity and duration. Imagine the pragmatic meaning of the same word but produced in low intensity and with short duration.

It is rather common for this supporting agreements/acknowledgments to overlap with other speaker's turn in a similar way that supporting head nods might be expressed in face-to-face conversations during the other speaker's turns. Hence, timing is not aligned to fit in the breaks of the interlocutor's speech, as other turns typically are.

The quality of vowel in this turn is also worth mentioning. 'Yeah' belongs among several words with highly variable and individual realizations. This is partly because it is functionally loaded, i.e. might express several pragmatic meanings such as (degree of) agreement, reservation, closure, etc., and these functions might be cued partly through different phonetic and prosodic realizations of the word. In this particular case, we do not really observe the diphthong [eə] or another common realization [æ], but rather something close to a prolonged low mid [a:].

- (9) ST: She was a collaborator of Prince's who died earlier this year.
 ʃi: wəz ə kə'læbə,reitəɪ əv 'prɪnsɪz hu daɪd ['ɜ:liəɪ ðɪs jɪəɪ] //

The entire turn here is realized as a single prosodic unit. The most highlighted word is 'collaborator' with a rising pitch accent (L+H* in ToBI), the other accented words are 'she', 'Prince's', 'earlier', and 'year', all with H* accents. The intonation at the end of the unit is difficult to assess due to the cross-talk from the other speaker but it sounds as a plateau signaling the speaker is not finish delivering the intended message.

The initial pronoun 'she' is produced with somewhat longer [i] due to the initial prominence realized with a pitch accent and we can thus describe this production as a strong form of this pronoun. The other candidates for weak forms in this turn such as 'a' and 'of' are produced as weak forms with schwas. Additionally, also 'who' has a very short vowel. We observe some, but not complete, devoicing of [v] in 'of' due to the regressive assimilation of voicing triggered by the following [p].

In segmental realizations it is interesting to look at two words we have analyzed in the previous turns: 'Prince', and 'died'. LH's production of the initial cluster [pr] in 'Prince' at 12.228 seconds

is quite different from this ST's production of the same cluster. Primarily, the aspiration, realized as r-devoicing, is markedly less salient and shorter in this ST's production than in the previous LH's production. This is very likely linked to the degree of prosodic emphasis on this word: LH wanted to make the word very prominent because it was mentioned for the first time and makes the main topic of the dialogue. This importance is signaled both prosodically with a very pronounced pitch accent as well as segmentally with aspiration/devoicing. On the other hand, ST does not need to highlight this word so much since it has already been introduced to the discourse and thus can be considered old, or given, information. Another word is new in the discourse ('collaborator'), and this is prosodically very salient. One of the functions of intonation is to mark this given/new status of words in discourse and in this comparison of the two words we see a nice example of that.

In a similar vein, ST has already produced the word 'died' before at 19.862 s and we said that the realization of the initial [d] was de-voiced. In the current realization of 'died' the initial [d] is almost fully voiced. This is another segmental aspect tightly linked to the prosodic aspects. In this unit, 'died' is not prosodically prominent, does not receive a pitch accent, and thus the production of this syllable might be weakened. Full voicing of [d] is one segmental aspect of this weakening. Note that this realization is also by almost one third shorter at 211 ms than the previous accented realization at 19.862 that was 292 ms long.

Finally, consider the word 'collaborator'. The word stress falls on the second syllable and we see a nice realization of the stressed [æ]. However, the schwa of the first syllable is completely devoiced due to the aspirated release of the initial [k] despite the fact that the first syllable is not stressed and aspiration thus should not be very salient. We might speculate that the prosodic prominence on this word based on the pragmatic and discourse needs affects the realization of the unstressed syllable so that these are also a bit more strengthened, and this strengthening is responsible for the somewhat greater aspiration.

(10) LH: My I'm glad you mention that.
[maɪ aɪm glæd ju:] 'menʃən ðæt //

This unit starts with a rather extended overlap with the previous speaker of approximately 720 milliseconds. It is plausible that LH thought that ST will end with 'She was a collaborator of Prince's' but did not expect the complemented clause 'who died earlier this year'. If there was no additional modifying clause, the LH's turn-initiation would come roughly after 270 ms after the completion of the ST's turn, resulting in a perfectly smooth transition.

Due to this extensive cross-talk, this unit possess a challenge for analysis. We might mention that there two clear pitch accents ('my', 'glad'), and I am less certain about the pitch accent on

'mention'. There are also weak forms of 'I'm' and 'you', and a strong form of 'that' due to both it being a demonstrative pronoun and also being at the end of a prosodic unit.

(11) ST: Also at fifty seven and that's the first thing I wanna get into with Prince is just how frequently and extensively he worked with and celebrated and provided songs for this incredible array of woman musicians. Uh he had women in his band throughout his career, you know, looking over a list of of his collaborators, and not only women he performed with or produced but women he wrote songs for. He wrote "I feel for you" by Shaka Khan, he wrote "Nothing compares to you" which Sinead O'Connor covered uh

This is the longest turn of the excerpt with almost 30 seconds of speech. It contains several prosodic units that we will discuss in turn. But obviously, a major challenge is to divide this long turn into prosodic chunks. A very good approach with such long turns is to look into Praat and identify silent pauses filled with in-breaths. I can count eight of them following these words successively: 'seven', 'is', 'extensive', 'for', 'musicians', 'produced', 'wrote', and 'covered'. This already gives us much more manageable chunks, called breath groups, to inspect if further division into prosodic units can be supported.

For easier exposition, I will take these breath groups and discuss further chunking with them and other connected speech and segmental aspects

(11a) Also at fifty seven
'ɔ:lsoʊ ə[?] 'fifti 'sevən //

This first breath group cannot be divided further and thus constitutes a single prosodic phrase is All three content words are accented with a downstepped accent on 'fifty' and the phrase ends with a medium plateau. This plateau is interesting since we would normally expect a fall in this phrase since the speaker adds new information that Vanity dies at 57, the same age as Prince was at the time of his death. The plateau plausibly signals the incompleteness of this turn and functions thus as a turn-holding signal in an effort to prevent LH from taking the floor as she did in the previous turn resulting in an extensive overlap. I thus speculate, that ST wanted to start a new and extended idea, and did not wish to be interrupted by LH at this time and used this intonational plateau to signal that.

There is only a single aspect of connected speech in this unit: the weakening of the preposition 'at'. Its realization can be transcribed as [ə[?]] and certainly not the strong form of [æt] used in the canonical transcription above. To explain the suggested transcription [ə[?]], the vowel is centralized and weakened to a schwa and the consonant is not produced in the oral cavity; i.e. the tongue does

not touch the alveolar ridge, but rather, the constriction is produced in the glottis; that means the vocal cords approach each other. However, even this glottal constriction is not complete as it does not result in a stop, which would be visible as a brief interval of mostly white space (i.e. no sound). Rather, what we see in the spectrogram is an interval of about 50 ms (29.893-29.95) with now familiar vertical striations characterizing glottalization. Hence, the muscles in the larynx make the vocal cords tensed and approaching each other so that they cannot vibrate regularly, but not enough to cause the complete glottal closure, which would prevent their vibration completely.

Segmentally, the link between prosodic and segmental aspects can be observed in the realization of ‘seven’ with a seemingly surprising full realization of a schwa in the second syllable and not a syllabic consonant replacing the schwa.

(11b) And that’s the first thing I want to get into with prince is
 əŋ ðætʰs ðə fɜːɪs θɪŋ a ‘wʌnə ɡer ‘ɪntu / wɪð prɪns / ɪz //

The next breath group offers some evidence for further division and placing a boundary after ‘into’ and ‘Prince’. However, we should also consider the variability in speech rate. The breath group starts rather fast (nine syllables in about one second till ‘get’ including), but ends in a very slow hesitant pace (five syllables in almost two seconds for the last four words of the unit). Hence, the first part is almost three times as fast as the second part. This is probably a result of cognitive processes of thinking and planning of what to say and how to say it. Since the lengthening in ‘into’ is rather extensive, and a complete silent pause follows ‘Prince’, we posit minor prosodic boundaries despite the absence of overwhelming evidence from f0.

This unit is also starting a new discourse segment or in other words, a new topic of discussion. As a result, the planning the speaker makes involves an extended chunk of language, as is clear from the length of the turn. Another prosodic marker of this new discourse segment is the high pitch at the beginning of the unit. It is known that when we start an idea, our voice is rather high in pitch and loudness, the levels of pitch and loudness gradually decrease towards the end of the idea or a discourse unit. A new reset of pitch and loudness to high values takes place when a new topic starts again. The pitch reaches 267 Hz, which is well higher than the highest points in previous ST’s turns and compares to 305 Hz with which he starts this excerpt. Hence, in addition to various other functions of intonation we have exemplified so far, pitch level might be used to facilitate the perception of chunking the discourse into meaningful units and help interlocutors ‘reveal’ the speaker’s mental processing regarding his/her message planning.

Intonationally, we see a familiar pattern by now of a first high pitch accent (on ‘that’s’) and then a succession of downstepped accents on ‘first’, ‘get’. The slow tempo of the second part of this unit gives an impression that each word is prosodically highlighted, and it is a combination of full

vowels and prosodic boundaries after each word that produces this perception labeled with low pitch accents and low boundaries.

Linking of individual words into connected speech is clearly bound to the tempo changes and this linking is very tight in the first part and very loose in the second part. The first word ‘and’ is extremely shortened and produced with a weak form [ən̩]. The diacritic under [n̩] marks the dental quality of this sound resulting from the regressive nasal place assimilation in which the dental place of the first sound of ‘that’s’ triggers the change from the alveolar to the dental place for [n]. Interestingly, despite a major rising pitch accent on ‘that’s’, and thus low likelihood of weakening this word, the vowel is somehow centralized and definitely not realized as a full [æ].

In linking ‘first’ and ‘thing’ we see a clear prototypical example of elision in which a word-final [t] disappears as it is in the medial position of a CCC cluster [stθ]. The pronoun ‘I’ is also weakened and produced as a monophthong [a]. In the word ‘get’ we have another word-final [t] in fast speech. We saw previously that depending on context, it might be elided or glottalized, but here we see another process of flapping the [t] across a word boundary. Hence, the words are so closely linked that the flapping that applies within words applies also here.⁵

The aspects of connected speech in the rest of the unit are minimal due to the slow speech rate. I may mention glottalization marking the boundary between ‘into’ and ‘with’, and lengthening of the second syllable of ‘into’. Both of these suggest that there is a minor prosodic boundary between these words, as indicated by a single slash in the IPA transcription above and in the ToBI labeling in the textgrid file.

(11c) just how frequently and extensively
dʒʌst haʊ 'frikwəntli ən ɪk'stensɪvli //

The next breath group is ‘just how frequently and extensively’. The speaker continues with the slow tempo from the end of the previous unit. There is also a minor phrase boundary between ‘how’ and ‘frequently’ marked with extensive lengthening of ‘how’. All words but ‘and’ receive prosodic prominence through being pitch accented on their respective stressed syllables with essentially high accents.

In connected speech there is only a weak form of ‘and’ realized as [ən], and segmentally I again point to the voicelessness of [h] in ‘how’, and [i] quality of the final sounds of both adverbs.

⁵ I do not want to say that flapping is strictly word-internal process and this case means that ‘get into’ is one word. Rather, it has been known for a long time that the domain of flapping in American English is prosodically defined. Simplifying, it applies unless there is a salient prosodic break.

- (11d) he worked with and celebrated and provided songs for
 hi wɜ:ɪkt wɪð en 'selə,breɪrəd en prə'vaɪdəd sɔŋs fɔ:ɪ //

In this next breath group the speaker continues with a relatively slow delivery and thus almost every word is pitch accented, which is especially clear in the final three downstepping accents on 'provided songs for'. Despite multiple 'and' conjunctions and rather slow tempo, we can treat the entire breath group as a single intonational phrase with the forward looking plateau completion of the boundary tone similar to the previous boundaries.

The accenting of phrase-final 'for' leads to the clear example of a strong form production of this preposition. Similarly, the initial 'he' is produced in its full strong form. The two realizations of 'and' can be characterized as slightly but not completely, weakened since the final 'd' is not present, but the quality of the vowel is not schwa and resembles [e] rather than a schwa. Despite a CCC [ktw] cluster between 'worked' and 'with', the past-tense marker [t] is fully realized and thus no elision took place. Finally, regressive assimilation of voicing can be observed and the final sound of 'songs', which should be [z], is realized as [s] due to the effect of the following voiceless [f].

Segmentally, I can draw the attention to three things in this unit. The first one is the flapping of [t] in 'celebrated' since it starts an unstressed syllable. The second one is certain degree of devoicing of [r] in 'provided' despite the fact that the initial syllable is not stressed. And the third one is the very perceivable nasalization of the vowel in 'songs'.

- (11e) this incredible array of women musicians
 ðɪs ɪn'kredɪbəl ə'reɪ əf 'wɪmən mju'zɪʃənz //

In this breath group the speaker finally completes the sentence he started at the outset of this turn syntactically. This syntactic completion is signaled also prosodically by a very low pitch target at the end in the final fall (L-L%). In fact, the target is so low that the speaker produces a vocal fry and the creak is very perceptible in the final syllable of 'musicians'. Again, all content words of this unit are accented. Although I think it is ok to label this breath group as a single unit, an alternative hypothesis might be to posit a minor break preceding 'of'. and a high accent followed by a downstepped one could be used for marking prosodic prominences.

In connected speech, there is a single point of interest: the transition 'of women'. The reason is that we see unexpected behavior. Normally, the fricative of the preposition should be a voiced [v]. The word is followed by another voiced consonant [w]. Hence, there is apparently no reason why 'of' should be pronounced with [f] as we see on the spectrogram and hear in the sound file. The first thing to notice is that 'of' is not produced in a weak form with schwa, but the quality of the vowel is closer to [ɑ] than to schwa. Hence, this preposition is produced in its strong form and thus

prosodically highlighted. This highlighting comes mostly from lengthening due to slower tempo and a potential minor prosodic boundary preceding ‘of’. Hence, I would speculate that this devoicing of [v] is a marker of a weak prosodic break and not so tight linking in the juncture ‘of women’.

Segmentally, I can draw attention to an extended period of aspiration after the stressed [k] in ‘incredible’. However, this is partly also due to almost non-existent complete closure between the tongue dorsum and the velum and the stop [k] thus undergoes fricativization and sounds as [x].

- (11f) Uh he had women in his band throughout his career you know looking over a list of of his collaborators, and not only women he performed with or produced

ə hi: hæd 'wɪmən ɪn hɪz bænd // θru' aʊt hɪz kə' rɪəɪ // jə noʊ / 'lʊkɪŋ 'oʊvəɪ ə lɪst
 əv / əv hɪz kə' læbə'reɪtəɪz // ən nɑr 'oʊnli 'wɪmən hi: pəɪ' fɔ:ɪmd wɪð əɪ prə'dʌst //

The following breath group in (11f) above is rather long and inevitably has to be split to smaller prosodic chunks. We apply again the steps as before, checking first the silent pauses and then the pitch and temporal patterns to identify potential prosodic breaks. After inspecting the file and listening to it several times, I propose that this breath group is divided into three intonational phrases with major boundaries following ‘band’, ‘career’, and ‘collaborators’. All three junctures have clear lengthening of the pre-boundary syllables, a very clear tonal targets in high boundaries, and the last two additional have minor silent pauses reinforcing the perception of these boundaries. We will discuss these three prosodic units in turn below.

In the first phrase ‘uh he had women in his band’, we see a change in the intonational patterns the speaker uses. There are two main pitch accents on ‘he’ and ‘band’ and high rising plateau melody before the boundary. Hence, in contrast to the mostly downstepped contours and low plateaus before, here we see an initial high, followed by slight sagging of regular physiological f0 declination and then followed by another high accent on ‘band’. It is also instructive to observe the content words that are *not* accented, as the word ‘women’ in this unit. This de-accenting is another example of the discourse function of intonation in that the new information is often accented while the old information might be de-accented. The fact that the interlocutors talk about women around Prince has been established before and thus belongs to the common ground of the speakers, and does not have to be accented. This is not to say that words conveying old/given information are never accented, but the reason for their accenting might fall under rhythmical reasons for example.

The connected speech aspects are closely linked to the segmental aspects in this unit since we concentrate on the initial sounds. Three words start with the [h] sound: ‘he’, ‘had’, ‘his’. Only in the first case, however, the [h] sound is voiceless and in the other two cases [h] is voiced. This is

due to tight linking in ‘he had’ and ‘in his’ with no perceivable boundary while there is a potential minor break between ‘uh’ and ‘he’, accompanied also by slight glottalization. Finally, the [b] sound, although phonologically characterized as ‘voiced’, is realized, as in several cases of [d] discussed above, as a devoiced unaspirated [p]. This is again in part due to the great prosodic prominence placed on this word with a strong pitch accent, and this melodic prominence is complemented also with a strong segmental realization of the initial sound.

The next intonational phrase of this breath group is ‘throughout his career’. It has an identical melodic realization with a rising final accent on ‘career’ followed by a high plateau as the end as the previous unit. The initial word, despite normally receiving word stress on the second syllable, is realized with both syllables roughly equally prominent and the final [t] is almost not there, maybe neutralized into a brief light glottal stop. The word ‘career’, also stressed on the second syllable, is realized with quite strong aspiration of the [k] in the first syllable, probably due to high prosodic prominence of this word.

The third prosodic phrase of this breath group is ‘you know, looking over a list of of his collaborators’. An alternative chunking is possible and the phrase can actually be divided into three smaller units with minor breaks after ‘know’ and in between the two ‘ofs’. The first break might be supported by a salient fall in f_0 and thus the presence of an L- phrasal tone. In support of the second break, it is signaled not only by the repetition and a short interval of silence but also by the difference in the duration of the two prepositions. While the first one is about 340ms, the second is about 100 ms shorter. This prolongation is a signal of a minor boundary. Cognitively, it is likely that the speaker could not find the proper word for ‘collaborator’ right away and this repetition is used to stall for time until the desired lexical item is accessed. There are three major pitch accents: ‘looking’, ‘list’, and ‘collaborators’, and a very weak prominence on ‘you’.

Other notable realizations in this unit include the weak form of ‘you’ and the masking of schwa with aspiration in the first syllable of ‘collaborators’. The syllable is again unstressed, and thus should have only weak aspiration, but it does receive a major pitch accent and thus also the segmental realization is strengthened.

The fourth, and final, unit of this breath group is ‘and not only women he performed with or produced’. This unit is again fairly regular rhythmically with four prosodic prominences (‘not’, ‘women’, ‘performed’, ‘produced’) and a slight local slow-down on the pronoun ‘he’. The speaker signals he’s not finished completing the thought by a mid-range plateau intonation at the end of the unit.

The initial ‘and’ is weakened to [ən]. The next transition ‘not only’ shows flapping of [t] across the word boundary in a very tight transition as we saw on several occasions above already. Additionally, the local slowdown on ‘he’ mentioned above is responsible for the strong form realization of this

pronoun with [i:] rather than the weak form [ɪ]. Finally, the conjunction ‘or’ is extremely reduced to a very short schwa.

Segmentally, consider the first syllables of ‘performed’ and ‘produced’. They both include the same segmental make up: [p], schwa, and [r], but what is different is the affiliation of the r-sound. While in the first case it provides the r-coloring as it stands in the syllable coda, in the second case it is part of the onset cluster. If you play just these two intervals (51.631s - 51.803s for ‘performed’ and 52.527s - 52.744s for ‘produced’), you would notice the small difference that the affiliation of [r] causes.

(11g) but women he wrote songs for he wrote
bə² 'wɪmən hɪ ru:t sɔŋz fɔ:ɪ / hi: ru:t //

The next breath group in (11g) above shows a nice challenge for further chunking. Mostly, the difficulty lies in the fact that this material shows a nice example of differences between syntactic/semantic structures on the one hand and the prosodic realization on the other. Many times, the breaks and boundaries signaling the prosodic chunking of speech coincide (more or less) with structural boundaries determined by syntax (phrases, clauses) and semantic/pragmatic structures (information chunks, pragmatic moves (request, questions,...)). However, this link between prosody and the other structures is never perfect and areas of disagreements are ubiquitous in everyday conversational speech. Specifically, we would expect a prosodic break at the completion of the syntactic clause ‘but the women he wrote songs for’. Yet, the break the speaker makes here is much weaker than the break he makes after ‘wrote’ in the middle of the next syntactic clause.⁶ I will not go into the aspects that affect prosodic chunking and speech planning in detail here but I just wanted to show that there are many other aspects beyond the syntax and semantic/pragmatic structures that influence the way speakers chunk their stream of speech into units.

After this short ‘tangent’, we may continue with the analysis of this breath group. As mentioned above, there is a prosodic boundary between ‘for’ and ‘he’. This boundary is signaled by a low pitch target (and thus primarily falling contour of the first part) associated with a very salient interval of creakiness physiologically stemming from this low target. This low target provides the feeling of completeness, and signals syntactic completion while the plateau of the following ‘he wrote’ signals pragmatic incompleteness. Hence, despite what I just said above, about the mismatch

⁶ This break also represents a syntactic boundary between a verb phrase and a noun phrase and thus a potential for a prosodic break. Moreover, to make a prosodic break here is totally natural. The point I am making is that the break before ‘He’ is much stronger than after ‘wrote’ in terms of syntax and that this relationship between these two syntactic breaks is not reflected prosodically. And that this mismatch between syntax and prosody is interesting to study and opens up space for hypotheses for other linguistic and extra-linguistic hypotheses (e.g. the speaker needed to breathe, wanted to think better of what to say, etc).

between prosody and syntax/semantics, there are also prosodic cues that do match syntax/semantics. I could only re-iterate how fascinating, yet complex and multivaried, the study of the prosodic aspects of speaking is.

In the first unit ‘but women he wrote songs for’ there is only one clear major pitch accents on ‘songs’ while in the second unit ‘he wrote’ there is accent on ‘wrote’. The accenting in the first unit is interesting. One would expect an accent on ‘women’, which would yield a very regular metrical beat. However, I do not perceive salient prominence on this word and rather, it seems that the word ‘wrote’ is somewhat prosodically highlighted. This, on the other hand, makes little sense rhythmically since it would put two prominent syllables (‘wrote’, ‘songs’) next to each other, which creates a so called clash and English speakers might employ several strategies to avoid such rhythmical clashes.⁷ I would speculate that the reason for de-accenting ‘women’ and keeping minor prominence (though not labeled as a pitch accent in my ToBI labeling) on ‘wrote’ instead lies in the discourse function of intonation. ‘Women’ is old/given information, this topic has been established in this discourse segment before. On the contrary, ‘wrote’ is new, since writing songs for women has not been mentioned, and stands with clear pragmatic contrast against ‘performing’ with women and ‘producing’ their songs, both of which were established in the discourse in the previous prosodic unit.

The aspects of connected speech in these two units include the extremely weakened realization of the initial ‘but’ with a schwa and minute glottalization replacing the final [t] and a weak form of ‘he’, but the strong form of the preposition ‘for’ that has a full [ɔ:] vowel due to its phrase-final position and the associated pre-boundary lengthening. We can also observe some de-voicing of the final [z] in ‘songs’ stemming from the voicelessness of the following [f] in the process of regressive voicing assimilation.

Segmentally, I would only point out the voiced realization of both [h] sounds in the two pronouns ‘he’. This is further evidence exemplifying the discussion of this sound we have had so far that prosodically strong syllables are phonetically realized with voiceless initial [h] and prosodically weak syllables tend to have some amount of voicing in the initial [h] sounds.

⁷ One example of stress clash avoidance comes from words with both primary and secondary stresses such as ‘Japanese’ or ‘overnight’. Normally, the pitch accent falls on the syllable with the primary stress as in ‘She is JapaNESE’ or ‘I stayed overNIGHT’. But to avoid stress clash, the major prosodic prominence associate with a word might move as in ‘She’s a JApAnese STUdent’ or ‘I took an OVERnight BAG’ to avoid adjacent stressed syllable of *‘She’s JapaNESE STUdent’ or *‘I took an overNIGHT BAG’. The asterisk indicate that native speakers would most likely not pronounce it this way and if heard it this way, they would find it funny or weird.

(11h) I feel for you by Shaka Khan he wrote nothing compares to you which Sinead O'Connor covered uh

ai fi:l fəɪ ju baɪ 'ʃakə kɑ:n // hɪ ruʊr 'nʌθɪŋ kəm'peəɪz tə ju: / wɪtʃ ʃɪ'neɪd ʊs'kənəɪ 'kʌvəd / ə

We will take the final breath group of the turn in (11h) together with the final filler 'uh' despite the in-breath preceding it for easier discussion. In terms of further division into prosodic units, there is a clear boundary after 'Khan' signaled by pre-boundary lengthening and the low boundary tone. The only other boundary is the in-breath between 'covered' and the final 'uh'.

In the first unit 'I feel for you by Shaka Khan', there are two clear high pitch accents on 'feel' and 'Khan'. I am not sure if 'Shaka' also receives a low pitch accent or just represents a low target for the rising pitch accent on 'Khan'. The completion is again signaled by a low boundary tone realized with glottalization.

In connected speech aspects I noticed the weak form of the preposition 'for', which you might compare with the same preposition 'for' of the previous unit and you clearly can hear the difference in the vowel quality in that the current realization of 'for' is schwa-like whereas the previous one was not and had [ɔ] quality. The following 'you' is not a full strong form with a long [u:] but I would definitely not classify it as weak [jə]. Hence, I would opt for a strong form characterization here.

Segmentally, consider the pronunciation of the name Shaka Khan. The second syllable of the first name 'Shaka' is completely devoiced and phonetically is realized basically only as a weak release of the preceding /k/ with no discernible vowel. The initial /k/ of the second name 'Khan' is, on the other hand, released forcefully with great energy burst and aspiration resulting from voice onset time of around 150 milliseconds.

The last intonational phrase of this breath group and the entire turn is 'he wrote nothing compares to you which Sinead O'Connor covered' followed by a filled pause 'uh'. Let me start the discussion here with the end of the turn and describe the turn management situation. ST, after saying 'which Sinead O'Connor covered' signals that he wants to continue talking since the intonation is plateau-like indicating incompleteness of the thought. Then he takes the breath and indicates further with a filled pause 'uh' that he wants to keep the floor and does not want to yield the floor to LH. However, the filled pause comes too late for LH since she started talking roughly 150 milliseconds after the onset of the filled pause and this is the time that it roughly takes people to react to a voice stimulus from the interlocutor. Hence, she has planned well before what she wanted to say and started executing her plan and the floor-keeping effort with the filled pause came too late. We may

only speculate if LH did not register (or registered too late) the plateau intonation signaling incompleteness of the ST's turn or did notice that and went ahead and interrupted anyway.

Returning now to the last unit of ST's turn, we can divide it into two chunks with a minor boundary between 'you' and 'which'. Although there is no silent pause, and 'you' is even a bit truncated, we clearly perceive a falling pitch following a high pitch accent on 'you'. Moreover, the prominence on 'you' is emphasized by ST's choice of a high, rather than a downstepped accent. Other pitch accented words in this unit are 'wrote', 'nothing', and 'compares'. In the following unit, O'Connor and 'covered' are clearly accented, the prosodic prominence on the first name 'Sinead' is less clear. The plateau boundary contour on 'covered' has been already mentioned.

The connected speech aspects include slightly weakened 'he', fully reduced and weak 'to' and the strong form of 'you'. Regarding 'to', notice how the release is again aspirated and not even a schwa, let alone a [ʊ], can be observed here. The final [t] of 'wrote' is flapped and voiced showing the weakness of the boundary between 'wrote' and 'nothing'.

The segmental realization that is worth noticing is with 'compares'. In this word, the lexical stress falls on the second syllable. As a result, the first syllable is weak and realized as schwa, and aspiration after the release of the first [k] is much weaker than aspiration after [p]: voice onset time of [k] is about 30 millisecond while that of [p] is almost three times as much as at around 85 milliseconds. This is again a very instructive word to practice for Slovak speakers of English since we have a tendency to stress the first vowel, realize it as a full [o] due to its spelling, and aspirate it more than the vowel in the second syllable.

(12) LH: he wrote uh when you were mine which Cindy Lauper covered I know
hɪ ruʊ? ə wen jə wɛɪ maɪn / wɪtʃ / 'sɪndi 'lɔ:pəɪ 'kʌvəd [aɪ nəʊ] //

The turn of LH in (12) is composed of two main chunks, or prosodic phrases, with the boundary between 'mine' and 'which' separating the two chunks but note the break following 'which' that is present as well. The first chunk ends in a continuation rise on 'mine' (L-H%) signaling that the speaker is not finished and wishes to continue. It is interesting to see that the filler 'uh' is used presumably to provide extra time for lexical access of the title of the song, but also that the prosodic pattern of downward declination is not interrupted by this filler and is thus integrated within the prosodic structure of the chunk. We can clearly identify pitch accents on 'when' and 'mine' but the initial accent is difficult to ascertain since 'wrote' is slightly truncated and placing the pitch accent of the filler is also somewhat questionable in this particular situation.

Following 'which', LH makes a clear break that is not realized as a silent pause but rather prolongation of the frication in the affricate [tʃ]. This break makes it difficult to identify the location

of the boundary between the two chunks of 1) ‘He wrote uh when you were mine’ and 2) ‘Cindy Lauper covered I know’; is it before or after ‘which’? My analysis would make a major boundary before which, for that we have both the f0 pattern in the presence of the continuation rise as well as the significant pre-boundary lengthening of ‘mine’. I would then treat the boundary after ‘which’ as a weaker boundary signaled temporarily and with forward-looking boundary rise (H-). Note also that following the accent on ‘which’, the only accented word is the singer’s surname with a low pitch accent, in fact so low that the speaker produces a creak, and the rest of the phrase is de-accented.

Finally, let us discuss the form and function of the final boundary tone of this final unit of the turn. First, consider its form. As I said above, there is a low pitch accent on the syllable ‘[lɔ:]’ and f0 continually rises for the rising boundary tone (H-H%) over the following five syllables [pəɪ 'kʌvəɪd aɪ noʊ]. Although the pitch tracker of Praat does not provide a clear picture for this rise due to the overlapped speech at the end of the turn, mere listening to it we can identify this rising pitch.

Second, consider the function of this unit in the communicative situation. Why would LH rise her pitch when providing a declarative statement that Prince also wrote another song covered by Cindy Lauper? Notice how this rising pattern is clearly different from the falling patterns that ST used in his previous two declarative statements mentioning the songs by Shaka Khan and Sinéad O’Connor. Obviously, I cannot be sure and can only speculate here but there are two potential explanations. The first one is that LH uses a clear example of “high rising terminal” or “uptalk”, i.e. the rising intonation on declaratives that was originally identified in Australian young females and since have spread to many English speaking communities and both genders. Originally, this intonation pattern was analyzed as female’s insecurity and seeking confirmation from her interlocutor, but the uses and implications have since widened considerably.

The other potential explanation is that LH uses a so called list intonation in which the items on the list are spoken with rising intonation indicating that there are some following items until the final item that show finality with a falling pattern. So LH might perceive that they are now both in the mode of listing Prince’s female collaborators and her rising intonation fits this pattern. This is corroborated by ST’s continuation of enumerating female singers Prince collaborated with (Sheena Easton, Sheila E, Apollonia).

While both of these analyses are plausible in my mind, I would vote for the first one that ST is simply used to produce uptalk intonation. Support comes from several following prosodic phrases where she uses uptalk on declaratives. The most clear example of this is the ending intonation in ‘that is literary the record that I am usually going to’ starting at 2:29.9 of the clip.

With the intonational pattern of this turn discussed now, let us turn our attention to the aspect of connected speech in this turn. The weak forms can be observed with ‘he’, ‘you’, and ‘were’. The final [t] of ‘wrote’ is either glottalized, or completely missing due to the truncation of the word.

In the segmental domain, I would like to discuss the production of the first name ‘Cindy’ [sɪndi], and in particular, the two vowels. The status of word-final [ɪ] in English is questionable since this vowel is typically shorter and less tense than [i:] in other word positions but at the same time not as short and lax as the other vowel [ɪ]. Moreover, the lax vowels such as [ɪ] in English distributionally do not appear in the open syllables (i.e. syllables without coda consonants) while the tense vowels like [i:] do. Both of these considerations support the transcription that we used with the first vowel [ɪ] and the second vowel [i] without the length mark.

We can now see if this difference also obtains in the production of this speaker. Praat allows for both the visualization (*Formants* → *Show formants*) and measurement (*Formant listing* at the cursor) of the formant frequencies that are the main acoustic indicators of vowel quality. We see by eyeballing that the second formant (F2) is visibly lower in the first vowel than in the second one. The measurement confirms this: around 2050 Hz and 2470 Hz respectively. The first formant is also different, but in the opposite direction: F1 of the first vowel is higher than in the second vowel (580 Hz vs. 380 Hz in the second vowel). Since the first formant is inversely related to the height of the tongue (the higher F1 the lower the tongue) and the second formant is related to the horizontal position of the tongue (the higher F2 the more front the tongue position), we can conclude that the first syllable is produced with a more lax vowel than the vowel in the second syllable: the second tense vowel with a higher and more front tongue body and thus more advanced tongue root than the first vowel. This is an example of experimental evidence for using the differences in IPA transcription.

(13) ST: He he wrote uh for Shena Easton and Sheila E and Apolonia and and this incredible array of women whose music he celebrated and you get into his own career his own body of just this unbelievable collection of songs.

With this longer turn we employ the same approach as with (11) and discuss individual breath groups. There are three clear in-breaths in the turn: between ‘and’ and ‘this’, between ‘whose’ and ‘uh’, and between ‘into’ and ‘his’.

(13a) He he wrote uh for Shena Easton and Sheila E and Apolonia and and
[hi] hi ru:t / ə fə:ɪ 'ʃi:nə 'i:stən / ən 'ʃi:lə i: / ən ,æpə'ləʊniə / ən ən

The initial repetition ‘he he’ is due to trying to grab the floor from his interlocutor LH and while the first ‘he’ overlaps with LH’s speech, the second one is timed to follow immediately after LH ends her turn. We will not treat this first ‘he’ as a separate phrase due to various complexities such as overlap and laugh. This breath group will be then divided into four intermediate (minor) phrases with the weaker boundaries following ‘wrote’, ‘Easton’, ‘E’, ‘Apollonia’. All of these intermediate phrases end in non-low f0 plateau (labeled H-) indicating non-finality and presenting a variation for the list intonation. Interestingly, the f0 pattern signaling the boundary after ‘wrote’, which can be assumed to stem from problems in lexical access of the first singer’s name due to a short filler following, is prosodically almost identical with the remaining three boundaries signaling the list intonation. Each of these intermediate units contain a single pitch accent on the word directly preceding the boundary (‘wrote’, ‘Easton’, ‘E’, ‘Apollonia’).

The aspects of connected speech here include the clear weak forms on ‘for’ and the first two mentions of ‘and’; while the third and fourth mentions of ‘and’ have a fuller non-schwa vowel. In all four renditions, however, the final [d] is omitted. While in ‘and Sheila’ the omission might be attributed to the CCC cluster with [d] flanked by two other consonants, i.e. a prototypical environment for the elision of /t,d/ in English, the following two cases have ‘and’ followed by a vowel-initial word. Hence, the d-elision can be attributed to the weakness of these conjunctions in this context rather than the phonetic environment in which they appear.

The segmental aspects that we will discuss in this breath group is the production of the last name ‘Apollonia’. It is a nice example of a word with both primary and secondary lexical stress on the third and the first syllables respectively. The secondary stress is primarily marked by the full realization of the vowel, hence not reducing it to a central schwa, in addition to slight increases in amplitude and duration. These phonetic markers of prominence are highlighted even more in syllables with primary stress. The stress pattern also determines other aspects of pronunciation such as aspiration. We see that the unstressed second syllable, starting with [p] is not only reduced to schwa in terms of vowel quality and also shorter and less loud than the flanking stressed syllables, but in addition, the aspiration on [p] does not take place. Hence, yet again, we see the multiple phonetic cues all signaling prosodic weakness of this syllable as a hallmark of redundancy of phonetic cues signaling functional contrasts; in this case of lexical stress.

- (13b) this incredible array of women whose
 ðɪs ɪn'kredɪbəl ə'reɪ əf 'wɪmən hu:z //

This is a very interesting breath group since it shows the very intricate and complex relationship between prosody and syntax/semantics proper of language. Intonationally, this phrase feels complete with a downstepping succession of pitch accents ending seemingly with a final fall. Yet,

this chunk is syntactically a fragment, lacking a verb, and ends with ‘whose’ clearly indicating that the continuation of the speaker’s thought is coming. Furthermore, the syntax would dictate that the prosodic break should come before ‘whose’ and not after it. And, the break after ‘whose’ is not just a minor break but a full pause during which the speaker takes breath. This is a nice example of the complexities of spontaneous speech and the intricate, yet still very close, relationship between prosody and other linguistic extralinguistic aspects of the conversation. This is especially interesting since there is a short filled pause ‘uh’ following the pause after ‘whose’ and thus the presence of the break might also in part relate to the difficulty in lexical access of the suitable words. We should also mention a minor hesitation break following ‘of’, again realized with [f] rather than expected [v], but we will not treat it as a minor boundary and the entire breath group thus constitutes a single intonational phrase with three different types of high accents: incredible (L+H*), array (!H*), and women (H*).

In the connected speech domain, we see a rather strong realization of the preposition ‘of’, which is plausibly related to the minor prosodic break following it. It is also worth mentioning the linking between words ‘array’ and ‘of’. While the linking [j] is naturally present stemming from the high front final sound of ‘array’, there is also observable glottalization commonly associated with prosodic breaks.

Segmentally, I only mention ‘incredible’ here. The complete closure for [k] is so brief or maybe non-existent at all. This is not surprising with [k] whose fricativization is quite frequent. But despite the lack of the closure, the stressed transition [kre] is strongly aspirated such that [r] is completely devoiced.

(13c) uh music he celebrated and then you get into’
 ə 'mjuzɪk hɪ 'seləˌbreɪrəd / əŋ ðen ju: get² 'ɪntu: //

The next breath group in (13c) is divided into two prosodic units with a prosodic break before ‘and’ labeled as L-. The evidence for this break, again, comes from the glottalization preceding the initial vowel of ‘and’, and a pitch reset for the accent on ‘than’. The in-breathing pause following ‘into’ fall under the same characterization of syntax-prosody relationship as described in the previous chunk. It is also worth pointing out that the first part of the chunk is sped up by the speaker whereas the second part slows down, setting up a trend of slowed articulation for the remainder of this turn. As a result, in addition to pitch accenting and highlighting ‘music’, ‘celebrated’, and ‘then’, also ‘get’, and ‘into’ are perceived as prosodically prominent as well.

There are several aspects of connected speech present in this chunk. First, let’s discuss the weak and strong forms. The pronoun ‘he’ is significantly weakened such that the initial [h] is almost

omitted and the vowel is lax and shortened. On the other hand, the pronoun ‘you’ is produced in a strong form since the vowel is quite tense and relatively long.

Second, notice the linking of adjacent words. In ‘and then’ we see the elision of [d], which is quite common with a weak form ‘and’ but we also see the assimilation of the nasal place of articulation since the alveolar nasal is produced dentally anticipating the following dental consonant [ð]. Finally, the link between ‘get’ and ‘into’ seems to be produced with a glottalized [t] but in such a way not to produce a prosodic break between the two words. Of several choices the speaker has at his disposal in linking these two words, such as flapping, full voiceless alveolar release, or a clear glottal stop preceding the vowel, he selects the one most suited to convey the sense of disclosure between the two words since all the above mentioned ways would convey a different perception of the prosodic boundary.

Segmentally, I only mentioned two additional realizations of spelled /t/ in ‘celebrated’ and ‘into’. While the first one is a nice example of a flap with a closure shorted that the one for the following /d/, the second one is quite strongly aspirated, which is a bit unusual for the unstressed syllable. However, the aspiration seems to be an accompanying signal of strengthening of this syllable due to its pre-pausal position, and thus the environment in which syllables are lengthened in speech.

(13d) his own career his own body of just this unbelievable collection of songs
hɪz oʊn kə'riəɪ / hɪz oʊn 'bɔ:ri əf // dʒʌst' ðɪs / ,ʌnbə'li:vəbəl kə'leɪksən əf sɔ:ŋz //

We will discuss the remainder of this turn as a single prosodic breath group that is, however, divided into several units with breaks discussed below. First, as mentioned in discussing the previous chunk, the speaker slows down and several words are produced with a strong emphasis. As a result, some links between words might be perceived as minor boundaries, for example between ‘own’ and ‘body’ or between ‘this’ and ‘unbelievable’; both of which are produced with a clear silent pause. However, in these places we do not see a pitch reset or a clear boundary tone and I would just describe these as stemming from the slowed down speech rate. Of course, the alternative with low phrasal tones (L-) for minor boundaries is also acceptable. On the other hand, the boundary between ‘of’ and ‘just’ would qualify as a major prosodic boundary since it has all four elements of a major break (pause, pre-boundary lengthening, pitch reset, and final fall).

Note how we again see a major prosodic boundary at a place where neither syntactically nor semantically we would normally expect one. It seems that the pragmatic reason for the presence of this boundary here is to provide an additional cue for the extra strong prosodic emphasis on ‘just’ and ‘unbelievable’ since separating them from the previous material complements the perception of prosodic prominence already cued by pitch accents and durational adjustments.

Speaking of pitch accents, almost every word in this chunk receives one but while the accents in the first prosodic unit ‘his own career his own body of’ follow a general downstepping pattern, with a slight reset after ‘career’, the accents on ‘just’ and ‘unbelievable’ have a clear rising pattern, which is used for extra emphasis and represent points on the scale in Pierrehumber and Hirschberg’s proposal. The speaker clearly expresses a very high point on such scale.

The most interesting is the situation with ‘unbelievable’ that I would analyze as having two clear pitch accents on both the primary and secondary stressed syllables. In some theoretical frameworks this would be difficult to describe but as Stephanie Shattuck Hufnagel (p.c.) has maintained for long time, double accented words are possible in English.

The connected speech issues in this chunk are also interesting. Take for example the two renditions of the pronoun ‘his’. The first one is accented while the second one is not. This prosodic difference translates into more than twice as long duration of the first one compared to the second one. Note that this is not achieved by elision of any segments (the elision of the initial [h] is permissible and quite common) but by general shortening of all three segments. In ‘body of’ we should mention slight glottalization, which again could have been replaced with a linking [j] between [i] and [ɑ] if the boundary was weaker but the glottalization signals a perceivable disjuncture. Also the final devoicing of [v] of ‘of’ is likely linked to the pre-pausal position of this vowel. Finally, the [t] of ‘just’ is not elided due to the extreme prominence of this word but is also not released and smoothly transfers into the following fricative.

Among the segmental aspects we can again mention rather strong aspiration of the first syllable of ‘career’ despite being unstressed, which is possibly linked to the emphasizing of this word. Next, the flapped realization of [d] in ‘body’ nicely shows that flapping, that affects both /t/ and /d/ in American English, is not really produced as [d] as many Slovak speakers of English would think but this flap is much weaker than a regular [d] in that the contact of the tongue with the alveolar ridge in the flap [ɾ] is either extremely brief, as in this case, or sometimes the tongue might not even touch the alveolar ridge.

5. Conclusion

Similarly to the conclusion of Pat 1, I would like to congratulate all those who have fully engaged in the theoretical notes in Sections 2 and 3, and the practical analysis of a dialogue in Section 4. It was only 80 seconds of speaking and 13 speaking turns that were covered in this analysis. But the space that was devoted to such a seemingly short period of speaking illustrates how exploring the speaking patterns of spontaneous speech may be both challenging in showing the complexities and rewarding in discovering the system and regularities that were previously not part of your conscious awareness.

We have analyzed turns of various lengths, pragmatic intentions, and prosodic realizations, and this breath and variability offers great material for discussing and practicing the meanings expressed through the variability in the segmental and suprasegmental aspects speaking. The sound file includes another roughly 80 seconds of the dialogue that is included primarily for you as the material you should use to try your own analyses and also the material for imitation and practice to improve the flow of your non-native English production and lowering the perception of non-native accent.

I believe that sharpening your analytical skills by walking through the guided analysis of this book and attempting your own analysis of the remainder of the dialogue provides a valuable experience, skill acquiring and self-awareness process. Further effort in imitating the flow of the native speakers in the turns will improve not only your productive ability in decreasing the non-native accent in your English but also your perception ability in training your ear in recognizing prosodic prominence and other aspects covered in this book that will improve your proficiency to readily recognize spontaneous conversations of native English speakers.

References

- Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. (2006). The original ToBI system and the evolution of the ToBI framework. In: Sun-Ah Jun (ed.) *Prosodic models and transcription: Towards prosodic typology*. Oxford University Press.
- Beňuš, Š., Bírová, J., Chebenová V., Molnárová, A. (2010). *Výučba výslovnosti cudzích jazykov pomocou porovnávacej akustickej analýzy*. Nitra: UKF
- Brennan S. E., Clark H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 1482–1493.
- Giles, H., Coupland, J., Coupland, N. (1991). *Accommodation Theory: Communication, Context, and Consequence*. In Giles, Howard; Coupland, Justine; Coupland, N. *Contexts of Accommodation*. New York, NY: Cambridge University Press.
- Kráľová, Z. (2011). *Slovensko-anglická zvuková interferencia*. Žilina: Fakulta humanitných vied Žilinskej univerzity v Žiline.
- Levinson, S. C. (2016). Turn-taking in human communication, origins, and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6-14.
doi:10.1016/j.tics.2015.10.010.
- Mennen, I., Schaeffler, F., Docherty, G. 2007. Pitching it differently: A comparison of the pitch ranges of German and English speakers. *Proc. 16th ICPhS Saarbrücken*, 1769-1772.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation (Massachusetts Institute of Technology).
- Pierrehumbert, J., Hirschberg, J. (1990). The Meaning of Intonation Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack (eds.) *Intentions in Communication*. Cambridge, MA: MIT Press, pp. 271 – 311.
- Roach, P. (2000). *English Phonetic and Phonology: A Practical Course* (3rd ed.). Cambridge University Press, Cambridge UK.
- Sacks, H., Schegloff, E.A., Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language* 50(4), 696-735.

Turton, D. (2016). Some /l/s are darker than others: Accounting for variation in English /l/ with ultrasound tongue imaging. *University of Pennsylvania Working Papers in Linguistics* 21.

Štefan Beňuš
Practice material for English Phonetics 2:
Comprehensive analysis of spontaneous speech
Nitra: Univerzita Konštantína Filozofa v Nitre
2018