7. Syllables and Sequences

Key Concepts

- Syllables are intuitive units of speech planning.

- Languages place constraints on syllable structure.

- The dynamics of articulation mean that speech must be planned to ensure appropriate sounds are produced at normal speaking rates.

- Planning involves anticipatory and carry-over contextual effects: coarticulation, assimilation and elision.

- The articulatory and acoustic form of phonological material depends on context.

- There are no necessary acoustic cues to segment, syllable or word boundaries in the signal.

Learning Objectives

At the end of this topic the student should be able to:

- describe the general structure of English syllables

- list some of the limitations on the form and content of syllables in English

- explain some differences between stressed and unstressed syllables

- give an explanation and examples of coarticulation, assimilation and elision

- interpret spectrograms of simple syllables

- explain why a "piecewise stationary" account of speech is insufficient.

Topics

1. English Syllable Structure

We have seen that there are about 19 phonological choices for vowels in English, and about 24 phonological choices for consonants. English syllables are made up from one vowel surrounded by zero or more consonants, but there are limitations on the complexity and combinations of the consonant sequences.

Conventionally syllables are divided into elements: onset, rhyme, nucleus and coda, as shown in the diagram below. The justification for this is that many restrictions occur as to what phonological elements can occur within these elements, but few restrictions occur across elements.
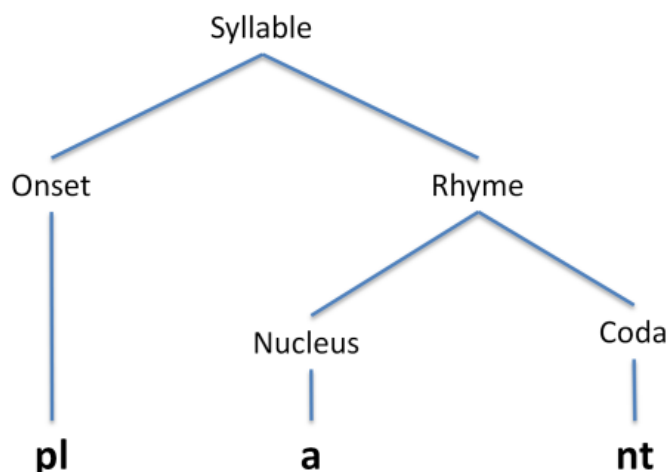
- **Ons**ets. In English the onset can consist of zero, one, two or three consonants. For example "a" /eɪ/, "ray" / reɪ/, "pray" /preɪ/, "spray" /spreɪ/. In onsets with three consonants, the first consonant must be /s/. Also approximant sounds can only occur in the last position next to the vowel, so we get /njuː/ but never /jnuː/, or /fliː/ but never /lfiː/. If the syllable begins with /s/+ a plosive, then there is never a distinction between voiced and voiceless plosives, that is if there is a word /sp-/ then there is not a word /sb-/. The sound /ŋ/ does not occur in onsets.

- **Nucleu**s. In English the nucleus is usually a single vowel, whether that is a long or a short monophthong or a diphthong. Under certain limited situations, other sonorants can take the place of the vowel, for example "puddle" as /pʌdl̩/. We write a syllable mark underneath the consonant in transcription to indicate that it is acting as a syllable nucleus.

- **Cod**a. In English the coda can consist of zero, one, two, three or four consonants. For example: "lee" /liː/, "lean" /liːn/, "link" /lɪŋk/, "length" /leŋkθ/, "lengths" /leŋkθs/. In four consonant codas, the last consonant can only be /s/ (in fact it can only be the plural morpheme -s or the possessive marker 's). The only approximant sound that can occur in a coda is /l/ (except for *rhotic* accents, which also allow /r/). As in onsets, approximant sounds in the coda are always found adjacent to the vowel. The sound /h/ does not occur in codas.

There are about 50 different onsets, 25 different nucleii and 80 different codas in English, which gives about 100,000 possible syllables. However only about 15,000 of these actually occur in English words, and only about 10,000 of these occur as mono-syllabic words. In contrast, Japanese is said to have only 100-200 syllables.

2. Combining Syllables

English words may be made up from one or more syllables. Generally listeners are adept at counting the number of syllables in a word. Since each syllable contains exactly one vowel sound, it is essentially a case of counting vowels (sounds not letters).

- here (1), hero (2), heroine (3)

- stick (1), sticky (2), stickiness (3), stickinesses (4)

- establish (3), disestablish (4), disestablishment (5), disestablishmentarian (8), disestablishmentarianism (10), antidisestablishmentarianism (12).

In polysyllabic words, we often hear that one syllable is more prominent than the others. For example "often",

not "often". We call the prominent syllables *stressed* syllables, and these are often longer, more well articulated and with clearer vowel quality. The weaker syllables are called *unstressed* syllables and are often shorter, less well articulated and with a more central vowel quality. Compare "<u>pho</u>tograph", "pho<u>to</u>graphy", "photo<u>graph</u>ic".

The majority of two syllable words are stressed on the first syllable:

- <u>comm</u>on, <u>na</u>tion, <u>o</u>pen, <u>stu</u>dy, <u>so</u>rry

The majority of words with three or more syllables are stressed on the antepenultimate syllable:

- <u>u</u>niverse, <u>ar</u>ticle, <u>re</u>lative, de<u>mo</u>cracy, e<u>co</u>nomy

but there are many words which do not follow these patterns!

3. Articulation of syllables and sequences

Before we look at the articulation of sequences of sounds, let's consider some simple aspects of the dynamics of moving objects.

An example of a simple dynamical system is a mass connected to a spring responding to an applied force. When a force is applied, the mass accelerates in the direction of the force and travels until the applied force is equal to the restoring force of the spring. Consider a step change in an applied force:
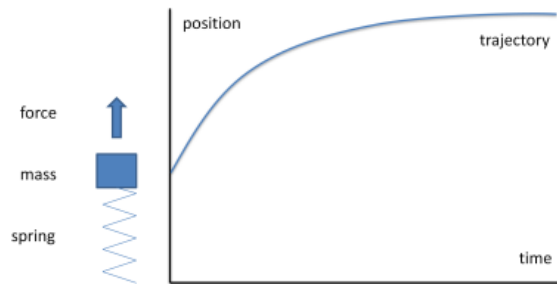

The object moves from its current position to a new position under the action of the force. The path that it takes depends on the degree of damping in the system (what frictional forces apply). If damping is too low then the mass overshoots the new rest position; if it is too high, then the mass takes a long time to get to the new position. If the damping is just sufficient to get to the new position without overshoot, we say the system is *critically damped*.
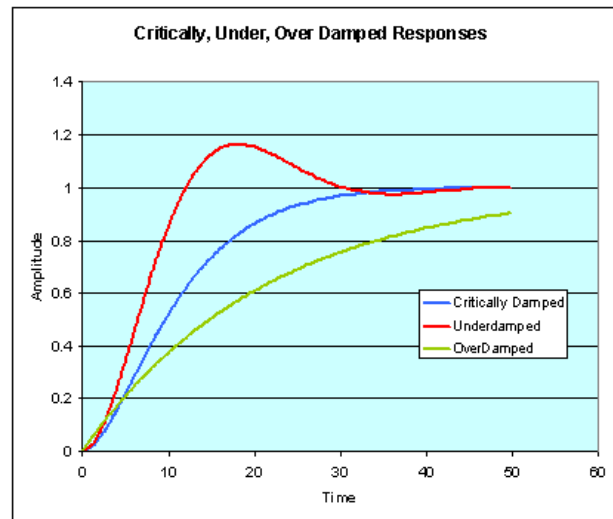
Critically damped movement is the fastest way to get to the target without overshoot. A reasonable assumption (and one accepted by various articulatory theories) is that articulator movement is critically damped.

However, if movement of the object is critically damped and if there is not enough time for it to travel to its target location under the applied force, then *undershoot* can arise, as seen in this diagram:
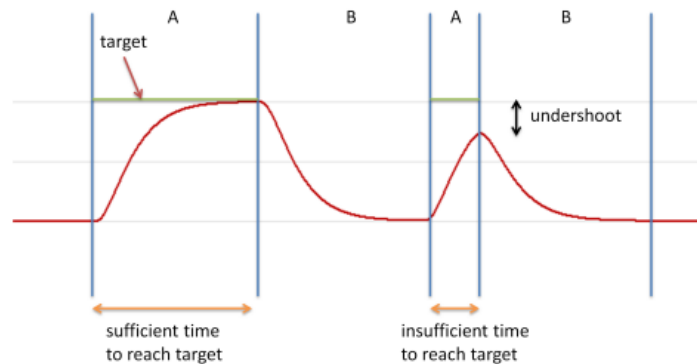
Here we see a sequence of targets A, B, A, B. For the first A target, the object has sufficient time to reach the target value before it heads off for the following B target. For the second A target, however, insufficient time is available, and the object heads on to the second B target before it reaches the A target position. This behaviour is called *undershoot*.

From this discussion we need to draw the conclusion that speaking sequences of speech sounds - involving multiple articulators each with different dynamical properties - is going to take a lot of *planning*. We cannot just throw the articulators around and hope that they will be in the right position at the right time. We need to allow sufficient time for each articulator to reach its target location and we need to plan asynchronous starts to ensure that the slower moving articulators are in position for when they are needed to synchronise with the faster moving articulators.

The articulatory planning to ensure rapid articulation of still recognisable sounds causes *contextual effects*, these are changes to the articulation and acoustics of sounds depending on the context - where they are in the word, syllable or syllable element and what their neighbouring sounds are.

4. Types of contextual effects

In the literature you will find a range of names for different types of contextual effects. Let us take them in turn, with examples.

- **Coarticulation**: Coarticulation means 'joint articulation', that is a situation where one articulator is trying to satisfy more than one segment. A classic example is in the articulation of "car" and "key"; in the first, anticipation of the vowel moves the body of the tongue low and back and so the velar closure occurs further back in the mouth $[\underset{\sim}{k}\alpha]$; while in the second, anticipation of the vowel moves the body of the tongue high and front and so the velar closure occurs further forward in the mouth $[\underset{+}{k}i]$. The contextual variation of the two [k] sounds is caused by the joint articulation of the plosive and the vowel. Another example of coarticulation is audible in the quality of /h/ before different vowels. When articulating [h], the tongue anticipates the position for the following vowel and so the vocal tract tube that filters the turbulence generated by the glottal fricative has the frequency response of the following vowel. Indeed we could say that varieties of /h/ are as different to one another as varieties of vowels.

- **Assimilation**: Sometimes the efficient planning of articulator movement leads to articulation that is ambiguous in terms of what phonological segment sequence it represents. Typically the place, manner or voicing of a consonant is altered to make it easier to produce in combination with adjacent consonants.

  In "good boy" or "good girl", for example, the alveolar plosive /d/ at the end of "good" is tricky to articulate in front of the bilabial plosive /b/ in "boy", or the velar plosive /g/ in "girl". If the place of the plosive is merged with the following plosive we will get respectively: $[g\upsilon bb\mathfrak{I}]$, $[g\upsilon gg\mathfrak{I}\mathfrak{I}l]$. In "good boy" we hear a long /b/ sound instead of the /db/ sequence, while in "good girl" we hear a long /g/ sound instead of the sequence /dg/. It is *as if* the identity of the first plosive has been changed to that of the second, although really what is happening is that the place of the plosive has become ambiguous.

  This loss in phonological distinctiveness is called *assimilation*.

  Another common form of assimilation is *devoicing*, particularly of voiced plosives and voiced fricatives before voiceless sounds. Since all English voiced plosives and fricatives have voiceless counterparts, a loss of voicing can always be interpreted as a change in the identity of the segment itself. It is common to hear "have to" spoken as $[hæft\upsilon]$ for example.

- **Elision**: In the assimilation examples above, we see how the place, manner or voice elements of a consonant can be modified because of the demands of efficient planning. In "good boy", we say that the distinctive place element of the plosive has been lost, so that it takes on the plosive of the following /b/.

  There are other situations in which whole segments can be lost, perhaps simply because there is not enough time to articulate them, or because meaning is preserved without them. This dropping of whole phonological segments is called *elision*.

  Consider the phrase "next week", it is indeed possible to articulate this as $[nekstwi\mathfrak{I}k]$ but it is also very often heard as just $[nekswi\mathfrak{I}k]$. The [t] is tricky to articulate between consonants, and it is unlikely that the listener will interpret the result as "necks week".

  Simplification can go even further such that whole syllables can be dropped, for example in "library" as /laɪbri/ rather than the dictionary form /laɪbrəri/.

5. Spectrograms of Connected Speech

One of the biggest surprises when we view spectrograms of connected utterances is that there are often no obvious sound cues to the boundaries between words, syllables or phonetic segments. Just as the utterance is

spoken in smooth unbroken movements of the articulators, the sound pattern is also smooth and unbroken. When we look closely we can see how movements of the articulators from one segment to the next cause dynamic changes in sound quality in the spectrogram both within words and across words.

In the example below, we see the word "swim" /swɪm/. Notice that there is no sharp boundary between any pair of segments. Because the articulators move slowly from one segment to the next, we see "joins" or *transitions* between segments. If we look closely we see how the acoustic quality of /s/ is affected by the lip rounding anticipated for the following /w/, and the vowel /ɪ/ is affected by anticipation of the following nasal. Conversely, the /w/ is slightly devoiced by the preceding /s/, and the vowel /ɪ/ is slightly rounded by carry-over from the preceding /w/.
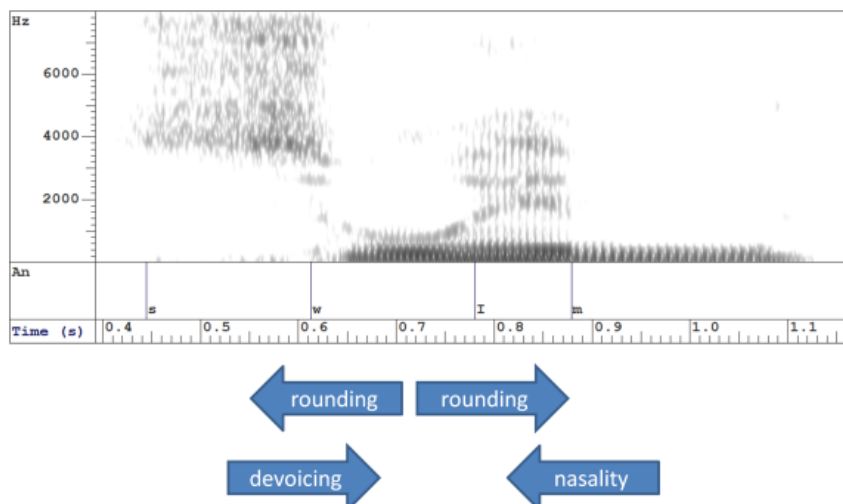
Although we treat phonological segments as discrete and symbolic, we cannot say the same of the sounds in the spectrogram that realise them. The spectrogram shows that speech sounds are affected by the context in which they arise as a consequence of the fact that the articulators move slowly and efficiently. The implication is that we cannot consider speech as a sequence of stationary, context-independent sounds.

Readings

Essential

- Coarticulation and connected speech processes, a web tutorial by Peter Roach.

Background

- Elizabeth Zsiga, "The Sounds of Language", Wiley Blackwell 2013. Chapter 11 - Phonotactics and Alternations. [available in library]

Laboratory Activities

In this week's lab session we will experiment with creating new utterances by gluing together the sound elements from old utterances.

Activities:

1. concatenation of individual segments

2. concatenation of syllable components

3. concatenation of words

4. concatenation using a "slot and filler" grammar

5. experimenting with text-to-speech conversion

Reflections

You can improve your learning by reflecting on your understanding. Come to the tutorial prepared to discuss the items below.

1. English spelling is rather unpredictable. Think of two words having the same spelling but different pronunciation. Think of two words with different spelling but the same pronunciation.

2. How do you know which syllable(s) should be stressed in a poly-syllabic word? Why might this be a problem for foreign learners of English?

3. In the word "sue", the /s/ has a different quality compared to the /s/ in "see". Why is this? Is this an example of coarticulation, assimilation or elision?

4. In my pronunciation of "keel" the /l/ has a different quality compared to the /l/ in "leek". Could this be an example of coarticulation?

5. The word "interesting" is most commonly spoken with some elements elided. What is a possible simplified form of this word?

Word count: 2315. Last modified: 14:20 21-Feb-2018.