Description of SpeechMark Syllabic Cluster Identification Program

Susan Worst and Harriet Fell, 1999; modified by Suzanne Boyce, 2014

This document describes the process by which the SpeechMark syllabic cluster analysis operates to group previously computed landmarks. The grouping algorithms were developed to deal with English-focused infant speech including babble—that is, speech whose intended lexical content is unknown (if it exists). Sequences that would be transcribed as an infant attempt at a speech syllabic cluster were identified, and empirical rules for separating these from the speech stream and from each other were developed based on landmark sequences and timing. It is important to remember that the syllabic cluster rules so developed are sensitive only to the speech AS UTTERED. They may or may not match syllabic clusters of speech as analyzed by transcription.

The specific syllabic clusters identified are roughly equivalent to the possible syllabic clusters of English. Only the following landmarks are considered for the purpose of identifying syllabic clusters: +b, -b, +g, -g, +s, -s. The +b and –b landmarks reflect the presence of obstruent-like acoustic characteristics in the speech signal. The +s and –s reflect the presence of sonorant-like acoustic characteristics and/or the transition from sonorant to sonorant. The +g and –g landmarks track the presence of periodicity at the point where it acquires, or loses, consistent harmonic power—that is, voicing that has an effective function for speech.

All of these landmarks are “articulator-independent”. That is, they do not respond to place information except as it affects the degree to which their acoustic criteria are present in the acoustic signal. For instance, voicing may begin differently for labial vs. velar stops, and this may affect the detection of +g landmarks, but the landmark itself carries no such information. Likewise, they respond to manner information only as it affects whether obstruent-like information is present in the acoustic signal. For instance, the +b and –b landmarks respond similarly to the burst portion of stop consonants and the onset of frication for fricatives, because they have similar acoustic characteristics.

Preprocessing

Before grouping SPEECHMARK landmarks into syllabic clusters, the program excludes some landmarks from consideration by marking their status as "axed" or "deleted." "Axed" landmarks occur in regions of the recording that include non-linguistic utterances, or in which the speaker’s voice has been obscured. The boundaries of axed regions are designated in the landmark files by the symbol "ax-h". A single "ax-h" is sometimes used to mark axed regions at the beginning or end of the file; otherwise, the symbols bracket the axed region.

The phrase “landmarked file” means a file with already-computed landmarks.

The program marks SPEECHMARK symbols as "axed" if

- A single "ax-h" symbol appears at the very beginning of the landmarked file, and the SPEECHMARK landmark timepoint falls before the ax-h timepoint;

- A single "ax-h" symbol at the end of the file, and the SPEECHMARK landmark timepoint falls after the ax-h timepoint;

- The "ax-h" symbols appear in pairs, and the SPEECHMARK landmark timepoint falls between the two ax-h timepoints;

- The symbol of the SPEECHMARK landmark is +b, +g, or +s, and the landmark appears immediately before an axed region. If there is an uninterrupted series of such landmarks, all of them will be excluded.

- The symbol of the SPEECHMARK landmark is -b, -g, or -s, and the landmark appears immediately after an axed region. If there is an uninterrupted series of such landmarks, all of them will be excluded.

An SPEECHMARK landmark is "deleted" if it is interpreted as an artifact of processing rather than an indication of a real speech event. The program omits landmarks that satisfy any of the following conditions:

- -g/+g pairs that are 30 milliseconds apart or less, in which neither the -g nor the +g is axed;

- +s/-s pairs that are 30 milliseconds apart or less, in which neither the +s nor the -s is axed;

- a +b that is followed by something other than -b or +g, or that appears at the very end of the file;

- a +b that is followed by an axed +g, or a +g that is more than 100 milliseconds away;

- a +b/-b pair that is followed by something other than an unaxed +g that is less than 100 milliseconds away from the +b;

- a -b that is preceded by anything other than an unaxed -g or +b, or that appears at the very beginning of the file;

- a -b that is preceded by a -g that is axed, or more than 100 milliseconds away.

Syllabic cluster Identification

The program identifies sequences of landmarks as syllabic clusters based primarily on their order. Forty possible syllabic clusters are recognized, plus a catchall category of "other". Eleven recognized syllabic clusters begin with +g:

+g/-g
+g/-g/-b
+g/+s
+g/+s/-g
+g/+s/-g/-b
+g/+s/-s
+g/+s/-s/-g
+g/+s/-s/-g/-b
+g/-s
+g/-s/-g
+g/-s/-g/-b

Each of these eleven syllabic clusters may have a prefix of +b (yielding an additional eleven syllabic clusters) or by +b/-b (providing eleven more).

Five syllabic clusters begin with +s:

+s/-g
+s/-g/-b
+s/-s
+s/-s/-g
+s/-s/-g/-b

Finally, two syllabic clusters begin with -s:

-s/-g
-s/-g/-b

In some cases, the distance between landmarks is used to separate sequences of landmarks into syllabic clusters; or, conversely, to “bind” adjacent landmarks so that they are part of the same syllabic cluster. Specifically:

- Deleted and axed landmarks are not included in syllabic clusters. Deleted landmarks are skipped over as if they did not exist; axed landmarks bring the current syllabic cluster to a stop.

- If a +b precedes a +g, and the +b has not been deleted, then the +b begins the syllabic cluster. (Recall that a +b is deleted if it is more than 100 ms away from a following unaxed +g.)

- If a +b/-b pair precedes a +g, and the +b/-b pair has not been deleted, then the +b/-b begins the syllabic cluster. (A +b/-b pair is deleted when the distance between the +b and the +g exceeds 100 milliseconds.)

- If a -b follows a -g, and the -b has not been deleted, then it is part of the syllabic cluster. (A -b is deleted when it occurs more than 100 milliseconds after a -g.)

- +g/-g, +g/-s, and +s/-g syllabic clusters (or syllabic clusters containing these sequences) may be of any length.
  +s/-s syllabic clusters must exceed 30 milliseconds; if they do not, they will have been deleted.

- If a +g/+s sequence appears in a syllabic cluster, and the distance between +g and +s is greater than 200 milliseconds, then the syllabic cluster ends at the +s, and a new syllabic cluster will begin with the same +s. If the distance between the +g and +s is less than or equal to 200 milliseconds, then the syllabic cluster continues on from the +g and +s.

- If a -s/-g sequence appears in a syllabic cluster, and the distance between the -s and the -g is less than or equal to 200 milliseconds, then the -s and the -g will remain part of the syllabic cluster. Otherwise, the current syllabic cluster will end with the -s, and a new syllabic cluster will begin with the -s/-g.

If a landmark sequence does not precisely match any of the syllabic clusters listed above, then the cluster is classified as "other." The "other" syllabic cluster will begin at the landmark following the previous defined syllabic cluster, and it will end at the next -g or -b; at the landmark before the next +b or +g if no -g or -b is found; at the last unaxed landmark; or at the last landmark in the input data. In data analyzed so far, the most common "other" syllabic cluster is +g/+s/+s/-g, in which the +g and +s are less than 200 milliseconds apart; this syllabic cluster is classified as "other" because the existing syllabic cluster list does not provide for the possibility of two +s in a row.

Statistics

Once a syllabic cluster is identified, then the tallies for that syllabic cluster (one tally for the file being examined, another for the entire group of files) are incremented; the length of the syllabic cluster is added to a running total used to compute the mean, and the syllabic cluster length is stored in a linked list used to calculate the standard deviation. The length of a syllabic cluster is the difference in the timepoints of the first and last landmarks in the syllabic cluster, except for syllabic clusters in the following pattern:

* /-s +s/*

where * can be any valid sequence of landmarks. In these cases, the syllabic cluster boundary is taken as the average of the timepoints of the -s and +s.

Utterances

The program computes the number of utterances in each file, and in the group of files as a whole, as well as the average number of syllabic clusters per utterance. An utterance is a sequence of syllabic clusters in which gaps between syllabic clusters are no more than 200 milliseconds long. When computing the number of syllabic clusters per utterance, syllabic clusters that do not conform to one of the 40 recognized types are excluded.