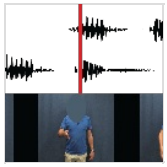


Sep 30, 2020

Audiovisual stimuli creation

HansRutger.Bosker [1]

[1]Max Planck Institute for Psycholinguistics

1 Works for me dx.doi.org/10.17504/protocols.io.bmv3k68n

HansRutger.Bosker

ABSTRACT

This protocol describes three ways of creating audiovisual stimuli that include (1) manipulations of the suprasegmental cues to lexical stress and (2) two different gestural alignments. These methods were used in Bosker & Peeters (submitted; https://doi.org/10.1101/2020.07.13.200543).

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Bosker, H. R., & Peeters, D. (submitted). Beat gestures influence what speech sounds you hear. *bioRxiv*. https://doi.org/10.1101/2020.07.13.200543

DOI

dx.doi.org/10.17504/protocols.io.bmv3k68n

PROTOCOL CITATION

HansRutger.Bosker 2020. Audiovisual stimuli creation. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bmv3k68n

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Bosker, H. R., & Peeters, D. (submitted). Beat gestures influence what speech sounds you hear. *bioRxiv*. https://doi.org/10.1101/2020.07.13.200543

CREATED

Sep 30, 2020

LAST MODIFIED

Sep 30, 2020

PROTOCOL INTEGER ID

42651

ABSTRACT

This protocol describes three ways of creating audiovisual stimuli that include (1) manipulations of the suprasegmental cues to lexical stress and (2) two different gestural alignments. These methods were used in Bosker & Peeters (submitted; https://doi.org/10.1101/2020.07.13.200543).

Experiment 1A (+2)

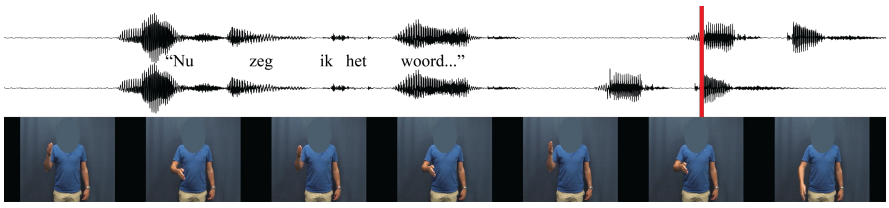
1 For Experiment 1A, we created 12 disyllabic pseudowords that are phonotactically legal in Dutch (e.g., "wasol" /ʋa.sɔl/; see Table below). In Dutch, lexical stress is cued by three suprasegmental cues: amplitude, fundamental frequency (F0), and duration (Rietveld & Van Heuven, 2009). We created 7-step lexical stress continua varying in mean F0 while keeping amplitude and duration cues constant, thus ranging from a 'strong-weak' (SW; step 1) to a 'weak-strong' prosodic

pattern (WS; step 7).

	pseudowords Experiments 1A, 2, and 3
1	/ʊa.sɔl/
2	/ɛl.pɑt/
3	/kla.fɔs/
4	/lo.sɛp/
5	/nu.fa/
6	/plo.sɪm/
7	/pra.bɔp/
8	/rɔs.kɪl/
9	/rʏ ŋ.ka/
10	/stra.dɔt/
11	/tu.sa/
12	/dɛm.rɔf/

2 Continua were created by recording an SW (i.e., stress on initial syllable) and a WS version (i.e., stress on last syllable) of each pseudoword from a male native speaker of Dutch. We first measured the average duration and amplitude values, separately for the first and second syllable, across the stressed and unstressed versions: mean duration syllable 1 = 202 ms; syllable 2 = 395 ms; mean amplitude syllable 1 = 68.1 dB; syllable 2 = 64 dB.

3 Then, using Praat (Boersma & Weenink, 2020), we set the duration and amplitude values of the two syllables of each pseudoword to these ambiguous values. Subsequently, F0 was manipulated along a 7-step continuum, with the two extremes and the step size informed by the talker's originally produced F0. Manipulations were always performed in an inverse manner for the two syllables of each pseudoword: while the mean F0 of the first syllable decreased along the continuum (from 159.2 to 110.6 Hz in steps of 8.1 Hz), the mean F0 of the second syllable increased (from 95.7 to 141.3 Hz in steps of 7.6 Hz). Moreover, rather than setting the F0 within each syllable to a fixed value, more natural output was created by including a fixed F0 declination within the first syllable (linear decrease of 23.5 Hz) and second syllable (38.2 Hz). Pilot data from a categorization task on these manipulated stimuli showed that these manipulations resulted in 7-step F0 continua that appropriately sampled the SW-to-WS perceptual space (i.e., from 0.91 proportion SW responses for step 1 to 0.16 proportion SW responses for step 7).

4 

To create audiovisual stimuli, the first author was video-recorded using a Canon XF205 camera (50 frames per second; resolution: 1280 by 720 pixels) with an external Sennheiser ME64 directional microphone (audio sampling frequency: 48 kHz). Recordings were made from knees up on a neutral background in the Gesture Lab at the Max Planck Institute for Psycholinguistics. The speaker produced the Dutch carrier sentence: "Nu zeg ik het woord... kanon", "Now say I the word... cannon", with lexical stress on the second syllable of "kanon". Concurrently, he produced three beat gestures in this sentence, with their apex aligned to the syllables "Nu", "woord", and "–non". The sentence-final beat gesture's stroke phase lasted 120 ms, the post-stroke hold was 100 ms, and the recoil phase lasted 340 ms. The manipulated auditory pseudowords described above were combined with this audiovisual recording using ffmpeg (version 4.0; available from http://ffmpeg.org/) by (1) removing the original "kanon" target word, (2) inserting a silent interval, and (3) inserting the manipulated pseudowords. By modulating the length of the intervening silent interval, the onset of either the first or the second vowel of the target pseudoword was aligned to the final beat gesture's apex. Furthermore, the facial features of the talker were masked to hide any visual articulatory cues to lexical stress. Finally, in order to mask the cross-spliced nature of the audio, combining recordings with variable room acoustics, the silent interval and pseudoword were mixed with stationary noise from a silent recording of the Gesture Lab. The same audiovisual stimuli were used in Experiment 1A with an explicit lexical stress categorization task as well as in Experiment 2 with a shadowing task (which itself

Citation: HansRutger.Bosker (09/30/2020). Audiovisual stimuli creation. https://dx.doi.org/10.17504/protocols.io.bmv3k68n

generated the auditory stimuli for Experiment 3).

Experiment 1B

5 To test the generalizability of our findings with pseudowords to more naturalistic stimuli, Experiment 1B was identical to Experiment 1A except that we used 5 Dutch minimal word pairs that only differ in lexical stress (e.g., *Plato* /ˈpla:.to/ “Plato” with stress on the first syllable vs. *plateau* /pla:.ˈto/ “plateau” with stress on the second syllable; see Table below).

	real Dutch minimal word pairs Experiment 1B
1	Plato /ˈpla:.to/ “Plato” – plateau /pla:.ˈto/ “plateau”
2	Servisch /ˈsɛr.vis/ “Serbian” – servies /sɛr.ˈvis/ “tableware”
3	voorruit /ˈvɔ:r.œyt/ “wind shield” – vooruit /vɔ:r.ˈœyt/ “forwards”
4	canon /ˈka:.nɔn/ “canon” – kanon /ka:.ˈnɔn/ “cannon”
5	voornaam /ˈvɔ:r.na:m/ “first name” – voornaam /vɔ:r.ˈna:m/ “distinguished”

6 The same male talker as in Experiment 1A was recorded producing each of the members of all 5 minimal pairs. After manual annotation of syllable onsets and offsets, we measured the values of the suprasegmental cues to lexical stress (F0, amplitude, and duration) in all tokens. Only the tokens with stress on the first syllable were selected for manipulation, which followed a similar procedure as used for the pseudowords in Experiment 1A.

7 We manipulated duration and amplitude cues to lexical stress by setting these to ambiguous values, while varying the F0 to create an F0 continuum from a strong-weak pattern (SW) to a weak-strong pattern (WS) using PSOLA in Praat. First, the duration of the initial syllable was set to the mean duration calculated across the stressed and unstressed version of the initial syllable; similarly, the final syllable was also set to the mean duration value calculated across the stressed unstressed version of the final syllable, thus setting the duration cues to ambiguous values (mean duration syllable 1 = 211 ms; syllable 2 = 293 ms). Similarly, the amplitude cues were set to ambiguous values (mean amplitude syllable 1 = 66.01 dB; syllable 2 = 66.14 dB). Subsequently, F0 was manipulated along a 7-step continuum, using the procedure and the same F0 values as applied in Experiment 1A. Pilot data from a categorization task on these manipulated stimuli showed that these manipulations resulted in 7-step F0 continua that appropriately sampled the SW-to-WS perceptual space.

8 Finally, these manipulated words were spliced into the audiovisual stimuli from Experiment 1A, creating audiovisual stimuli with two different alignments of the last beat gesture to either first vowel onset or second vowel onset. Like in Experiment 1A, facial features of the talker were masked, and silent intervals as well as manipulated words were mixed with stationary noise to match the room acoustics across the sentence stimuli.

Experiment 4 (+ S2)

9 For Experiment 4, we created 8 new disyllabic pseudowords, with either a short /ɑ/ or a long /a:/ as first vowel (e.g., *bagpif* /bɑx.pɪf/ vs. *baagpif* /ba:x.pɪf/; cf. Table below). This new set had a fixed syllable structure (CVC.CVC, only stops and fricatives) to reduce item variance and to facilitate syllable-level annotations.

	pseudowords Experiment 4
1	/b?x.pɪf/
2	/p?f.bʏx/
3	/b?x.kʏf/
4	/t?x.tɔs/

Citation: HansRutger.Bosker (09/30/2020). Audiovisual stimuli creation. https://dx.doi.org/10.17504/protocols.io.bmv3k68n

5	/t?f.pɛx/
6	/t?f.dɔs/
7	/p?g.dɔx/
8	/b?f.kix/

NOTE: The question mark in the IPA transcriptions in the table indicates the location of the 5-step vowel continua varying F2 from short /ɑ/ to long /a:/

10 The same male talker was recorded as before, producing these pseudowords in four versions: with short /ɑ/ and stress on the first syllable (e.g., *BAGpif*); with short /ɑ/ and stress on the second syllable (*bagPIF*); with long /a:/ and stress on the first syllable (*BAAGpif*); with long /a:/ and stress on the second syllable (*baagPIF*).

11 After manual annotation of syllable onsets and offsets, we measured the values of the suprasegmental cues to lexical stress (F0, amplitude, and duration) in all four conditions. Pseudowords with long /a:/ and stress on the first syllable were selected for manipulation. First, we manipulated the cues to lexical stress by setting these to ambiguous values using PSOLA in Praat: each first syllable was given the same mean value calculated across all stressed and unstressed first syllables (mean F0 = 154 Hz; original contour maintained; amplitude = 65.88 dB; duration = 263 ms), and each second syllable was given the same mean value across all stressed and unstressed second syllables (mean F0 = 159 Hz; original contour maintained; amplitude = 63.50 dB; duration = 414 ms). This resulted in manipulated pseudowords that were ambiguous in their prosodic cues to lexical stress. Since this included manipulating duration across all recorded conditions as well, it also meant that the duration cues to the identity of the first vowel were ambiguous.

12 Then, the first /a:/ vowel was extracted and manipulated to form a spectral continuum from long /a:/ to short /ɑ/. In Dutch, the /ɑ-a:/ vowel contrast is cued by both spectral (lower first (F1) and second formant (F2) values for /ɑ/) and temporal cues (shorter duration for /ɑ/; Bosker, 2017, Att Perc Psychophysics). We decided to create F2 continua since F2 is no cue to lexical stress in Dutch (in fact, in our original recordings, F2 values did not differ between stressed and unstressed conditions) while varying F2 does influence vowel quality perception. These spectral manipulations were based on Burg's LPC method in Praat, with the source and filter models estimated automatically from the selected vowel. The formant values in the filter models were adjusted to result in a constant F1 value (750 Hz, ambiguous between /ɑ/ and /a:/; value based on the original recordings) and 5 F2 values (step 1 = 1325 Hz, step 5 = 1125 Hz, step size = 50 Hz; values based on the original recordings). Then, the source and filter models were recombined. Finally, the manipulated vowel tokens were spliced into the pseudowords. Taken together, these manipulations resulted in pseudoword tokens that were ambiguous in lexical stress (average values of F0, amplitude, and duration), but varied in F2 as cue to vowel identity.

13 To create audiovisual stimuli, these manipulated pseudowords were spliced into the audiovisual stimuli from Experiment 1A. Once again, manipulating the silent interval between carrier sentence offset and target onset resulted in two different alignments of the last beat gesture to either first vowel onset or second vowel onset. Like in Experiment 1A, facial features of the talker were masked, and silent intervals as well as manipulated pseudowords were mixed with stationary noise to match the room acoustics across the sentence stimuli.

Citation: HansRutger.Bosker (09/30/2020). Audiovisual stimuli creation. https://dx.doi.org/10.17504/protocols.io.bmv3k68n