Listeners' predictions of sentence lengths are categorical, not gradient

Amanda Rysling (UC Santa Cruz), Charles Clifton (UMass Amherst), Nicholas Van Handel (UC Santa Cruz), Ria Geguera (UMass Amherst), Hannah Choi (UMass Amherst), & Anne Cutler (The MARCS Institute, Western Sydney University)
rysling@ucsc.edu

Evidence that listeners can predict the length of a sentence from its F0 declination first came from [5]'s demonstration that listeners can predict how many words followed a potential last word (PLW) of a sentence. Listeners heard words up to a PLW of an experimental sentence, underlined in (a). In the +0 condition, this PLW was the last word that the speaker had produced. In +3 or +6 conditions, (b) and (c), listeners heard words only up to the PLW, but those were taken from productions of sentences that continued for three or six words past the PLW. Early studies found that American English (AmE) native listeners reliably predicted more words for +6 than +3 than +0 conditions [5, 6], but native European French listeners did not [6]. F0 declination was implicated as the cue that listeners used after acoustic analyses found that only F0 range from beginning to PLW to reliably changed with sentence length. But recent work has failed to replicate different length predictions for +3 vs. +6 conditions by native Australian English (AuE), L1 Korean-L2 AuE, L1 Vietnamese-L2 AuE, native German, and L2 German listeners [7, 8]. This lack of difference is problematic for accounts that rely on F0 declination, because common speech production pressures make declination occur cross-linguistically.

We hypothesize that these failures to replicate arise, because the original +3 vs. +6 difference was due to Type I error, not because of special sensitivity to declination on the part of native AmE listeners. We present two identical pre-registered studies conducted at two North American research universities. The original study [5] employed a gating manipulation in which different-size portions of the PLW were presented. Since the largest effect of words after a PLW was found when the full PLW was presented, we used only that condition. A male native speaker of AmE produced all sentences, resulting in smaller F0 range than in [5, 6], but significant differences between declination ranges of different conditions (Table 1). Listeners (target 32 per study, as in [5]) heard all three conditions from each item, not just one from each item, allowing them greater familiarity with the speaker than listeners in [5, 6, 7]. Because [5, 6] found no sensitivity to a +6 vs. +9 contrast, +9 was not tested. On each trial, participants heard up to the PLW of an item, and saw text from corresponding +0, +3, and +6 conditions on a screen. They indicated whether the production they heard came from a +0, +3, or +6 version.

In two studies, native listeners of AmE demonstrated a clear insensitivity to the +3 vs. +6 contrast, but they could correctly state that no more words followed the PLW in +0 conditions (Figure 1 and Table 2; statistical analyses for the first instance of a condition from each item, as in [5, 6], were not different). Results and stimulus acoustic analyses (Table 1) from the present studies demonstrate that listeners do not simply use F0 declination to predict ultimate sentence lengths. Some F0 declination differences that are present, e.g. between +0 and +6 on the PLW, are apparently not leveraged. We argue that this is because listeners only generate prosodic predictions from cues of high validity [3]. Speakers can modulate their F0s very quickly, stopping or production-planning mid-sentence. Their rate of pitch change can thus be unreliable. Listeners' failures to leverage weaker sources of evidence to form gradient predictions parallels previous work, which demonstrated that native listeners of English do not leverage regularities in English stress for word segmentation, in contrast to native listeners of Dutch and Spanish, who use qualitatively comparable patterns of higher cue validity [2, 3, 4, 9]. We thus argue for an understanding of prosodic prediction under which listeners only form predictions for which they can have relatively high degrees of certainty.

Example experimental item

(a) **+0 condition**: Earlier my sister took a dip.
(b) **+3 condition**: Earlier my sister took a dip in the pool.
(c) **+6 condition**: Earlier my sister took a dip in the pool at the club.

	+0	**+3**	**+6**	**Significant differences**
F0 declination to last word in sentence (Hz)	50.4	42.4	36.0	+0 vs. +3, +0 vs. +6, +3 vs. +6
F0 movement on PLW (Hz)	1.2	11.6	16.5	+0 vs. +3, +0 vs. +6

Table 1 - Acoustic measures of F0 declination in stimuli sentences

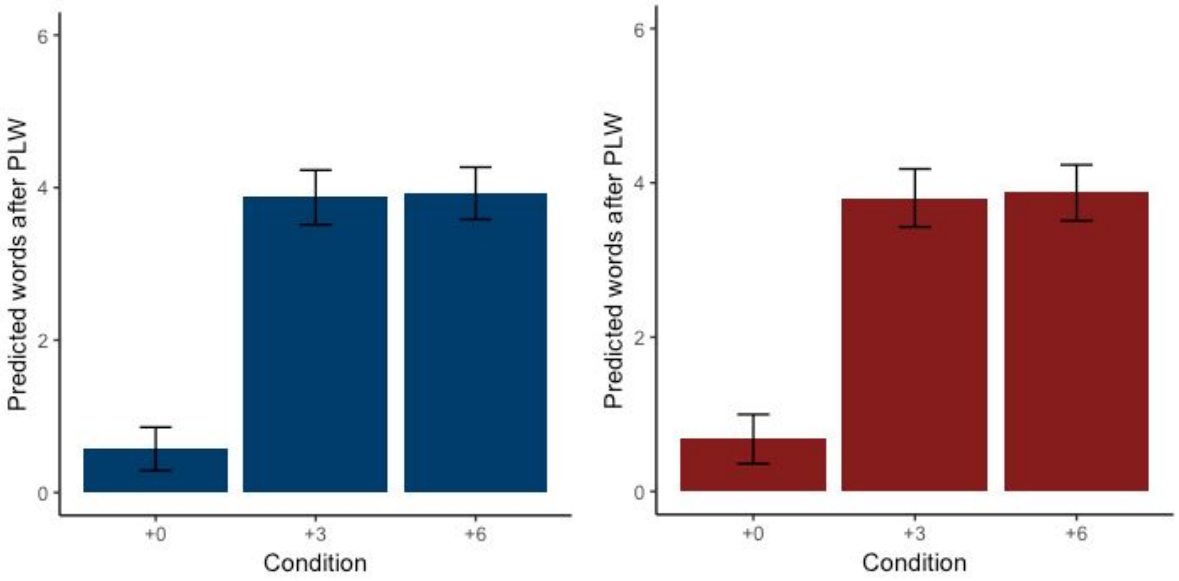

Figure 1 - Mean(s.e.) predicted words from replications 1 and 2, respectively

	Replication 1				**Replication 2**			
	Estimate	**s.e.**	**z**	***p***	**Estimate**	**s.e.**	**z**	***p***
0 vs. (+3 and +6)	**1.82**	**0.18**	**9.86**	**<0.001**	**1.34**	**0.05**	**28.75**	**<0.001**
+3 vs. +6	**-0.013**	**0.05**	**0.27**	**0.78**	**-0.03**	**0.04**	**0.77**	**0.44**

Table 2 - Fixed effects estimates from ordinal mixed effects regression [1]

[1] Christensen, R.H.B. (2012). [2] Cooper, N., Cutler, A., & Wales, R. (2002). [3] Cutler, A. (2012). [4] van Donselaar, W., Koster, M., & Cutler, A. (2005). [5] Grosjean, F. (1983). [6] Grosjean, F., & Hirt, C. (1996). [7] Jeske, J., Kember, H., & Cutler, A. (2016). [8] O'Brien, M. G., & Hendricks, A. K. E. (2013). [9] Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001).