THE USE OF SYNTACTICALLY REDUNDANT PHONETIC CUES IN SPEECH PERCEPTION

Clara Cohen

Glasgow University Laboratory of Phonetics, University of Glasgow
clara.cohen@glasgow.ac.uk

ABSTRACT

Subsegmental phonetic detail can be highly informative about upcoming word structure. For example, shortened stems tend to signal that suffixes or additional syllables will follow. Yet preceding morphosyntactic context can reduce the informational value of such cues. This study asks whether listeners still attend to phonetic detail in such contexts.

In a visual world eye-tracking experiment, 21 listeners distinguished singular target nouns named in audio recordings of English sentences from plural or two-syllable competitors. In the sentences, noun stems were lengthened or shortened, and positioned either after an agreeing determiner (*this/that*), rendering durational cues to plural suffixes redundant, or after the neutral determiner *the*, rendering them informative. GAMM analysis of gaze traces revealed that lengthening the stem facilitated perception, but only after agreeing determiners. This suggests not only automatic processing of redundant phonetic detail, but a requirement that such detail can be predicted from context before it can be processed.

Keywords: speech perception; eye-tracking; pronunciation variation; morphophonetics

1. INTRODUCTION

The duration of an initial syllable can carry substantial information about the structure of the rest of the word. In a pattern known as polysyllabic shortening, the first syllable in words like *captain* or *hamster* are shorter than they would be if they formed the stand-alone words *cap* and *ham* [1, 2]. Since morphological affixation often adds additional syllables, polysyllabic shortening can signal the presence of upcoming morphological structure, such as derivational suffixes.

Listeners are capable of drawing on these durational patterns during perception. They can use them to distinguish single-syllable words from multi-syllabic competitors [1–3], and simple words like *clue* from complex words like *clueless* [4–6]. Even when the addition of a suffix does not increase the syllable count of a word, subsegmental cues can still provide information about morphological structure [7–9]. The first goal of this study, therefore, focuses on English nouns, whose stems are shorter when a plural *-s* suffix is added [10], even when it does not change the syllable count. Can listeners use this durational cue to aid them in distinguishing singular nouns from plurals?

The second goal of this study concerns the role of morphosyntactic context in modulating the informational value of stem duration cues. Consider an utterance *Don't trip over the dogs!* This sentence is segmentally ambiguous regarding the number of canine obstacles until the final suffix is produced, and so keen attention to durational detail might afford listeners a head start on identifying challenges to safe locomotion. By contrast, if the warning took the form *Don't trip over those dogs!*, the listener would know even before the onset of the noun that multiple tripping hazards are in play. In this context, then, the stem-durational cue, which signals an upcoming plural suffix, is rendered redundant by the number information in the agreeing determiner. Since processing fast-moving, low-level phonetic detail is cognitively costly [11, 12], listeners may cease to attend to such stem duration cues in contexts where they offer no new information beyond what is already known from preceding morphosyntax.

An alternative perspective springs from research on incremental prediction. Listeners process real-time speech incrementally, and preactivate representations that are highly predictable in context [13–15]. When those predictions are violated, comprehension is impeded while the listener reanalyses the speech stream. Neurophysiological evidence suggests that this predictive preactivation may include not only semantic and phonological features, but also phonetic expectations [8, 16–18]. Such predictions may well include knowledge about durational variation patterns in noun stems. For example, if preceding morphosyntactic context leads listeners to expect a singular noun, then part of the expected phonetic form of a singular noun would include a lengthened stem. If so, then lengthening a singular noun stem would aid listener perception *more* in contexts where such lengthening is predictable, and hence informationally redundant, because it better matches the listener's expectations.

The current study tests these predictions with a visual world eye-tracking experiment. Listeners were presented with a four-quadrant visual display, containing a one-syllable singular target noun, a competitor noun (either plural or two syllables), and two distractors. They were asked to select the picture mentioned in a recorded sentence (e.g., *My view was blocked by* ***the cart*** *in front of the door*). In each sentence, the target singular noun (*cart*) was either lengthened, to match the expected duration of a singular one-syllable noun, or shortened; and positioned either after an agreeing determiner (*that cart*) or a non-agreeing determiner (*the cart*).

Lengthening the singular noun stem should help listeners distinguish one-syllable nouns from two-syllable competitors in all contexts, replicating previous findings on listeners' use of polysyllabic shortening cues during perception [1–6]. However, for plural competitors the perceptual benefits of lengthening may depend on context. If listeners disregard redundant phonetic detail, then lengthening the noun stem should help more in contexts when such detail is informative—i.e., after non-agreeing *the*. However, if phonetic detail forms part of the pre-activated predictions that assist incremental perception, then lengthening the noun stem should still facilitate perception even after agreeing determiners like *this/that*—perhaps even more strongly than after *the*, which does not preactivate phonetic expectations about singular noun stem duration.

2. METHODS

2.1. Materials

Sentences were built around a set of 84 one-syllable noun stems, which could appear in one of three forms: a singular (e.g., *cart*); a plural (*carts*); or a two-syllable carrier (*carton*), in which the singular was always lexically embedded as the first syllable.

All of these target words were presented in sentences, such as *My view was blocked by that cart in front of the door*. Each sentence contained a preamble (*My view was blocked by. . .*), determiner (*the/that/those*), target word (*cart/carts/carton*), and postamble (*in front of the door*). Sentences with singular targets formed the critical items, while plural and carrier targets served as fillers.

Each experimental list was formed by rotating a given item across eight experimental conditions formed by fully crossing three binary factors: Context, Competitor Type, and Duration Match. Context (agreeing/non-agreeing) encodes the determiner preceding the target noun. Agreeing determiners explicitly signaled the number of the following noun, taking the form *this* or *that*. Non-agreeing determiners were always *the*. Competitor Type (plural/carrier) encodes the competitor image in the visual display. A Plural competitor depicted two examples of the target noun (e.g., two *carts*), while the Carrier competitor a single example of the carrier word (e.g., one *carton*). Each list contained 40 critical and 40 filler sentences.

All sentences were produced by a native speaker of Scottish English, and recorded in six variants. These variants were created by fully crossing the three possible target noun forms (singular/plural/carrier) with the two determiner forms (agreeing *this/that* or *these/those* and non-agreeing *the*). The Duration Match (match/mismatch) manipulation was then generated by manipulating and splicing these raw recordings. First, the raw duration of the singular noun and its intended competitor type (plural or carrier) were averaged together. Then, 40ms were added to that average to calculate the target duration for Match stimuli, and 40ms were subtracted to calculate the target duration for Mismatch stimuli. Each singular noun stem was then adjusted by Praat script [19] to produce Match and Mismatch versions, which thus differed in duration by 80ms—a difference determined from pilot work to be perceptible without veering too far from actual patterns of polysyllabic shortening. Finally, the determiner, manipulated noun stem, and postamble were spliced onto a preamble from a third recording of the sentence, so as to remove any early cues as to the original nature of the target.

Filler sentences, containing plural and carrier targets, were manipulated in the same way, except that all filler targets contained only the shortened version of the nouns stem, to support the listener's expectation that stems with suffixes or second syllables would be shortened.

This procedure generated 8 versions of each sentence, which were rotated across experimental lists in a Latin Square, ensuring a fully crossed 2× 2 × 2 within-item and within-subjects design.

2.2. Participants and Procedure

Twenty-one native speakers of Scottish English (mean age 19.29; 16 female, 5 male) were recruited

Table 1: GAMM summaries for Target Advantage in sentences with plural (left) and carrier (right) competitors. Levels for Duration Match (abbreviated *DM*) are Match (reference, *m*) and Mismatch (*mm*); levels for Context (*Con*) are Agreeing (reference, *ag*) and Non-Agreeing (*na*). Difference smooths for carrier competitor GAMM were calculated on a four-level interaction between Context and Duration Match.

Plural competitors

Parametric	Est.	SE	*t*	*p*
Intercept	3.20	0.284	11.26	$< .001$
DM=mm	-1.14	0.17	-6.66	$< .001$
Con=na	-0.60	0.17	-3.49	$< .001$
DM=mm:Con=na	1.04	0.24	4.34	$< .001$

Difference smooths	edf	*F*	*p*
Window	4.64	20.76	$< .001$
Win:DM=mm	1.77	2.30	.14
Win:Con=na	2.55	3.18	$< .001$
Window, by subj	95.57	3.55	$< .001$

Carrier competitors

Parametric	Est.	SE	*t*	*p*
Intercept	1.53	0.23	6.78	$< .001$
DM=mm	-2.27	0.13	17.95	$< .001$
Con=na	0.82	0.13	6.45	$< .001$

Difference smooths	edf	*F*	*p*
Window	1.05	31.46	$< .001$
Win:DM=mm,Con=ag	4.41	13.46	$< .001$
Win:DM=m,Con=na	1.00	2.66	.10
Win:DM=mm,Con=na	3.38	5.44	$< .001$
Window, by subj	87.09	2.29	$< .001$

from the local university community. Participants started with a short demographic questionnaire, before continuing on to the eye-tracking procedure. The procedure started with a nine-point calibration and seven practice trials. After pausing to check participant comfort and answer questions, the tracker was recalibrated and the full experiment began. Each trial began with 500ms fixation cross, and then a 3500ms preview of the images, each of which was accompanied by a label. After 3500ms, the audio recording played, and participants selected the image named in the sentence. Each experiment contained 80 trials, presented in blocks of 20 with calibration break between each block.

Data were collected from the right eye by an Eyelink 1000+ at a 1000Hz sampling rate.

2.3. Analysis

Gaze data was analyzed from the onset of the determiner until 900ms after the offset of the target noun, when visual inspection of the raw data indicated that overall looks to target had peaked. Samples were binned into windows from the onset of the determiner. The proportion of looks to target and competitor within each bin was calculated by dividing the number of samples in which the gaze fixated on the target or competitor by the number of samples within the bin. Each determiner was divided into two bins, and each noun stem was divided into five bins. This ensured that the determiner offset was normalised to occur at the end of bin 2, and the noun offset occured at the end of bin 7. Post noun-offset, all bins were 50ms.

Proportions were converted to empirical logits [20], and eLog Target Advantage was calculated by subtracting the eLog looks to competitor from eLog looks to target in each window bin. Elog Target Advantage was then analysed with generalised additive mixed models from the `mgcv` package (version 1.8-40 [21]) in R (version 4.2.1 [22]). Duration Match and Context were included as both parametric terms and difference smooths, with random smooths included by subject. Terms were added sequentially to the null model, first as parametric terms and then as difference smooths, and only retained if their addition returned a significant difference in a Chi-squared test on the maximum likelihood scores when compared against the simpler model, as implemented in the `itsadug` package (version 2.4.1) [23]. Stimuli with carrier competitors were analysed separately from stimuli with plural competitors.

3. RESULTS

Table 1 provides final model summaries, and Figure 1 shows model predictions overlaid on observed Target Advantage across all time bins.

The plural competitor model revealed an interaction between Match and Context in the parametric terms (Table 1). In the parametric effects, Mismatch sentences showed overall lower Target Advantage than Match sentences when preceded by Agreeing determiners ($\beta = -1.13, p < .001$). Match sentences showed lower overall Target Advantage with Non-Agreeing determiners, relative to Agreeing determiners ($\beta = -0.60, p < .001$). The interaction between Match and Context, however, revealed that the disadvantage associated with Mismatch sentences was almost entirely undone with Non-Agreeing determiners ($\beta = 1.04, p < .001$). This can be observed in the left-hand panel of Figure 1. Although Agreeing contexts (top) show a decided Match effect, with higher Target Advantage for the

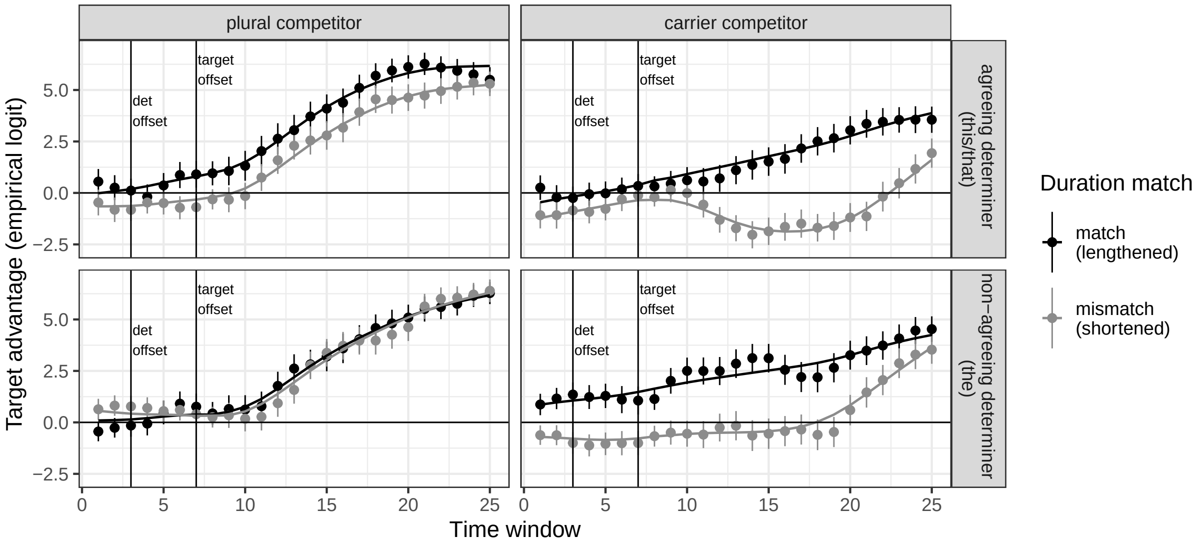

Figure 1: Target advantage over time for sentences with plural competitors (left) and carrier competitors (right), with targets placed after Agreeing determiners (top row) and Non-Agreeing determiners (bottom row).

black Match gaze trace than the grey Mismatch gaze trace, this effect disappears in Non-Agreeing contexts (bottom).

The carrier competitor model revealed a main parametric effect of Match and Context, but no interaction between them. The effect of Context reflects an overall advantage for sentences with Agreeing determiners ($\beta = 0.82, p < .001$), and an overall disadvantage for sentences with Mismatching—i.e., shortened—stem duration ($\beta = -2.27, p < .001$). The difference smooths revealed an interaction between those variables: Although there was no difference in curve shape between the gaze traces for Agreeing and Non-Agreeing sentences when they appeared in the Match condition, the gaze traces differed substantially in the Mismatch condition. As shown in Figure 1, Target Advantage for Mismatch sentences drops after the target offset in Agreeing sentences before rising again at the end of the interest period (top right), while Non-Agreeing sentences (bottom right) show a flatter trajectory and before the late rise.

4. DISCUSSION AND CONCLUSION

The results of this experiment show that listeners can use subsegmental detail to aid in processing both syllable structure and morphological structure. When distinguishing one-syllable nouns from two-syllable competitors, listeners showed a marked advantage when the target one-syllable nouns were lengthened according to patterns of polysyllabic shortening, replicating studies [2, 3]. This Duration Match effect did not interact with context, which is also expected. Since targets and carrier competitors were both singular, the presence of an agreeing determiner would not affect listeners' predictions or expectations about the phonetic or morphosyntactic properties of the upcoming noun.

When distinguishing one-syllable singular nouns from one-syllable plurals, however, the situation was different. Listeners were better at identifying the singular target when the stem was lengthened, but this pattern only obtained after agreeing determiners. After non-agreeing determiners—exactly the context where such detail might be expected to carry the most valuable information about the target noun's form—listeners seemed insensitive to the duration of the target noun stem.

This interaction is consistent with models of online speech perception which hold that listeners form phonetically detailed predictions about upcoming material [8, 16–18]. When those predictions are confirmed by the incoming speech stream, listeners are faster to identify the target word. After non-agreeing determiners, by contrast, listeners could not form predictions about the number of the target noun, and so could not effectively use the duration information in the stem.

In sum, these findings suggest that listeners readily process durational cues that signal upcoming lexical word structure. Cues which signal upcoming morphosyntactic information, however, are only processed when they are syntactically redundant, and only valuable inasmuch as they confirm pre-existing phonetic expectations.

5. REFERENCES

[1] M. H. Davis, W. D. Marslen-Wilson, and M. G. Gaskell, “Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition,” *Journal of experimental psychology: Human perception and performance*, vol. 28, no. 1, pp. 218–244, 2002. [Online]. Available: http://dx.doi.org/10.1037/0096-1523.28.1.218

[2] A. P. Salverda, D. Dahan, and J. M. McQueen, “The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension,” *Cognition*, vol. 90, no. 1, pp. 51–89, 2003. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0010027703001392

[3] K. B. Shatzman and J. M. McQueen, “Prosodic knowledge affects the recognition of newly acquired words,” *Psychological Science*, vol. 17, no. 5, pp. 372–377, 2006.

[4] L. J. Blazej and A. M. Cohen-Goldberg, “Can We Hear Morphological Complexity Before Words Are Complex?” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 41, no. 1, pp. 50–68, 2015.

[5] R. J. J. K. Kemps, M. Ernestus, R. Schreuder, and R. H. Baayen, “Prosodic cues for morphological complexity: the case of Dutch plural nouns.” *Memory & Cognition*, vol. 33, no. 3, pp. 430–446, 2005.

[6] R. J. J. K. Kemps, L. H. Wurm, M. Ernestus, R. Schreuder, and R. H. Baayen, “Prosodic cues for morphological complexity in Dutch and English,” *Language and Cognitive Processes*, vol. 20, no. 1/2, pp. 43 – 73, 2005.

[7] R. Smith, R. Baker, and S. Hawkins, “Phonetic detail that distinguishes prefixed from pseudo-prefixed words,” *Journal of Phonetics*, vol. 40, no. 5, pp. 689–705, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.wocn.2012.04.002

[8] A. Hjortdal, J. Frid, and M. Roll, “Phonetic and phonological cues to prediction: Neurophysiology of Danish stød,” *Journal of Phonetics*, vol. 94, sep 2022.

[9] M. Clayards, M. G. Gaskell, and S. Hawkins, “Phonetic detail is used to predict a word’s morphological composition,” *Journal of Phonetics*, vol. 87, jul 2021.

[10] C. Cohen and M. Carlson, “Shifting between storage and computation in lexical retrieval: Evidence from pronunciation variation,” in *Interfaces of Phonetics*, M. Schlechtweg, Ed. De Gruyter, to appear.

[11] M. H. Christiansen and N. Chater, “The Now-or-Never Bottleneck: A Fundamental Constraint on Language.” *Behavioral and Brain Sciences*, vol. 39, pp. 1–52, 2016. [Online]. Available: http://journals.cambridge.org/abstract_S0140525X1500031X

[12] G. R. Kuperberg and T. F. Jaeger, “What do we mean by prediction in language comprehension?” *Language Cognition & Neuroscience*, vol. 31, no. 1, pp. 32–59, 2016. [Online]. Available: http://dx.doi.org/10.1080/23273798.2015.1102299

[13] G. T. M. Altmann and Y. Kamide, “Incremental interpretation at verbs: Restricting the domain of subsequent reference,” *Cognition*, vol. 73, no. 3, pp. 247–264, 1999.

[14] M. Kutas and K. D. Federmeier, “Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP),” *Annual Review of Psychology*, vol. 62, no. 1, pp. 621–647, 2011. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev.psych.093008.131123

[15] A. Ito, M. J. Pickering, and M. Corley, “Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study,” *Journal of Memory and Language*, vol. 98, pp. 1–11, 2018. [Online]. Available: https://doi.org/10.1016/j.jml.2017.09.002

[16] R. L. Newman and J. F. Connolly, “Electrophysiological markers of pre-lexical speech processing: Evidence for bottom-up and top-down effects on spoken word processing,” vol. 80, pp. 114–121, 2009.

[17] R. C. N. D’Arcy, J. F. Connolly, E. Service, C. S. Hawco, and M. E. Houlihan, “Separating Phonological and Semantic Processing in Auditory Sentence Processing: A High-Resolution Event-Related Brain Potential Study,” *Human Brain Mapping*, vol. 22, no. 1, pp. 40–51, 2004.

[18] J. F. Connolly and N. A. Phillips, “Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences,” *Journal of Cognitive Neuroscience*, vol. 6, no. 3, pp. 256–266, 1994.

[19] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2015. [Online]. Available: http://www.praat.org

[20] D. J. Barr, “Analyzing ’visual world’ eyetracking data using multilevel logistic regression,” *Journal of Memory and Language*, vol. 59, no. 4, pp. 457–474, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.jml.2007.09.002

[21] S. N. Wood, “Thin-plate regression splines,” *Journal of the Royal Statistical Society (B)*, vol. 65, no. 1, pp. 95–114, 2003.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[23] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, “itsadug: Interpreting time series and autocorrelated data using gamms,” 2022, r package version 2.4.1.