EARLY GANONG EFFECTS

Amanda Rysling[a], John Kingston[a], Adrian Staub[b], Andrew Cohen[b], and Jeffrey Starns[b]

[a]Linguistics, [b]Psychological and Brain Sciences, University of Massachusetts
arysling or jkingston@linguist.umass.edu; astaub or acohen or jstarns@psych.umass.edu

ABSTRACT

In the first of two experiments measuring Ganong effects, listeners were permitted to respond when they wished, while in the second, they were prompted to respond within a short interval at varying delays after stimulus onset. Both showed that lexical items were activated early and that their activation strength grew with response time and persisted after the effects of the stimuli's acoustics began to fade. These results show that Ganong effects are not produced by late decision processes.

Keywords: Ganong effects; response times; response-signal paradigm

1. INTRODUCTION

Listeners prefer to identify ambiguous members of a word-nonword continuum between two phonological categories as the category corresponding to the word endpoint [4, 11]. This lexical preference or "Ganong effect" has both informed understanding of how listeners apply their lexical knowledge in perceiving speech and fueled debate about whether lexical activation feeds back on pre-lexical processing. Ganong [4] himself conceived and interpreted his experiments as testing whether listeners' lexical knowledge altered their percepts of the stimulus's acoustic properties, as expected by an interactive model [6], or instead induced a post-perceptual response bias, as expected by an autonomous model [9]. He argued that their lexical knowledge altered their percepts because it shifted responses toward the word category only to intermediate, ambiguous stimuli and not also unambiguous stimuli near the continuum's endpoints.

Both the occurrence and strength of the Ganong effect has been shown to depend on when the listener responds, but not consistently. For initial consonant targets, the size of the lexical effect on response proportions has been shown to be stronger for slow RTs than for fast RTs [3, 8], while for final consonant targets, the lexical preference is instead stronger in fast than slow responses [7]. In a more recent study [11], however, the strength of the lexical preference didn't vary with response time for three initial consonant targets and was only stronger in fast than slow responses for one out of two final consonant targets. Although one study [2] has found word responses to be faster than non-word responses to ambiguous, intermediate stimuli, but not to differ for unambiguous endpoints, another [11] found word responses to be faster than nonword responses to endpoint and not intermediate stimuli. These inconsistencies undermine any attempt to use these results to support either interactive or autonomous models.

Table 1: *Freq*uencies (counts per million), frequency-weighted neighborhood *dens*ities, *for*ward [p(aɪ|C)], *back*ward (p(C|aɪ)) CV transitional probabilities, Subtlex, IPhOD [1, 12].

word	*freq*	*dens*	CV	*for*	*back*
file	44.04	2975.63	faɪ	0.0584	0.0616
side	200.92	2065.53	saɪ	0.0241	0.0445

We undertook two studies to resolve these inconsistencies. In the first, listeners could choose when to respond, while in the second, they were prompted to respond at particular moments and were severely limited in the amount of time they had to respond. By varying when the prompt occurred, we could collect responses that differed in how much of the stimulus listeners had heard and thus how much information they had obtained about which category made a word.

2. EXPERIMENT 1

2.1. Method

2.1.1. Stimuli

The target sounds in this experiment consisted of 12 steps selected from a 20-step /s–f/ continuum. They occurred before one of three –VC rimes: *–ide* (/aɪd/), where /s/ but not /f/ made a word (*side, *fide*), *–ile* (/aɪl/), where /f/ but not /s/ made a word (**sile, file*), and *–ime* (/aɪm/), whether neither /s/ nor /f/ made a word (**sime, *fime*). These materials were chosen to minimize the influence of various lexical statistics on listeners' responses (Table 1).

The rimes, *–ide, –ile, –ime*, were recorded with

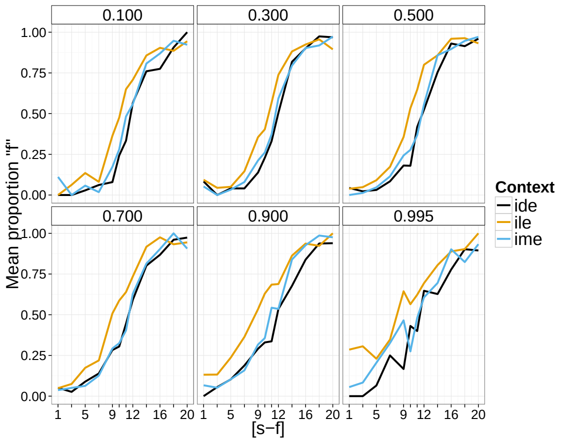


Figure 1: Mean proportion of "f" responses by RT quantile bin from fastest (0.005-0.1, upper left) to slowest (0.9-0.995, lower right), context, and step.

the onset /h/, i.e. *hile, hide, hime*, so there were no formant transitions into the vowel that would convey information about the preceding consonant's place of articulation. Information about the final consonant became audible 215-225 ms into the rimes. Listeners could have used vowel nasalization in the nucleus of the *–ime* rime, and a higher pitch in the *–ile* to anticipate the final consonant; however, nasalization must persist for a relatively long period to be detected [13], and the pitch difference between the *–ile* rime and the other two rimes was most pronounced about the same time as acoustic evidence about the final consonant became audible. The /s–f/ continuum was made by mixing in complementary proportions tokens of /s/ and /f/ that exhibited the largest differences in energy between a lower frequency 0-4000 Hz band and a higher frequency 4000-8000 Hz band: higher – lower = 15.99 dB for /s/ and −5.92 dB for /f/. The fricatives were shortened to 150 ms, so the final consonant that informed the listener as whether /s/, /f/, or neither made a word became audible 365-375 ms after stimulus onset.

2.1.2. *Procedure*

Each participant was trained with correct answer feedback on the endpoints in the no-bias *–ime* context. In testing, each listener responded 8 times to each endpoint, 16 times to the two stimuli nearest the endpoints, and 24 times to the remaining eight intermediate stimuli in each of the three contexts, for a total of 720 test trials. Correct answer feedback was also given to endpoint stimuli during testing. 27 adult native speakers of North American English dialects participated. None reported any speaking or hearing disorders. They provided informed consent.

Table 2: Fixed effects estimates, including all two-way interactions of step, context, and RT quantile bin.

	Est	*se*	*z*	*p*
Int	-0.575	0.232	-2.480	0.013
Step	2.790	0.177	15.736	< 2e-16
RT	0.291	0.089	3.264	0.001
Step:RT	-0.435	0.038	-11.529	< 2e-16
s- vs no-	-0.264	0.119	-2.215	0.0268
f- vs no-	0.985	0.211	4.665	3.09e-06
Step:s- vs no-	-0.031	0.090	-0.348	0.728
Step:f- vs no-	-0.187	0.087	-2.145	0.032
s- vs no-:RT	-0.049	0.056	-0.879	0.380
f- vs no-:RT	0.079	0.057	1.377	0.169

2.2. Results

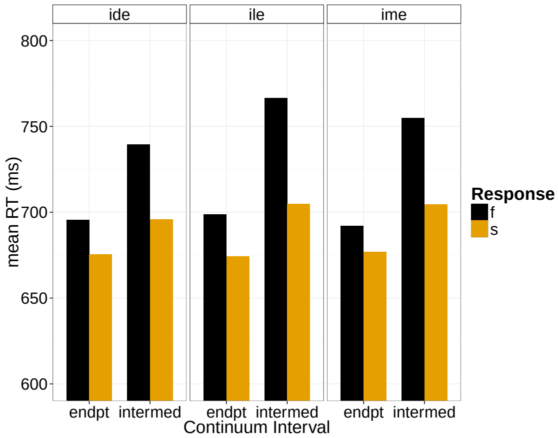
2.2.1. *Response proportions*

Response time (RT) quantile bins (0.005–0.1, 0.1–0.3, 0.3–0.5, 0.5–0.7, 0.7–0.9, 0.9–0.995) were calculated separately for each listener and step, and responses faster than the 0.005 quantile and slower than the 0.995 quantile were discarded. Figure 1 shows that in all RT quantile bins, listeners responded "f" considerably more often in the f-bias *–ile* context than the other two contexts. They also responded "f" less often in the s-bias *–ide* than the no-bias *–ime* context, but this difference was smaller. Slower responses were less categorical: near the endpoints, responses corresponded less reliably to the nearby endpoint category, and they crossed over less abruptly from one category to the other.

A mixed effects logistic regression model including step, treatment coding of context (s-bias *–ide* vs no-bias *–ime* and f-bias *–ile* vs no-bias *–ime*), RT quantile bins and all two-way interactions between these fixed effects was fit to the relative proportion of "f" responses. All continuous fixed effects were centered (by subtracting their means) and scaled (by dividing by their standard deviations), in this and all subsequent models. Random effects of listener on the intercept and the slopes of the individual fixed effects, but not their interactions, were included, in this and all subsequent models.

This model fit better than one lacking these interactions ($X^2(6) = 178.78, p < 20.2e - 16$), and the fit was not improved by adding the three-way interactions between the fixed effects ($X^2(2) = 1.5658, p = 0.4571$). Besides the main effects of step and lexical preference expected from Figure 1, the estimates in Table 2 show that listeners responded "f" more often as they responded slower, but less often for slower responses as the fricative became more /f/-like. The

Figure 2: Mean response times (ms) by context, endpoint versus intermediate continuum intervals, and response.



effect of step didn't differ between the s-bias *–ide* and the no-bias *–ime* contexts, but "f" responses decreased as the stimulus became more /f/-like in the f-bias *–ile* compared to the no-bias *–ime* context.

2.2.2. Response times

The continuum was divided into a continuous interval spanning intermediate, ambiguous stimuli, steps 7-12, and a discontinous interval including more unambiguous stimuli, steps 1-5 near the /s/ endpoint and steps 14-20 near the /f/ endpoint. Figure 2 shows that RTs were uniformly slower for intermediate than endpoint stimuli across the three contexts, cf. [10], that they are also uniformly slower for "f" than "s" responses across contexts and within both endpoint and intermediate intervals, and that the difference in RT between "f" and "s" responses was larger for intermediate than endpoint stimuli.

A mixed effects linear model including the endpoint versus intermediate intervals, context, and response, and the interaction between intervals and response as fixed effects fit the log RTs better than a model without this interaction ($X^2(1) = 50.954, p = 8.457e-13$) and was only slightly improved by adding the other two-way and three-way interactions ($X^2(6) = 12.898, p = 0.04469$). Moreover, none the estimates for the additional interactions reached significance ($|t| < 2$). In the model with just the single interaction, the estimates for endpoint versus intermediate steps and for "f" versus "s" responses were both significantly positive, as was that for their interaction, thus confirming that "f" responses were slowed more than "s" responses for intermediate compared to endpoint stimuli.

2.3. Discussion

The lexical preferences were greater for slower responses, cf. [3], but clearly present in the fastest responses, too. The effect of the target sounds' acoustics also diminished as responses slowed. This combination suggests that words were activated quickly, and that their activation persists after memory of the sounds' qualities has faded. RTs were not faster for word than nonword responses for either endpoint or intermediate stimuli, but instead uniformly slower for intermediate than endpoint stimuli, and for "f" than "s" responses. The latter difference was greater for intermediate than endpoint stimuli.

3. EXPERIMENT 2

The second experiment is, so far as we know, the first speech perception experiment to use the response-signal paradigm. In our use of this paradigm, listeners were prompted (signaled) to respond at different times on every trial and were limited in the amount of time they had to respond. The earliest prompt was chosen to tap listeners' perceptual experience at the moment when lexically relevant acoustic information first became audible, while later ones successively tapped the subsequent evolution of that experience.

3.1. Method

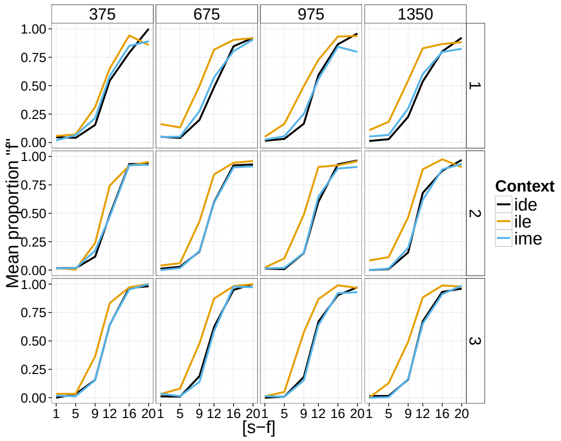
3.1.1. Stimuli

The endpoints and four intermediate stimuli were reused from the continuum used in Experiment 1.

3.1.2. Procedure

The only substantial change in procedures was that responses were prompted randomly at 375, 675, 975, or 1350 ms after stimulus onset (henceforth "delays") by displaying "Go!" in the middle of the screen, and listeners had just 300 ms to respond after this prompt appeared. The earliest prompt coincided with when lexically relevant acoustic information about the final consonant became audible, the next earliest right after the stimulus ended, and the two later ones at increasingly noticeable delays afterward. These intervals approximate the mean 0.005, 0.25, 100 ms after the 0.75, and the 0.995 RT quantiles, respectively, from Experiment 1, and thus sample the range from fastest to slowest responses in that experiment. Listeners were informed after their response whether it was on-time, too fast, or too slow, and the cumulative total of on-time, too fast, and too slow responses was displayed at the end of each test block. In three one-hour sessions on different days, the endpoint stimuli were presented 8 times and the

Figure 3: Mean "f" response proportions by response prompt delay, session, context, and step.



intermediate stimuli 25 times in each of the three contexts at each of the four response prompt delays, for a total of 1392 test trials/participant. 29 participants who met the same criteria as Experiment 1 gave informed consent and participated.

3.2. Results

Only on-time responses were included in the analysis, overall a little more than 0.83 of the responses.

Figure 3 shows that the lexical preference for "f" responses in the *–ile* context influenced responses even at the earliest delay, that this lexical preference was nonetheless greater at all later delays, that responses became less categorical as the delay got later, much as they did as responses got slower in Experiment 1, but that they became more categorical at those later delays in later sessions, and finally that the lexical preference for "s" responses in the *–ide* compared to the *–ime* context disappeared by the second session.

A mixed effects logistic regression model including step, context, delay, and session and the two-way interactions between context and delay, context and session, and delay and session fit the relative "f" response proportions better than one lacking these interactions ($X^2(5) = 54.121, p = 1.979e-10$). Besides the effects of step and lexical preference expected from Figure 3, the estimates in Table 3 show that listeners responded "f" more often in the f-bias *–ile* vs no-bias *–ime* context at longer delays. The marginal interactions between delay and session and between s-bias *–ide* vs no-bias *–ime* and session reflect the trends visible in Figure 3 toward more categorical responses for later delays in later sessions and the disappearance of the lexical preference for "s" responses in the s-bias *–ide* context, respectively.

Table 3: Fixed effects estimates for a model including the two-way interactions context, delay, and session.

	Est	*se*	*z*	*p*
Int	-0.657	0.309	-2.127	0.034
Step	4.109	0.301	13.639	$<$ 2e-16
s- vs no-	-0.043	0.185	-0.232	0.816
f- vs no-	1.613	0.221	7.289	3.13e-13
Del	0.003	0.042	0.065	0.948
Sess	0.080	0.103	0.776	0.438
s- vs no-:Del	0.026	0.049	0.534	0.594
f- vs no-:Del	0.308	0.050	6.198	5.72e-10
Delay x Sess	-0.039	0.020	-1.950	0.051
s- vs no-:Sess	0.080	0.048	1.665	0.096
f- vs no-:Sess	0.019	0.049	0.386	0.700

3.3. Discussion

The lexical preference for "f" responses in the *–ile* context was stronger at later response prompt delays, but it was nonetheless clearly present at the earliest delay. Otherwise, as with longer RTs in Experiment 1, the perceptual effects of lexical preferences persisted as delays got later while those of stimulus acoustics faded.

4. GENERAL DISCUSSION

These experiments show that lexical items were activated early, as soon as the relevant acoustic information became audible. These lexical preferences both strengthened over time and lasted longer than those influenced by the stimuli's acoustics themselves. In an eye tracking experiment with the same stimuli [5], novel fixations to the letter corresponding to the word category increased significantly as soon as the relevant acoustic information became audible, too, but that lexical preference was outlasted by the influence of the stimuli's acoustics.

The immediacy of the lexical preference might appear to support interpreting Ganong effects as modifications of pre-lexical processing rather than post-perceptual decisions, cf. [4, 6], but nothing in an alternative autonomous model [9] requires lexical activation to be slow. The models instead differ architecturally: in TRACE, activation of a lexical item feeds *back* on the pre-lexical processing and thus alters the activation strength of phonological categories, while in Merge, the signal feeds strictly *forward* through pre-lexical to lexical processing, which feeds forward in turn to a task-specific category decision process. These results do not, therefore favor interactive over autonomous models, even if they rule out treating Ganong effects as products of late decision processes.

5. REFERENCES

[1] Brysbaert, M., New, B. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods* 41, 977–990.

[2] Connine, C. M., Clifton, C. E. 1987. Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 13, 291–299.

[3] Fox, R. A. 1984. Effects of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance* 10, 526–540.

[4] Ganong, W. F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6, 110–125.

[5] Kingston, J., Staub, A., Levy, J. submitted. Eye movement evidence for an immediate Ganong effect.

[6] McClelland, J., Elman, J. L. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18, 1–86.

[7] McQueen, J. M. 1991. The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance* 17(2), 433–443.

[8] Miller, J. L., Dexter, E. R. 1988. Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 14, 369–378.

[9] Norris, D., McQueen, J. M., Cutler, A. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23, 299–370.

[10] Pisoni, D. B., Tash, J. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15, 289–290.

[11] Pitt, M. A., Samuel, A. G. 1993. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 19, 699–725.

[12] Vaden, K. I., Halpin, H. R., Hickok, G. S. 2009. Irvine phonotactic online dictionary, version 2.0.

[13] Whalen, D. H., Beddor, P. S. 1989. Connections between nasality and vowel duration and height: Elucidation of the Eastern Algonquian intrusive nasal. *Language* 65, 457–486.