Chapter 2

Title: Now you hear me, now you don't: perception of highly lenited Chilean Spanish approximants and its implications for lexical access models[1]

Short Title: Perception of lenited Chilean Spanish approximants: implications for lexical access models

Título: Ahora me escuchas, ahora no: percepción de consonantes aproximantes altamente elididas del castellano chileno y sus implicaciones para modelos de acceso léxico

Abstract: Chilean Spanish is special in that it displays particularly high degrees of lenition and elision of [β], [ð̞] and [ɣ̞] (Pérez, 2007). Interestingly, Chilean Spanish listeners can recover elided units effortlessly, which challenges the assumptions of some lexical access models, such as strong bottom-up abstractionist models (Mitterer & Ernestus, 2006). This proposal reports on a series of perception experiments in which synthetic continua from full approximants to elided variants were presented in several informational conditions. Results showed that increasing the amount of acoustic information and the number of semantic cues had a significant effect on listeners' responses, enabling lexical effects and minimizing phonological recovery. Moreover, these effects were different for the three consonants being tested, probably due to existing links between production and perception. These findings are discussed in light of previous research on lexical effects and recovery, and lexical access models in general.

Resumen: El castellano hablado en Chile es singular porque presenta grados particularmente altos de lenición y elisión de [β], [ð̞] and [ɣ̞] (Pérez, 2007). Curiosamente, los oyentes pueden recuperar estas unidades elididas sin esfuerzo, lo que problematiza las asunciones de algunos modelos de acceso léxico, como de los fuertemente abstraccionistas y de abajo-hacia-arriba (Mitterer & Ernestus, 2006). Esta propuesta reporta los resultados de una serie de experimentos en los que se presentaron continuos desde consonantes aproximantes hasta variantes elididas, en varias condiciones informacionales. Los resultados mostraron que aumentar la cantidad de información acústica y el número de claves semánticas tuvo un efecto significativo en las respuestas de los oyentes. Además, estos efectos fueron diferentes para las tres consonantes evaluadas, probablemente dados ciertos vínculos entre los dominios de producción y percepción. Estos hallazgos se discuten a la luz de literatura sobre efectos léxicos, recuperación fonológica y modelos de acceso léxico en general.

[1] This chapter is based on "Chapter 6. Recovery, lexical effects and lexical access in /b d g/" of the first author's Doctoral Dissertation (see Figueroa Candia, 2016).

Keywords: highly lenited units, lexical access, perception, phonological recovery, lexical effects

Palabras clave: unidades altamente elididas, acceso léxico, percepción, recuperación fonológica, efectos léxicos

Author 1 (correspondence author):

Name: Figueroa Candia, Mauricio A.

Affiliation: Universidad de Concepción, Chile

Email: maufigueroa@udec.cl

Author 2:

Name: Evans, Bronwen G.

Affiliation: University College London, United Kingdom

Email: bronwen.evans@ucl.ac.uk

2.1. Introduction

Under normal circumstances, listeners are often required to achieve lexical access for word forms for which they lack sufficient acoustic evidence (Mitterer & Ernestus, 2006). Highly lenited and elided variants are indeed the norm in conversational speech (e.g., Janse, Nooteboom, & Quené, 2007; Torreira & Ernestus, 2011; Brown, 2011), and one consequence of this is that the acoustic variables cueing some segments are often poorly represented in the signal or completely absent from it. Despite these challenges, communication does not seem to be hindered by elision or lenition (Ernestus, 2014); on the contrary, listeners are capable of interpreting the perceptual input most of the time.

This raises a series of questions pertaining to the sources of information and strategies that listeners employ to attain lexical access. One important strategy to aid perception is phonological recovery, whereby underlying representations for missing segments are formed despite lacking full prelexical support, provided certain conditions are met (Samuel, 1981; Samuel, 1996). For example, studies have shown that listeners can resort to coarticulatory cues to aid perception, taking advantage of coarticulatory information from segments preceding or following missing or masked ones in order to recover them (Yeni-Komshian & Soli, 1981; Repp, 1983). Listeners can also employ semantic and

syntactic cues to achieve phonological recovery, especially when the acoustic cues are unreliable (e.g., Kemps, Ernestus, Schreuder, & Baayen, 2004; Mitterer & Ernestus, 2006).

Describing the conditions required for listeners to recover missing units from lenited and elided word forms is relevant because it has direct consequences for models of lexical access and speech perception[2]. For instance, proponents of episodic models of speech perception such as LAFS (Klatt, 1979; 1989) and Minerva 2 (Hintzman, 1984; Goldinger, 1998) claim that episodic models are able to account for lexical access of lenited forms given that reduced word forms have their own episodic representations in long-term memory, which are activated to match the acoustic input when required, without having to resort to intermediate abstract representations or phonological recovery. These models, however, have been challenged by evidence showing that listeners are unable to recover highly lenited word forms unless additional context is provided (e.g., Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004). On the other hand, bottom-up abstractionist models such as Cohort (e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987) and Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997), with no top-down lexical feedback to prelexical stages of speech processing, are also challenged by phonological recovery evidence, because these models require intermediate abstract representation to be formulated based on reliable acoustic evidence, which is not available in highly lenited forms. Interactive models of lexical access such as TRACE (McClelland & Elman, 1986a, 1986b), and hybrid models such as Goldinger's Complementary Learning Systems (CLS) (Goldinger, 2007), Pierrehumbert's Exemplar Dynamics (ED) (Pierrehumbert, 2001, 2002) and POLYSP (Hawkins & Smith, 2001; Hawkins, 2003), seem to fare better at accommodating experimental evidence in favour of recovery by positing that top-down feedback from the lexical level can inform prelexical stages of speech processing.

Although the hypothesis of phonological recovery itself is still debatable, the fact remains that listeners are overwhelmingly successful at dealing with highly lenited forms when complementary sources of information compensate for the lack of reliable acoustic evidence. The specific way in which these additional sources of information contribute to achieve lexical access has been the focus of research on highly lenited forms from conversational speech. For example, Ernestus, Baayen and Schreuder (2002) presented word forms that varied in their degree of lenition and carefully controlled the amount of acoustic and semantic cues available to the listeners. The results demonstrated that listeners were able to recognize highly reduced word forms when phonetic, semantic and syntactic contexts facilitated

[2] For some useful reviews on lexical access models and models of speech perception we can recommend McQueen (2005), McQueen, Cutler & Norris (2006), Ernestus (2014) and Figueroa Candia (2016).

lexical access. Another experiment by Kemps, Ernestus, Schreuder and Baayen (2004), showed that listeners were able to recover underlying /l/ from highly reduced instances of the Dutch suffix "-(e)lijk" [(ə)lək] in phoneme monitoring tasks, but only when the reduced suffixes were presented in a context of several words. Mitterer and Ernestus (2006) also conducted a series of perception experiments in which they presented synthetic instances of Dutch /t/ in coda position with varying degrees of lenition in several phonetic contexts, and in several semantic and syntactic contexts; their results showed that listeners utilized both bottom-up (phonological context and sub-phonemic detail) and top-down sources of information (lexical status) in order to attain lexical access.

Chilean Spanish spirant approximant variants of /b d g/ provide a novel testing ground to explore some of these issues surrounding the nature of phonological recovery, and how listeners weight different cues depending on their availability. This is because [β̞], [ð̞] and [ɣ̞] display natural continua of realizations from open approximants to elided variants, with /d/ displaying the highest rates of lenition, and /g/ the lowest (Cepeda & Poblete, 1993; Pérez, 2007). While most of the research on highly lenited forms has presented listeners with categorical increments of acoustic information (e.g. Ernestus, Baayen & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004), spirant approximants of Spanish allow the amount of acoustic evidence available in perception to be carefully controlled. Experiments can thus be designed in which increasing amounts of acoustic detail can be presented, enabling exploration of the role of fine-grained phonetic detail in the perception and phonological recovery of lenited forms.

Another important advantage of this particular form of variation is that, for some minimal pairs, both the presence of an approximant consonant and its absence constitute words. For example, eliding the approximant consonant from the word *dudo* [ˈdu.ð̞o] ("to doubt") renders [ˈdu.o], which can be interpreted by the listener as *dudo*, with lenited /d/, or as *dúo* ("duet"). Minimal pairs such as these allow the creation of ecologically valid continua from consonant presence to absence, while at the same time controlling for some lexical effects on speech perception, which would otherwise bias perception towards words as opposed to nonsense words (Ganong, 1980). Crucially, how listeners process these minimal pairs offers a transparent way to determine whether they actually perform phonological recovery, since the underlying phonological unit is the only difference between the two items.

In order to systematically evaluate the way in which listeners are able to process and interpret varying acoustic information, synthetic continua from full approximants to elision were presented to subjects in three tasks (a phoneme monitoring task, an identification task, and a discrimination task), each one with

four informational conditions: the first, “segmental level”, contained only acoustic information about the target segments and their immediate phonetic context; the second, “word level”, contained minimal lexical and semantic information; and the third and fourth ones, “primed approximant” and “primed elision”, included semantic primes for both ends of each continuum. In summary, this chapter aims to, first, establish an auditory perceptual baseline for the perception of approximant consonants of /d/ and /g/. Second, to determine the effect of increasing the number and type of acoustic and non-acoustic cues in the perception of these continua. Third, to determine whether there is evidence of phonological recovery in the perception of /d/ and /g/ in any of the experimental conditions. Finally, the results will be discussed in light of lexical access models and models of speech perception.

2.2. Methods

2.2.1. Participants

Sixty one native monolingual Chilean Spanish speakers (mean age 21.1 years; 42 females and 19 males) took part in the experiments, which consisted of two sessions lasting around 1 hour each and taking place on different days. Participants were undergraduate students, residents of large Chilean urban centres. Participants received an information sheet, and were required to give informed consent before the start of the experimental sessions. None of the participants reported having any cognitive, hearing, language or speech impairment. Participants were paid for their participation.

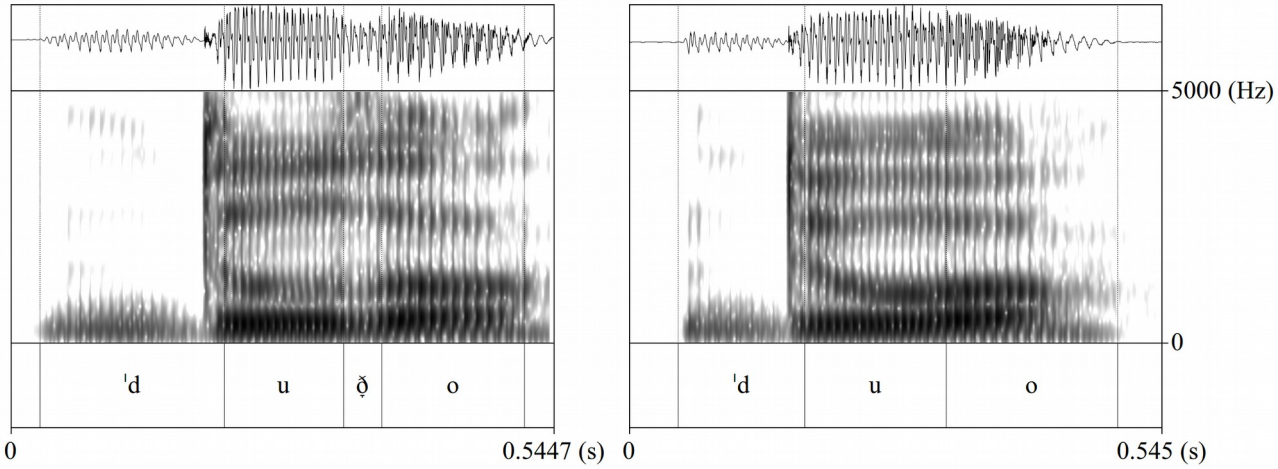
2.2.2. Stimuli

Several minimal pairs such as *dudo* [ˈdu.ð̞o] (“to doubt”) and *dúo* [ˈdu.o] (“duet”), in which eliding an intervocalic approximant consonant from the first word results in a different lexical unit, were identified for /d/ and /g/. These pairs and each item's relative lexical frequency were extracted from the Spanish lexical frequency list CREA (Real Academia Española, 2014). Minimal pairs in which both items shared the same lemma were excluded, as well as particularly unusual word forms. Lexical frequency for the members of a minimal pair was controlled so that the relative lexical frequency for the most frequent item did not exceed more than two times that of the less frequent item. Five semantic associates were selected for each lexical item to serve as primes in some experimental conditions. The primes were submitted to an online word association task in which 20 monolingual native Chilean Spanish speakers quantified the strength of the association between the target (e.g., *dudo*, “to doubt”) and a given prime (e.g., *titubear*, “to hesitate”). The primes with the highest semantic association index

were selected for each target word, but care was taken to ensure that the associates of both words (e.g., of *dudo* and *dúo*) had a similarly strong association.

Several instances of all selected target words and their associates were recorded by a monolingual native Chilean Spanish speaker, in a sound-isolated booth, using a Rode NT1A condenser microphone along with an RME Fireface UC interface connected to a PC. Recordings were made at a frequency of 44100 Hz and 16 bit depth, and were filtered with a Hann band-stop filter from 0 to 60 Hz. All words were then excised manually in *Praat* (Boersma & Weenink, 2015). For both the full form (e.g., *dudo*) and the elided targets (e.g., *dúo*), the segments of interest were segmented manually into TextGrids using visual cues from waveforms and spectrograms, and following auditory inspection of the signals (see Figure 2.1).

Figure 2.1. Left-hand side panel: instance of *dudo* [ˈdu.ð̞o], a full form target; the intervocalic approximant consonant [ð̞] is visible in both the waveform and spectrogram. Right-hand side panel: an instance of *dúo* [ˈdu.o], the elided target.



An acoustic model was built for the approximant consonants and their neighbouring segments (e.g., for [ˈu.ð̞o], from *dudo*), as well as for the corresponding elided counterparts (e.g., [ˈu.o], from *dúo*). To build these models, the time domain was divided into 100 equally distanced samples where f_0, intensity, oral formants from F1 to F3, and bandwidths for F1 to F3 were queried from acoustic objects in *Praat*. KlattGrid objects were then created for each section and populated with the acoustic parameters to match the acoustic models. Eight equally-distanced intermediate steps were created between the full approximant and elided targets. The resulting 10 steps were synthesized into sounds with a sampling frequency of 44100 Hz, using Klatt synthesis (Klatt & Klatt, 1990; Weenink, 2009). Examples of the

endpoints for a continuum can be seen in Figure 2.2. Two conditions –word and primed word– required the resulting synthetic sections to be spliced into a broader phonetic context. For these conditions, each synthetic VCV or VV section was spliced back with overlap to the remaining unaltered section taken from the original full approximant target (see Figure 2.3 for an example). Afterwards, all stimuli including semantic associates were subjected to a Hann band-pass filter from 0 to 5000 Hz in order to match the maximum frequency and overall quality of the synthetic sections with the rest of the natural stimuli. Also, mean intensity was homogenized to 70 dB SPL.

Figure 2.2. Waveform and spectrogram for the synthesized endpoints of the [ˈu.ð̞o] to [ˈu.o] continuum. The approximant consonant is visible between the vowels in the left-hand side panel, and is fully lenited in the right-hand side panel.

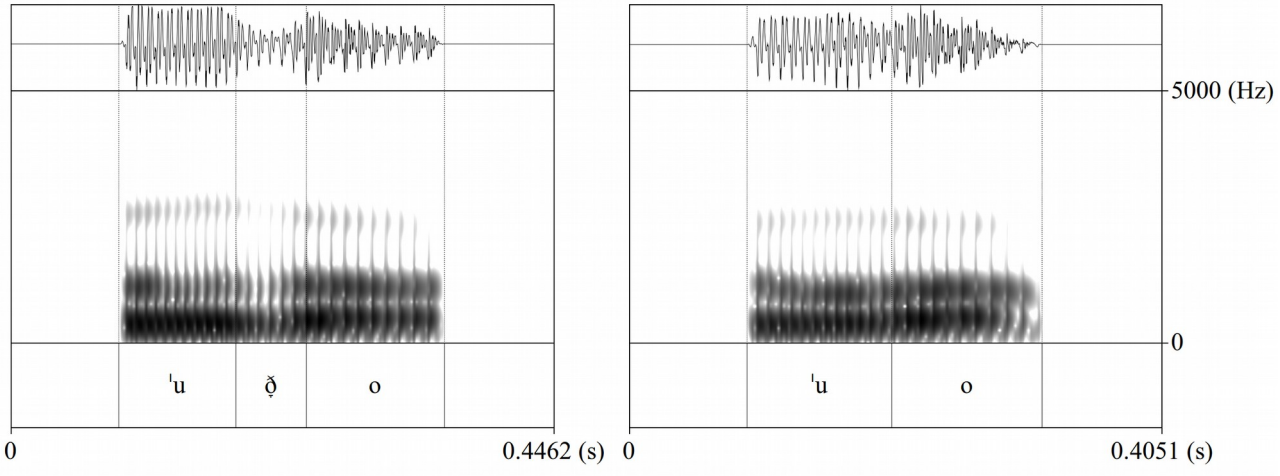
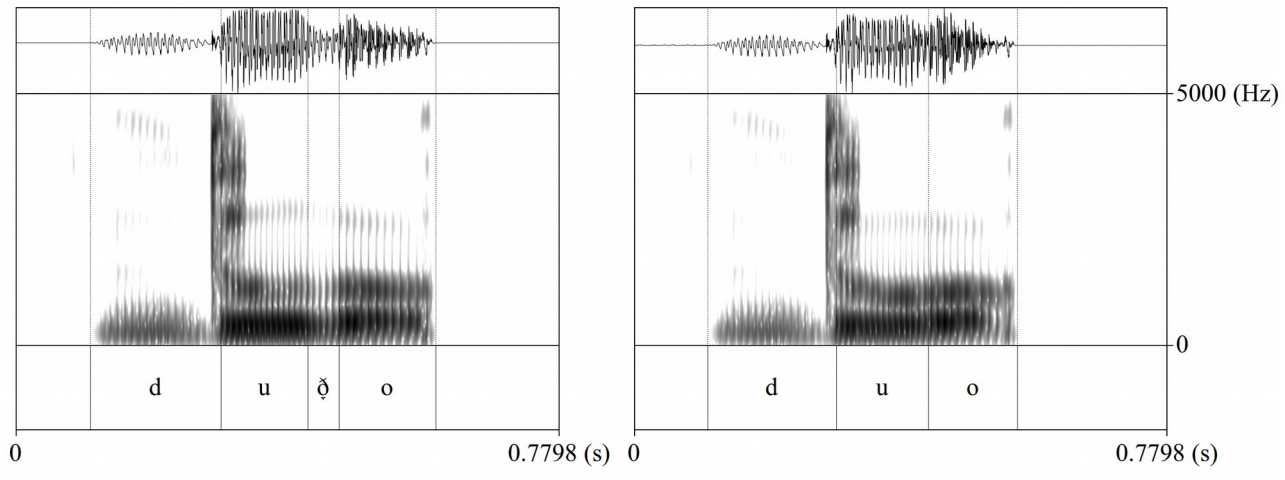


Figure 2.3. Waveforms and spectrograms for the synthesized endpoints for the [ˈdu.ð̞o] to [ˈdu.o] continuum; the approximant consonant is visible in the waveform and spectrogram in the left-hand side panel, while in the right-hand side panel the consonant is fully elided. In both cases, the stimuli have been cross-spliced to a word-initial [d], from the full form *dudo*.



Two continua were selected for /d/ and two /g/, one for the tasks and another for the practice sessions. The continuum for the practice sessions for /d/ was created from the words *callado* (“silent”) [ka.ˈja̯.ð̞o] and *Callao* (“Callao”, the Peruvian port) [ka.ˈja̯.o]. The semantic prime for the full form *callado* was *enmudecer* (“to silence”), and for the elided endpoint *puerto* (“port”). The continuum for the tasks for /d/ was created from the words *dudo* (“to doubt”) [ˈdu.ð̞o] to the word *dúo* (“duet”) [ˈdu.o]. The prime for the full form *dudo* was *titubear* (“to hesitate”) and the semantic prime for the elided endpoint was *pareja* (“couple”). The continuum for the practice sessions for /g/ was created from the words *mega* (“mega”) [ˈme.ɣ̞a] and *mea* (“to urinate”, informal) [ˈme.a]. The semantic prime for the full form *mega* was *grande* (“big” or “large”), and for the elided endpoint *orinar* (“to urinate”, formal). The continuum for the tasks for /g/ was created from the words *boga* (“fashionable”, “trendy”) [ˈbo.ɣ̞a] to the word *boa* (“boa constrictor”) [ˈbo.a]. The prime for the full form *boga* was *actualidad* (“presently” or “current”) and the semantic prime for the elided endpoint was *constrictor* (“constrictor”).

2.2.3. General procedures

Data collection was conducted by graduate students, formally trained in experimental phonetics, and in the purpose of the project, tasks and procedures. Participants were seated in quiet rooms in front of computers, in carefully selected locations in Santiago, Concepción and Temuco. Stimuli were presented via an experiment interface in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and participants listened over Sennheiser HD201 headphones. The responses were entered using the computer mouse and registered in databases. Each participant completed three tasks for each consonant: the phoneme monitoring task, an identification task and a discrimination task, each one with four conditions (see below). The order of the tasks was counterbalanced between sessions across participants.

2.2.4. Procedures for phoneme monitoring

Condition: Segmental

Listeners were presented with VCV to VV continua (e.g., from [ˈu.ð̞o] to [ˈu.o]), for /d/ and /g/. They were told that they would hear sound sequences and that on each trial they had to decide whether a given consonant, which they saw written on the screen, was present or not. Two buttons labelled "Sí" (yes) and "No" (no) were made available for each trial. Listeners completed a practice session, including examples from the full length of the practice continuum. For the task, participants were presented with each 10 step continuum twice, with stimuli being presented in a randomized order, amounting to 40 trials in total (10 steps * 2 repetitions * 2 consonants). The order of the consonant blocks was counterbalanced across participants.

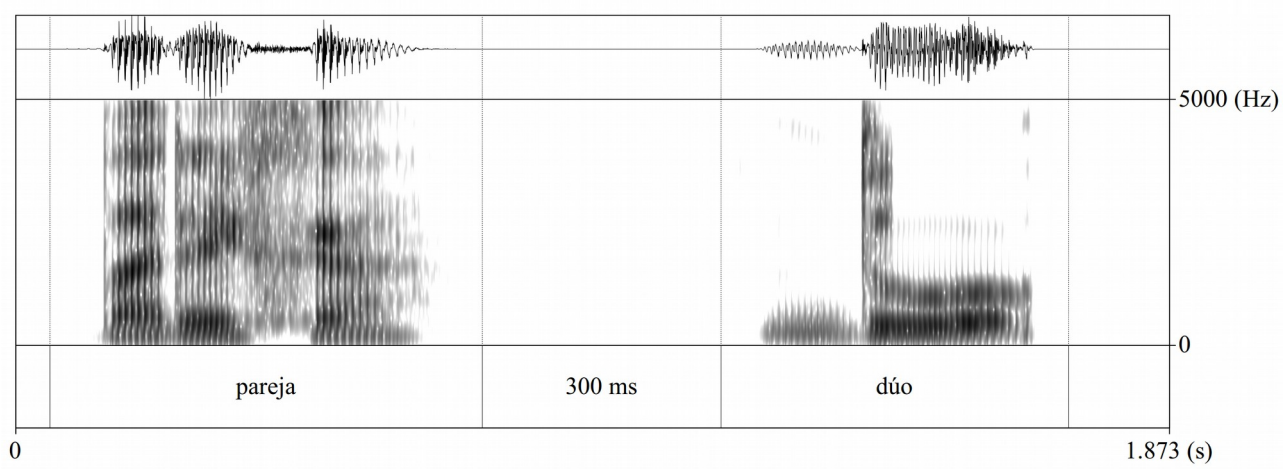
Condition: Word-level

For the most part, this condition was identical to the segmental condition. It differed in that, instead of VCV to VV sequences, listeners were presented with word-level continua (e.g., from *dudo* [ˈdu.ð̞o] to *dúo* [ˈdu.o]). Participants completed 40 trials (10 steps * 2 repetitions * 2 consonants).

Conditions: Primed approximant and Primed elision

These conditions were similar to the word-level condition, but they differed in that semantic primes were presented 300 ms before each word (see Figure 2.4), half of the time in favour of the full approximant interpretation and half of the time in favour of the elided interpretation. Listeners were told that they were going to hear two words in a sequence for each trial and that they had to monitor for a consonant in the second. A practice session was completed for each consonant block, including both primes. The task consisted of a 10 step continuum for each consonant, presented two times for each prime type (full approximant priming and priming for elided variants) in a randomized order, amounting to 80 trials in total (10 steps * 2 repetitions * 2 primes * 2 consonants). Consonant blocks were counterbalanced across participants.

Figure 2.4. Semantic associate (prime) *pareja* ("couple") and target elided endpoint *dúo* ("duet") for phoneme monitoring, primed word condition.



2.2.5. Procedures for identification

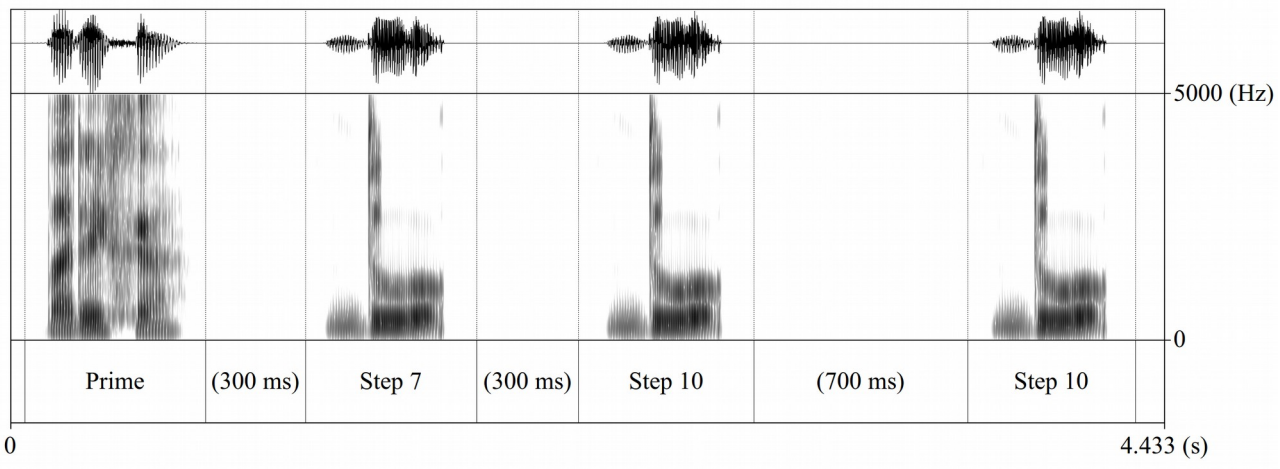
Sub-tasks, stimuli, general presentation conditions and most instructions were the same as for the phoneme monitoring task, however, listeners were told that they would hear VCV or VV sequences, and that they had to identify the sound (not monitor for it). They gave their responses by clicking on two buttons containing an orthographic transcription for the two endpoints of the continuum (e.g., "ega", for [ˈe.ɣa], and "*ea*" for [ˈe.a]). In the case of the primed word condition, listeners were explained in writing that they were going to hear two words in a sequence and that they had to identify the second word from two options transcribed orthographically on two buttons.

2.2.6. Procedures for discrimination

All discrimination tasks were designed under the ABX paradigm (Liberman, Harris, Hoffman, & Griffith, 1957; Creelman & Macmillan, 1979), which can be understood as consisting of a 2IFC subtask, in which the listener determines the order of the first two elements, and a yes-no subtask for the third item based on the decision made for the 2IFC subtask (Macmillan, Kaplan, & Creelman, 1977). One important advantage of the ABX design over alternatives such as the AX paradigm is that, listeners know that two of the three stimuli that they are perceiving are identical, and that one of the two possible answers (either "A" or "B") is indeed correct. This helps to overcome the possible bias towards answering "same", which can be found in AX designs. The stimuli used for the ABX task were adapted from those used for phoneme monitoring and identification. Each 10 step continuum was transformed into 7 discrimination pairs with 2 step interval distances. The resulting pairs were 1-4, 2-5, 3-6, 4-7, 5-8, 6-9 and 7-10. A 300 ms silence was inserted between the first two items, and a silence 700 ms long was inserted between the last two.

In the segmental condition, listeners were instructed that they would hear 3-item-long sound sequences, interspaced with pauses, and that they had to determine whether the third element matched the first or the second. Participants were also instructed that the first item would always be different from the second, despite both being potentially very similar, and that the third one would always match either the first or the second. After each trial, participants saw two buttons to enter their responses: "Primero (A)" (first) and "Segundo (B)" (second). Again, consonant blocks were counterbalanced across participants. A brief practice session was completed for each consonant block. For the main task, participants were presented with 7 discrimination trials from the task continuum, also in a randomized order, including all permutations for the ABX structure (ABA, BAA, ABB and BAB); this was done to control for bias towards the second element of the sequence, and to prevent primacy and recency effects (Greene, 1986), as well as fatigue effects (Van der Linden, Frese, & Meijman, 2003). As a result, the task comprised 56 trials in total (7 pairs * 4 permutations * 2 consonants). In the case of the word-level condition, for the most part, it was identical to the segmental condition, with the only difference that word sequences were presented to participants and instructions were adjusted accordingly. In the case of the primed word condition, the semantic primes were presented 300 ms before each ABX sequence (see Figure 2.5), and listeners were told to determine whether the fourth sound matched the second or third one.

Figure 2.5. ABX trial for the discrimination task, primed condition. In this example, the semantic prime *pareja* is presented 300 ms before the ABX group, with trial stimuli from the continuum from *dudo* to *dúo*. In this case, steps 7 and 10 are presented first, and then step level 10 again after 700 ms, completing an ABB trial.



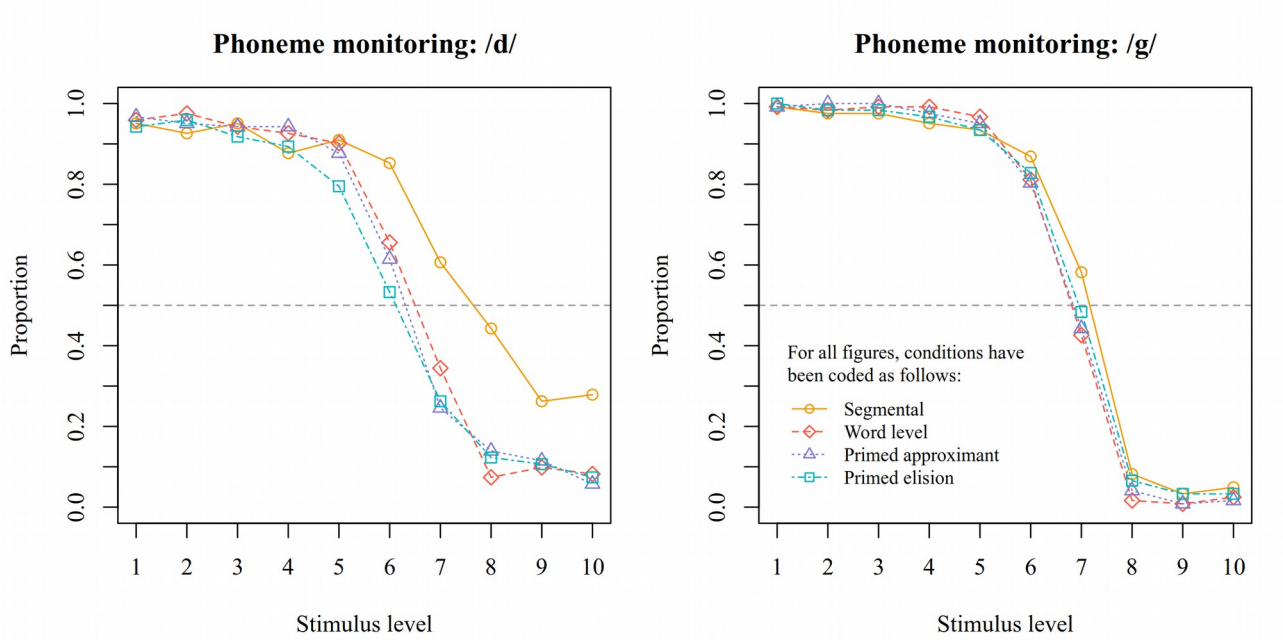
2.3. Results

2.3.1. Phoneme monitoring

Phoneme monitoring: /d/

The results for the segmental condition resembled a cumulative binomial distribution (see left-hand panel from Figure 2.6). Perception of [ð̩] started with values around 95% and remained close to this level for the first 6 stimuli, at which point perception decreased and crossed the 50% chance level between stimulus step 7 and 8. The lowest levels for the segmental condition were 30% of perception for the last two steps. In general, perception of [ð̩] was high and did not reach floor despite the acoustic evidence being very scarce or even absent in the last steps of the continuum.

Figure 2.6. Phoneme monitoring results for /d/ and /g/, shown as averaged responses across participants (for each consonant, n = 4880). Proportion of reported presence of each consonant is shown as a function of stimulus level; chance level is shown as a dashed horizontal line.



The results for the word-level condition also displayed a cumulative binomial distribution, but when compared to the segmental condition, perception of [ð̩] in the word-level condition was lower for the second half of the continuum. Overall, responses from the primed approximant condition were very similar to those from the word-level condition, with the exception of step number 7, where the

responses for the primed approximant condition were lower, and step 8, where the inverse pattern was observed. In this condition, it was expected that the category boundary would shift towards perception of [ð̞] when compared to condition word-level, but this was not completely substantiated by the observed responses. In the case of the primed elision condition, responses followed a similar pattern as word-level and primed approximant conditions, but perception was lower than these two conditions in the group of stimuli immediately around the category boundary. In this case, the prediction of lower perception of [ð̞] with respect to the word-level condition was met.

A GLMM analysis was conducted on the results of phoneme monitoring for /d/. The best-fitting model for this analysis included *response* as the dependent variable, *experimental condition* and *stimulus level* as fixed factors, their interaction, *subject* as a random factor, and *stimulus level* as a random slope. All subsequent GLMM models were built in the same way. A significant main effect of *condition* was found ($\chi^2(3) = 24.626$, $p < 0.001$), along with a significant main effect of *stimulus level* ($\chi^2(1) = 106.361$, $p < 0.001$) and a significant interaction between *condition* and *stimulus level* ($\chi^2(3) = 72.581$, $p < 0.001$). Wald *z* statistics exploring the differences in *response* variable for the interaction between *condition* and *stimulus level* are provided in Table 2.1. As is customary, only the details pertaining the interaction were included. Regarding the interaction, the analyses showed that the relationship between *stimulus level* and *response* were significantly different only when the segmental condition was compared to all other conditions.

Table 2.1. Wald *z* statistics for differences in *response* in the phoneme monitoring task for /d/ for the interaction between *stimulus level* and *condition* (SE = standard error).

Baseline	**Comparison**	**Estimate**	**SE**	***z***	***p***	
Segmental	Word-level	-0.438	0.063	-6.995	< 0.001	***
Segmental	Primed app.	-0.419	0.062	-6.806	< 0.001	***
Segmental	Primed elision	-0.313	0.057	-5.453	< 0.001	***
Word-level	Primed app.	0.018	0.070	0.264	= 2.376	
Word-level	Primed elision	0.124	0.066	1.872	= 0.183	
Primed app.	Primed elision	0.106	0.066	1.617	= 0.318	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

Phoneme monitoring: /g/

The results for the segmental condition for /g/ showed a cumulative binomial distribution with values that centred on around 95% for the first 5 steps and then descended abruptly to cross the 50% chance level between stimuli 7 and 8. Perception stabilized close to 10% perception for the last three steps (see right-hand panel from Figure 2.6). When compared to all other conditions, the segmental condition displayed slightly higher values of [ɣ̟] perception for the steps around the category boundary shift. The distributions of the other experimental conditions are very similar to that of the segmental condition. Regarding their differences, in the case of the word-level condition, the first and last sections of the continuum reached values slightly closer to ceiling and floor perception, and the higher levels of [ɣ̟] perception were sustained longer than for segmental condition. The results of both primed conditions did not differ in any noticeable way from word-level condition, perhaps with the exception of the primed elision condition in step 7, which presents values more similar to those found in the segmental condition. No clear semantic priming effect was observed.

A GLMM analysis with *response* as the dependent variable failed to reach significance for *condition* ($\chi^2(3) = 1.1319$, $p = 0.77$), but did show a significant main effect of *stimulus level* ($\chi^2(1) = 160.1113$, $p < 0.001$), as well as a significant interaction between *condition* and *stimulus level* ($\chi^2(3) = 160.1113$, $p < 0.001$). Wald *z* statistics for the interaction (see Table 2.2) show that the relationship between *stimulus level* and *response* was significantly different when the segmental conditions were compared to the word-level and primed approximant conditions, but not between the segmental and primed elision condition, probably due to the latter being slightly more similar than the former than other conditions.

Table 2.2. Wald *z* statistics for differences in *response* in the phoneme monitoring task for /g/ for the interaction between *stimulus level* and *condition* (SE = standard error).

Baseline	**Comparison**	**Estimate**	**SE**	***z***	***p***	
Segmental	Word-level	-1.201	0.359	-3.348	< 0.01	**
Segmental	Primed app.	-1.274	0.365	-3.487	< 0.001	***
Segmental	Primed elision	-0.440	0.302	-1.458	= 0.435	
Word-level	Primed app.	-0.074	0.429	-0.171	= 2.592	
Word-level	Primed elision	0.761	0.377	2.015	= 0.132	
Primed app.	Primed elision	0.834	0.384	2.172	< 0.1	.

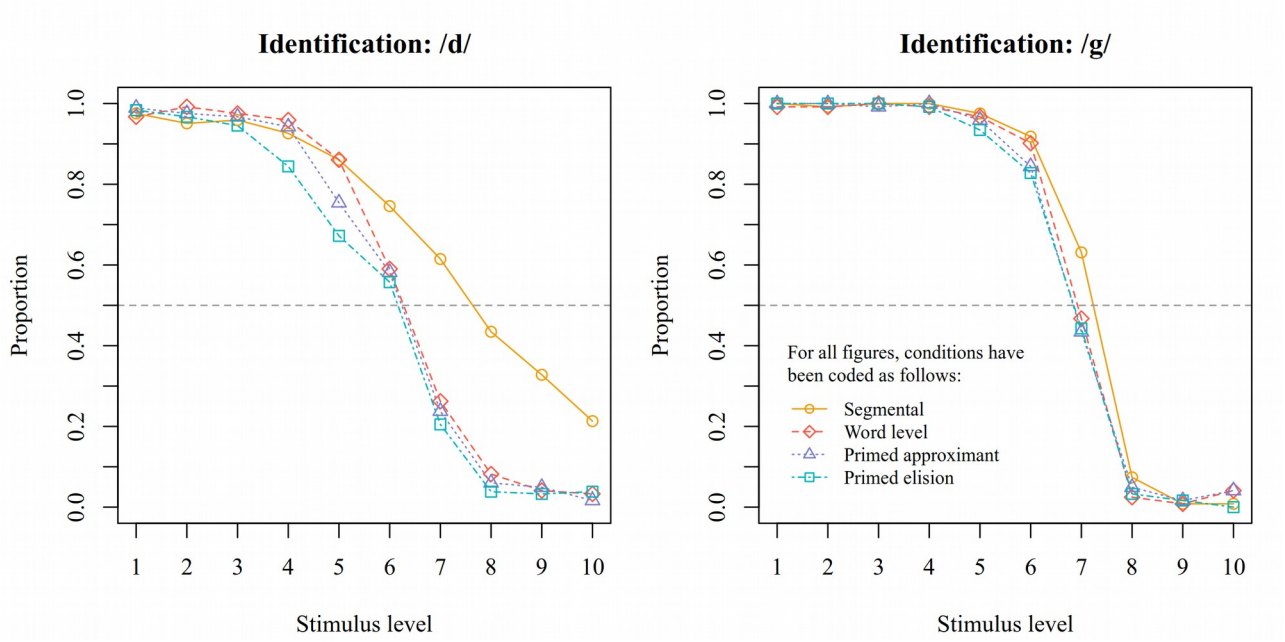
Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

2.3.2. Identification

Identification: /d/

The results for the segmental condition show that the first 4 steps reached around 95% of [ð̞] identification (see left-hand panel from Figure 2.7). Values then decreased gradually and the category boundary crossing of the 50% chance level occurred close to step 8. The last two steps still displayed high values, with responses around 35% to 25% identification. Overall, [ð̞] identification was high when compared to other conditions, and the continuum did not reach a plateau close to zero. In the word-level condition, results were organized in a cumulative binomial distribution. For the first 4 steps, [ð̞] identification approached ceiling, and then identification decreased gradually until it crossed the 50% chance level at step 6, and stabilized for the last 3 steps with values around 5% identification. From stimulus 5 onwards, identification values were lower than in the segmental condition. The primed approximant condition displayed similar results to those from the word-level condition, with the exception of stimulus 5, which displayed lower values of identification. The prediction of higher identification for [ð̞] and a delayed 50% chance level category boundary crossing was not met. Finally, the results for primed elision condition also displayed a cumulative binomial distribution. These results were, in general, similar to those from the word-level and primed approximant conditions, although the identification values are lower between stimuli 4 and 5, and slightly lower in stimuli 3, 7 and 8. The prediction of less [ð̞] identification for this conditions was only met on steps 4 and 5.

Figure 2.7. Identification results for /d/ and /g/, shown as averaged responses across participants (for each consonant, n = 4880). Proportion of identification for each consonant is shown as a function of stimulus level; chance level is shown as a dashed horizontal line.



A GLMM analysis was conducted on the results of /d/. The results showed a significant main effect of *condition* on the dependent variable ($\chi^2(3) = 43.093$, $p < 0.001$), a significant main effect of *stimulus level* ($\chi^2(1) = 148.523$, $p < 0.001$) and a significant interaction between *condition* and *stimulus level* ($\chi^2(3) = 123.816$, $p < 0.001$). Wald *z* statistics for the interaction (see Table 2.3) showed that the way in which the dependent variable *response* changed over *stimuli level* was different when the segmental condition was compared to all others, but not for any other comparison.

Table 2.3. Wald *z* statistics for differences in *response* in the identification task for /d/ for the interaction between *stimulus level* and *condition* (SE = standard error).

Baseline	**Comparison**	**Estimate**	**SE**	***z***	***p***	
Segmental	Word-level	-0.620	0.075	-8.239	< 0.001	***
Segmental	Primed app.	-0.610	0.068	-8.907	< 0.001	***
Segmental	Primed elision	-0.474	0.062	-7.581	< 0.001	***
Word-level	Primed app.	0.011	0.085	0.126	= 2.700	
Word-level	Primed elision	0.147	0.080	1.828	= 0.204	
Primed app.	Primed elision	0.136	0.074	1.844	= 0.195	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

Identification: /g/

For the segmental condition there was a cumulative binomial distribution curve, with values close to ceiling identification for the first 5 steps, and then an abrupt decline with a cross of the 50% chance level between steps 7 and 8, that stabilized with values of 10% to floor identification for the last three steps (see right-hand panel from Figure 2.7). The remaining three conditions (word-level, primed approximant and primed elision) showed very similar results. The only differences were that, firstly, these three conditions crossed the 50% chance level slightly earlier than the segmental condition, and, secondly, that in both priming conditions [ɣ̟] identification decreased sooner than in segmental and word-level conditions (in steps 5 and 6). A GLMM analysis showed a significant main effect of *condition* on the identification results for /g/ ($\chi^2(3) = 18.642$, $p < 0.001$), and a significant main effect of *stimulus level* ($\chi^2(1) = 183.155$, $p < 0.001$), but no significant interaction, which means that responses changed between the levels of the continuum, and that these changes were different between some conditions, but that the shape of the distributions is not different. The results of Wald *z* statistics, provided in Table 2.4, showed that the segmental condition was significantly different from all other conditions, but not in other comparisons.

Table 2.4. Wald *z* statistics for differences in *response* in the identification task for /g/ between different levels of the variable *condition* (SE = standard error).

Baseline condition	**Comparison**	**Estimate**	**SE**	***z***	***p***	
Segmental	Word-level	-0.495	0.192	-2.571	< 0.05	*
Segmental	Primed approximant	-0.604	0.193	-3.134	< 0.01	**
Segmental	Primed elision	-0.805	0.194	-4.152	< 0.001	***
Word-level	Primed approximant	-0.110	0.191	-0.573	= 1.701	
Word-level	Primed elision	-0.310	0.191	-1.620	= 0.315	
Primed approximant	Primed elision	-0.201	0.191	-1.050	= 0.882	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

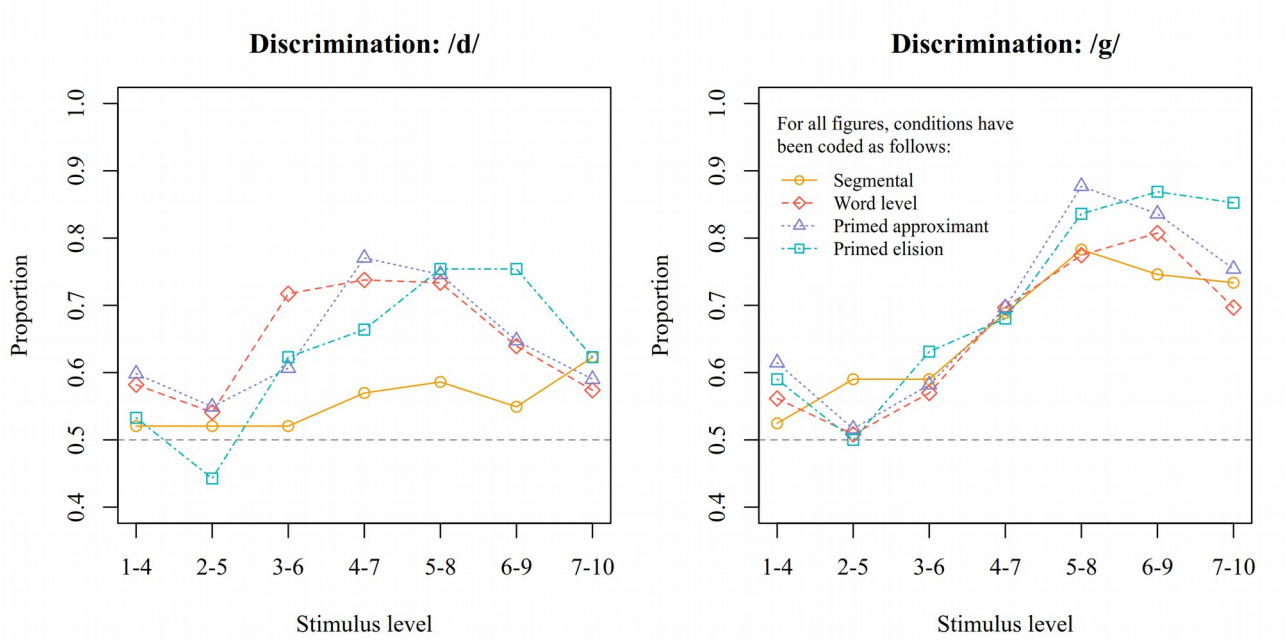
2.3.3. Discrimination

Discrimination: /d/

The results from the segmental condition showed low discrimination sensitivity across the stimulus pairs, with values virtually at chance level for the first 3 pairs and then increasing to values around 60%

discrimination, with no evident discrimination sensitivity peak (see left-hand panel from Figure 2.8). In the word-level condition, discrimination began with values close to chance level for the first two pairs and then they increased to around 75% discrimination for pairs 3 to 5. In the last two pairs, discrimination decreased to values around 60%. The shape of this distribution might be indicative of a category boundary around the fourth pair. The results for the primed approximant condition resembled those seen for the word-level condition, with the first pairs near chance discrimination, then an increase to values closer to 75%, and a fall towards 60% discrimination in the last two steps. The primed approximant condition showed a delay in the increase of discrimination sensitivity with respect to word-level condition (see step 3-6), and then surpassed the maximum level of discrimination seen in this latter condition. The prediction of a bias in favour of the full interpretation of the stimuli with respect to the word-level condition was only partially backed up by the results. In the case of the primed elision condition, the first stimuli pair showed values close to chance level and then discrimination fell below this threshold for the second step. Afterwards, discrimination sensitivity rose gradually until reaching a discrimination peak in stimuli pairs 5-8 and 6-9, with values close to 75%. Finally, discrimination descended to a value closer to 60% in the last pair. The prediction of a bias towards the elided end of the continuum had some support from the results, with the discrimination peak shifted towards the right with respect to the word-level condition.

Figure 2.8. Discrimination results for variants of /d/ and /g/, shown as proportion of discrimination averaged across participants as a function of stimulus level pairs (for each consonant, n = 5124). Chance level is shown as a dashed horizontal line.



A GLMM analysis was conducted on the results of discrimination from /d/. The results showed a significant main effect of *stimulus level* in *response* ($\chi^2(1) = 5.1921$, $p < 0.05$), a significant main effect of *condition* ($\chi^2(3) = 16.1107$, $p < 0.001$) and a significant interaction between *stimulus level* and *condition* ($\chi^2(3) = 9.2174$, $p < 0.05$). Wald *z* statistics for the interaction (see Table 2.5) showed that only the comparison between the word-level and primed elision conditions were statistically significant, most likely because of the sharp descent of the word-level condition results seen in pair 2-5, and a delay of the primed elision condition to reach lower sensitivity values in pair 6-9.

Table 2.5. Wald *z* statistics for differences in *response* in the discrimination task for /d/ for the interaction between *stimulus level* and *condition* (SE = standard error).

Baseline	**Comparison**	**Estimate**	**SE**	***z***	***p***	
Segmental	Word-level	-0.031	0.036	-0.867	= 1.158	
Segmental	Primed app.	-0.011	0.044	-0.257	= 2.391	
Segmental	Primed elision	0.102	0.044	2.305	< 0.1	.
Word-level	Primed app.	0.020	0.045	0.443	= 1.974	
Word-level	Primed elision	0.134	0.045	2.973	< 0.01	**
Primed app.	Primed elision	0.114	0.052	2.197	< 0.1	.

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

Discrimination: /g/

The results for the segmental condition started at chance level for the first pair and increased gradually until reaching a discrimination sensitivity peak in pair 5-8 with around 80% discrimination (see right-hand panel from Figure 2.8). Discrimination then decreased in the last two steps to values close to 75%. The results for the word-level condition showed discrimination starting slightly above chance level and then decreasing to chance level at the second pair. From pair 2-5 onwards, discrimination increased gradually until reaching a discrimination sensitivity peak around 80% in pair 6-9, after which it descended to around 70% discrimination. For the most part, responses showed a similar distribution than for the segmental condition. For the primed approximant condition, discrimination began around 60% and then decreased to chance level for the second pair. From the third stimulus pair onwards, discrimination increased until it reached a maximum value in pair 5-8, approaching 90%, after which it decreased to levels closer to 75% discrimination. In line with predictions, the discrimination peak preceded that in the word-level condition. The results for the primed elision condition showed that the first stimulus pair had a discrimination value close to 60%. Discrimination decreased to chance level in the second pair and then it increased gradually until reaching a maximum around 85% for the last three steps. This maximum of sensitivity for the last pairs matches predictions of a category boundary shift in favour of an elided interpretation of the continuum.

A GLMM analysis showed a significant main effect of *stimulus level* on *response* ($\chi^2(1) = 46.1917$, $p < 0.001$) and a significant interaction between *stimulus level* and *condition* ($\chi^2(3) = 8.8181$, $p < 0.05$), but not a significant main effect of *condition* ($\chi^2(3) = 2.0011$, $p = 0.57218$). Wald *z* statistics for the interaction (see Table 2.6) showed significant differences between the condition primed elision and the conditions segmental and primed elision, probably due to differing trajectories for the first and last two pairs of the discrimination continuum. The fact that the primed elision condition did not show a noticeable decrease in the last stimuli pairs is particularly important, as this also helps to explain the significant interaction when this condition is compared to the word-level condition.

Table 2.6. Wald *z* statistics for differences in *response* in the discrimination task for /g/ for the interaction between *stimulus level* and *condition* (SE = standard error).

Baseline	**Comparison**	**Estimate**	**SE**	***z***	***p***
Segmental	Word-level	0.012	0.038	0.311	= 2.268
Segmental	Primed app.	0.055	0.048	1.135	= 0.768

Segmental	Primed elision	0.139	0.050	2.791	< 0.05	*
Word-level	Primed app.	0.043	0.048	0.888	= 1.122	
Word-level	Primed elision	0.127	0.050	2.553	< 0.05	*
Primed app.	Primed elision	0.084	0.058	1.462	= 0.420	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

2.4. Discussion

2.4.1. Interpreting the results

The experiments controlled a very specific set of cues in order to explore what is required for listeners to perceive approximant variants of the consonants /d/ and /g/. The condition with the smallest amount of available cues was the segmental condition, in which only the approximant consonant and surrounding segments were present. In this condition only the acoustic variables directly cueing for the approximant or its absence were available, along a minimal phonetic context which contained coarticulatory cues. This condition, particularly so in phoneme monitoring, provided an auditory baseline of perception, since no semantic or syntactic information was available to listeners to cue for the presence of approximants.

Beginning with the results of the phoneme monitoring experiment for /g/ (see right-hand panel from Figure 2.6), listeners displayed responses that suggest that they take the acoustic evidence provided to them at face value, and that a continuum from presence to absence is perceived categorically (Liberman, Harris, Hoffman, & Griffith, 1957; Harnad, 1987). The hypothesis that the perception of /g/ is driven predominantly by acoustic cues gains support when results from identification and discrimination tasks in the segmental condition are also considered. In the case of identification, basically the same pattern of categorical perception was observed. In the case of discrimination, there was a coincidence between the location of a discrimination sensitivity peak for the segmental condition and the stimuli in which the chance level crossing occurred in phoneme monitoring and identification tasks, which is also consistent with a categorical perception account (see right-hand panel from Figure 2.8).

Results were quite different for /d/: it is still true that acoustic cues had a strong effect on perception, because the more acoustic cues were available for an approximant consonant, the more listeners reported perceiving it. However, [ð̞] was always perceived to some extent, even when no acoustic cues for it were available in the signal. Given that no lexical information was provided to listeners in this

condition, no lexical effect can explain the persistence of perception despite absence of acoustic information. Two alternative explanations can be provided instead. First, it may be the case that listeners, knowing that evidence for /d/ is scarce in natural perception, are particularly sensitive to small acoustic cues for [ð̞], and thus require less evidence for perception by the end of the continuum. Second, it may be the case that listeners are not particularly sensitive to small acoustic cues for /d/, but that, instead, knowing that evidence is scarce or unreliable, over-compensate for it after initial stages of speech processing, even when acoustic cues are completely absent from the signal (Mitterer & Ernestus, 2006; Janse, Nooteboom & Quené, 2007). When the results from discrimination are considered, the second explanation seems more likely, that is, an explanation which assumes phonological recovery, because no evidence of particularly high sensitivity to small differences was observed in the results of discrimination for the segmental condition in [ð̞] (see left-hand panel from Figure 2.8).

Overall, providing a wider phonetic context and semantic cues in the word-level condition had the effect of bringing participants' perception of [ð̞] closer to distributions consistent with a categorical perception account, in both the phoneme monitoring and identification tasks (see left-hand panels from Figure 2.6 and Figure 2.7). This can be explained as the result of two lexical representations, of similar lexical frequency, competing for perception in more or less equal terms, which results in two cancelling lexical effects on perception and a categorical treatment of the continuum. This would also explain why the results from the identification task are closer to a categorical perception distribution than in phoneme monitoring, given that a lexical effect should be stronger in a task where listeners are forced to process two lexical representations prior to providing their responses, as opposed to one in which listeners can ignore lexical representations.

Generally speaking, providing further semantic priming in two conditions in the phoneme monitoring and identification tasks provided very little evidence of a priming effect in the hypothesized directions, for both consonants. Weak effects of semantic priming were only detected for /d/, a consonant with particularly unreliable acoustic evidence in natural perception, which perhaps makes it more susceptible to priming. In any case, it seems to be the case that there is a limit to the effects that various cues have in speech perception. Some cues, such as adding a word-level context, have the effect of dramatically changing the way in which listeners perceive a continuum from consonant presence to absence, and the same can be expected of adding sentence level syntactic and semantic cues. Semantic priming, instead, barely had an effect. More powerful and direct priming alternatives might have shown

stronger effects of semantic priming on speech perception (e.g., Samuel, 1981; Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004).

As to the results of the discrimination tasks, in the case of /g/, sensitivity increased as a function of stimulus pair and reached a sensitivity maximum in those stimuli in which the chance level crossing occurred in phoneme monitoring and identification tasks (see right-hand panel from Figure 2.8). This is clear evidence of categorical perception (Harnad, 1987). In the case of /d/, the results displayed values close to chance level in the segmental condition. This could be interpreted as indicating that listeners were not sensitive to acoustic differences between variants of /d/ (see left-hand panel from Figure 2.8). Increasing the number of cues for the word-level condition had a dramatic effect on discrimination. In particular, it increased discrimination sensitivity considerably for those stimuli involved in the chance level crossing in phoneme monitoring and identification. Semantic priming further increased discrimination sensitivity peaks with respect to the word-level condition, and shifted each distribution in line with predictions with respect to the last peak of the word-level condition.

2.4.2. Implications for lexical access models

Episodic models of lexical access assume that numerous exemplars are stored in long-term memory, and that similar episodes aggregate into clusters. These “memory clouds” are then matched to lexical representations for lexical access. Episodic models deal with variation by assuming that exemplars for highly lenited forms are also stored in permanent memory. If these episodes are more frequent or recent, they will have a comparative advantage over alternative episodes. Under these models, it is expected that listeners have stored multiple exemplars of words containing /d/ and /g/, including episodes along the entire continua from approximants to elided variants. It is also expected that more exemplars exist for words containing the consonant's most frequent variants. For example, in the case of /g/, words containing the consonant should be better represented in exemplar clouds; in the case of /d/, absence of the consonant and very weak approximants should be better represented.

When continua from approximant to elided variants are presented in isolation (as in the segmental condition), listeners cannot match the input to lexical-sized episodes. Instead, they have to match the segmental input to segmental-sized acoustic episodes, and better matches ought to occur for those steps from continua where stimuli are better represented by said episodes. Episodic models would thus predict the results observed in the segmental condition from /g/, but also the results observed for /d/. Since most episodes for this consonant contain little acoustic evidence, not much is required to attain a match for [ð̞]; put another way, not much is required from a signal for it to be a good candidate of /d/.

In the case of word-level stimuli, listeners should find it easier to perceive inputs that agree with frequent episodes, although it is also expected that partial matchings to the input can occur, since there is more information available to make a lexical decision. These assumptions should apply more or less equally to both members of the minimal pair from each continuum, given that they have similar lexical frequencies. They predict that listeners will perceive the continuum from /g/ categorically, since the acoustic cues for [ɣ̟] tend to be well represented in the signal in natural perception (when [ɣ̟] is present, acoustic evidence is clear). Instead, the category boundary for [ð̟] should be shifted to the right with respect to [ɣ̟], since variants with less acoustic evidence are good exemplars for both words (although harder to perceive). Only the predictions for /g/ were confirmed by our results. It is harder to evaluate the results from semantic priming, since mostly null effects were found in our experiments.

Strong episodic models of lexical access (e.g., LAFS) have trouble explaining lexical effects on speech perception, since labels representing episodic clouds do not intervene in prelexical stages of speech processing. Models where some top-down influence can take place, such as Minerva 2, are better able to account for lexical effects and phonological recovery by assuming that the echo of integrated exemplars that is returned after an episodic probe has been sent to long-term memory can contain information not present in the input episode, in essence, working as a pseudo-abstract representation. This semi-abstract echo would allow for the activation of lexical items for which there is imperfect matching. In our data, evidence for phonological recovery and lexical effects was observed for /d/, and thus Minerva 2 is better suited to explain it. Crucially, however, this does not work for the segmental condition, unless the probability of a sequence conforming to a word could be evaluated against stored episodes. The influence of the two competing lexical items on perception, which made responses more categorical, was stronger in identification, a task in which evaluating the input against two underlying lexical representations was required before a response was provided. The fact that the two target lexical items were permanently activated in identification can explain a facilitatory effect in episodic models, in which the desired exemplar clouds would remain active throughout the task.

Given that in episodic models exemplars cluster naturally into groups defined by the similarity, frequency and recency of episodes, these groups can be thought to behave as prototypes, although they do not constitute abstract representations. If prototypes of some sort exist –albeit only functionally– then episodic models should have no problem accounting for categorical perception: listeners should be better able to tell apart examples from different episodic clouds from those in the same clouds. In our results, sensitivity to stimuli differences increased as the amount of acoustic evidence for the consonants decreased in all conditions of the discrimination task, but in some more clearly than in

others. In the case of /g/, sensitivity maxima tended to coincide with identification category boundaries. The results from the segmental condition for /d/, in which proportion of discrimination was low overall, could also be explained under episodic models by assuming that episodic clouds for the segmental level are not particularly strong or easy to match, since segmentation into intermediate abstract segmental categories is not an assumption of episodic models. When the acoustic evidence is provided in a word-level context, it becomes interpretable via matching to word-level sized episodes, and discrimination should improve, as was observed in the results.

Abstractionist models work under the assumption that the mental lexicon contains one abstract representation for each word, and that its structure consists of a string of abstract phonological segments. All models assume that the acoustic input has to be converted into a chain of some type of prelexical segmental-sized units that will later be compared to lexical abstract representations until an optimal match has been found and lexical access takes place. Beyond these commonalities, abstractionist models differ considerably in their posited processes and structures. For instance, most abstractionist models do not allow top-down feedback, i.e., Cohort, FLMP (Oden & Massaro, 1978; Massaro & Oden, 1980). Other models like RACE (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987) and Shortlist –both also autonomous– propose parallel and independent phonemic and lexical processing routes, whilst a minority implements top-down feedback directly (TRACE). Still others like Merge (Norris, McQueen, & Cutler, 2000) also have two parallel and independent processing routes like RACE, from prelexical processing units to lexical units, and from prelexical processing units to phoneme decision nodes, but also feedback from the lexical level to phoneme decision nodes (but not to prelexical acoustic processing nodes). Some abstractionist models such as Cohort and Shortlist make very specific predictions regarding how perception unfolds over time, allowing groups of candidates to compete depending on their degree of acoustic match to the incoming input. Lastly, different models defend alternative strategies to deal with lenited forms, ambiguous input, and to account for lexical effects and recovery. For instance, some models allow for underspecified features (Cohort), while others resort to independent lexical routes (RACE), facilitation from lexical levels to phoneme decision levels (Merge), or even to direct top-down feedback (TRACE).

Given that in our experiments all the sequences from the segmental condition were nonsense units, it is to be expected that only prelexical processing took place in their perception. Most abstractionist models of lexical access and speech perception are able to describe how listeners process segmental input by the means of prelexical processing modules that parse the acoustic input and extract the relevant features to build hypothesis of abstract phonological representations. Models like RACE or Merge

would predict categorical perception for the two consonants, since there is no clear way for an underlying phonological representation to contain degraded or partial features; consequently, they are not very good at accounting for the results of /d/ in the segmental condition. FLMP does accept partial featural and phonological matching, and thus it might be able to explain recovery for [ð̩] in this condition, because minimal evidence is still a better cue for [ð̩] than it is for its absence. Notice that inhibitory connections between competitors in feature and phonemic nodes, such as those modelled by TRACE, do not apply to the presence of an item competing with its absence, and thus they do not help explaining partial matching. All in all, abstractionist models of lexical access do not seem particularly well suited to explain the results from segmental conditions, which is not surprising considering that most of the reviewed models are models of lexical access and not of speech perception.

The results from word-level conditions can be better accounted for by abstractionist models, although not all models are well prepared to predict the results from tasks in which lexical processing is not mandatory, as in phoneme monitoring. In Cohort, given that lexical access is achieved by the activation of all possible lexical candidates for an input, listeners cannot ignore lexical levels of processing in phoneme monitoring and provide a purely auditory response. Consequently, Cohort would predict identical results for the phoneme monitoring and identification tasks, a prediction which was not supported by the results described here. Cohort also predicts that semantic priming facilitates activation of the primed candidate, even in the event of ambiguous input. The model would thus predict semantic priming effects for /d/ and /g/, but they should be stronger for /d/, which in turn was the only consonant in which semantic priming effects were clear.

In Shortlist, just as in Cohort, listeners cannot ignore lexical levels of processing, given that groups of lexical candidates are activated as soon as the acoustic input is received. In the case of identification, the model predicts that, as long as perception follows a prelexical route, the acoustic input will activate a group of candidates compatible with this input, and that the competing lexical items will remain as good candidates until the acoustic input disambiguates in favour of one or the other by means of inhibitory links. Given that Shortlist is a RACE model, an independent lexical route should also be able to provide an output and thus account for lexical effects. In any case, Shortlist should predict categorical perception in identification.

Assuming that listeners parse the input of phoneme monitoring tasks by resorting primarily to prelexical processing, Merge would predict that prelexical nodes will provide most of the information that phoneme decision nodes receive (although feedback from lexical processing nodes to phoneme

decision nodes cannot be ruled out), and that categorical response curves should be observed for a continuum from full approximant to elided variants. However, categorical perception would not be maximized for phoneme monitoring because ambiguous input in this model is not resolved at early stages of speech processing. In identification, lexical decision nodes receive input from the prelexical processing nodes, and lexical nodes provide feedback to the phoneme decision nodes, disambiguating the input and producing responses closer to categorical perception, as observed in the data.

In the case of RACE, the model would predict that listeners will resort primarily to the prelexical processing route to provide a response in phoneme monitoring, but that sometimes, ambiguous input might allow the lexical processing route to achieve certainty thresholds first. Given that the two lexical items in competition are comparable in frequency and identical with the exception of the approximant consonant, it is still more likely for the prelexical route to win in phoneme monitoring, which would result in categorical perception distributions of responses. In identification, in which the two competitors are primed via the response categories, and given that sections of the continua are ambiguous, the lexical route should win more often, and provide results closer to categorical perception.

Finally, in the case of TRACE, in which information can flow in any direction between feature, phonemic and lexical nodes, the model would predict that in phoneme monitoring, in which responses can be purely prelexical, the input is processed by featural nodes and the presence or absence of the consonant could be resolved at the phonemic level. However, sometimes listeners may choose to adopt a post-lexical strategy in phoneme monitoring. In both cases, the model would predict categorical perception results in phoneme monitoring, but they should be clearer in identification where there is a direct lexical effect in speech perception, as was seen in the results.

Hybrid models propose that both episodes and abstract representations exist and interact in speech perception and lexical access. In some models like Goldinger's CLS and Pierrehumbert's ED model, abstract representations always play a role in lexical access, while in other models like POLYSP, the retrieval of abstract linguistic units is an optional by-product of lexical access. Both Goldinger's CLS and POLYSP can be considered connectionist models, given that top-down feedback is possible, and consequently are particularly well suited to account for lexical effects on speech perception and phonological recovery. In contrast, Pierrehumbert's ED model is autonomous, given that the flow of information only goes from lower to higher processing levels. All these models have the advantage that

they include exemplars and episodic clouds, which can explain evidence that listeners utilize fine-grained phonetic detail during lexical access and learning, alongside abstract representations which account for evidence that listeners are able to perform speaker-normalization and that they perceive some contrasts categorically.

Hybrid models, in general, have no trouble explaining the results obtained in our perception experiments. To begin with, the results from the segmental conditions in phoneme monitoring can be explained as for episodic models (as a result of the nature and structure of the episodic clouds for each consonant, which include expectations regarding what is expectable in natural perception). Adding semantic cues in the word-level condition should provide better matches to lexical-sized episodes, and also allow for top-down feedback to inform speech perception if listeners choose to use a lexical route for their responses (in CLS and POLYSP, at least). This explains responses describing distributions closer to categorical perception in the word-level condition. Given that lexical processing is mandatory in identification, lexical effects and phonological recovery should be stronger than in phoneme monitoring for both consonants. Semantic priming should have the effect of increasing the relative advantage of the primed candidate by pre-activating the relevant episodic cloud, particularly when the acoustic evidence is unreliable, as was observed as a trend for /d/. The results from discrimination can also be explained by hybrid models. Firstly, increasing the amount of acoustic and semantic cues should facilitate discrimination, partly, because lexical-sized episodes can be compared. Consonants with relatively poor acoustic evidence in natural perception showed low discrimination values in the segmental level, which is expected for episodic clouds with relatively poor acoustic detail, and which should therefore be more difficult to discriminate.

Summing up, the fact that results differed for /d/ and /g/, with perception failing to reach floor in /d/ in the segmental conditions, even when acoustic evidence was absent, makes it unlikely that mapping the acoustic input to underlying phonological units is driving the perceptual process, even if underspecified features are possible, as in Cohort (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987; Lahiri & Marslen-Wilson, 1991), or if gradual featural matching is possible, as in the Fuzzy Logical Model of Perception (Oden & Massaro, 1978; Massaro & Oden, 1980). This is because in abstractionist models of lexical access underlying units are indifferent to the type of variation resulting from processes such as lenition, because that information is lost at early normalization stages (Ernestus, 2014). Moreover, normalization ought to be particularly strong in tasks in which no other information is able to disambiguate the input, such as in phoneme monitoring in the segmental condition.

These observations seem to constitute evidence in favour of an episodic memory account, at least for the data presented here, as these models propose a mechanism for the storage of fine phonetic detail. There ought to be a bigger effect of episodic memory in an auditory task, such as phoneme monitoring, in which the acoustic information itself is scrutinized. Conversely, in tasks in which top-down feedback becomes available, the relative strength of episodes should decrease, as in the identification task. Indeed, when additional cues become available, and the listener can tap into top-down information, postlexical feedback seems to override the effects of episodic traces seen in the segmental task. Lexical access models that display properties compatible with these requirements –both the storage of episodic information and an interaction between prelexical and postlexical stages of processing– are interactive episodic models such as Minerva 2, and interactive hybrid models such as Goldinger's CLS and POLYSP.

In conclusion, the fact that expectations have an effect on prelexical stages of lexical access suggests that episodes play a role in perception. The evidence is less clear regarding whether underlying abstract representations are also required or not. When additional cues become available, in particular word-level cues, lexical effects set in, overriding the effects that episodes have in tasks and conditions where post-lexical processing is not mandatory. Both episodes and underlying abstract phonological representations can account for these effects. In the case of episodes, the amount of overlap between the label of an episodic cloud and the input, considerably larger than in the segmental condition, ought to facilitate lexical effects and recovery. A very similar explanation can be elaborated for a chain of underlying phonological units: as soon as lexical level information plays a role, underlying units should be able to tolerate a degree of mismatch for a segment given clear evidence for the rest.

2.5. References

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.17, retrieved 20 August 2015 from http://www.praat.org/

Brown, E. (2011). Paradigmatic peer pressure: Word-medial, syllable initial /s/ lenition in Dominican Spanish. In *Selected proceedings of the 5th conference on laboratory approaches to Romance phonology* (pp. 46-58).

Cepeda, G., & Poblete, M. T. (1993). Retención y elisión de /β/ y /ð/ en sufijos y morfemas radicales. *Estudios Filológicos*, *28*, 87-96.

Creelman, C. D., & Macmillan, N. A. (1979). Auditory phase and frequency discrimination: a comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 146.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*(2), 141-177.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27-41.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1), 162-173.

Figueroa Candia, M. F. (2016). *Lenition in the production and perception of Chilean Spanish approximant consonants: Implications for lexical access models* (Unpublished doctoral dissertation). University College London – UCL, London, United Kingdom.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 49-54).

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, *99*(2), 221.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 1-52). New York, NY: Cambridge University Press.

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics*, *13*, 99-188.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3), 373-405.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, *22*(2), 161-200.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*(1), 117-127.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279-312.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*(2), 820-857.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245-294.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*(5), 452.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29-63.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71-102.

Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, *67*(3), 996-1013.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324.

McClelland, J. L., & Elman, J. L. (1986a). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McClelland, J. L., & Elman, J. L. (1986b). Interactive processes in speech perception: The TRACE model. *Parallel Distributed Processing*, *2*(58), 121.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*(1), 73-103.

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The Handbook of Cognition* (pp. 255–275). London: Sage Publications.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive science*, *30*(6), 1113-1126.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(03), 299-325.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191-243.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological review*, *85*(3), 172.

Pérez, H. E. (2007). Estudio de la variación estilística de la serie /b-d-g/ en posición intervocálica en el habla de los noticieros de la televisión chilena. *Estudios de Fonética Experimental*, *16*, 228-259.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*, *7*, 101-139.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Real Academia Española (2014). *Banco de datos (CREA)* [online]. Corpus de referencia del español actual. <http://www.rae.es> [20 August 2015]

Repp, B. H. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *Journal of the Acoustical Society of America*, *74*, 420.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes?. *Journal of Experimental Psychology: General*, *125*(1), 28.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2), 331-353.

Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45-65.

Weenink, D. (2009). The KlattGrid speech synthesizer. In *Interspeech*, *10*, 2059-2062. International Speech Communication Association.

Yeni–Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, *70*(4), 966-975.