The Strength of Segmental Contrasts: A Study on Laurentian French

Sophia D. Stevenson

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

Acknowledgements

There are many people to whom I am sincerely and eternally grateful for their role in bringing about the completion of this dissertation.

Thank you, firstly, to Marie-Hélène Côté, without whom I would not have begun. Thank you for your insight, interest and standing by me through the years. Your warmth, understanding and belief in me carried me through the worst moments of doubt. Thank you for all the time and effort you invested in me. It won't be forgotten. Thank you also to Tania Zamuner, without whom I would not be finishing. Thank you for your patience, for pushing me to always do better and go beyond myself, and for your dedication to seeing this through to the end. You have been a true *Dea ex machina*, arriving in the nick of time. I will always be grateful.

Thank you also to the various professors who have impacted my education in the best ways possible, imparting their love of their sub-fields on to me and nourishing my fledgling ideas, particularly Laura Sabourin (who put the "super" in "supervisor" during my second comprehensive exam), Ian MacKay, Marc Brunelle and Jeff Mielke. Thank you to my fellow students who shaped these past years in innumerable and inevitably enjoyable ways, but especially Marie-Claude Tremblay, Viktor Kharlamov, Santa Vinerte, Félix Desmeules-Trudel, Will Dalton, Allison Lealess, and the past and present members of the BALL lab. To Erin Hetherington, Cristal Loveng, Shannon Doyle, and Brittany Makort: your willingness to accommodate my crazy schedule and be there for me regardless of all the times I couldn't be there for you means more to me than words can say. You are the best exemplars of great friends.

I would be remiss if I did not also thank all the participants in this study who gave of their time so willingly. Your contribution made this work possible and is tremendously appreciated.

Most of all, I would like to thank the members of my family since this is as much their thesis as it is mine. To Annette O'Connor, I owe a debt of gratitude for all the support she has provided during these final months. This thesis would not have been completed without your unfaltering drive to see this through to the end, your easy countenance, your faith in my abilities, and your willingness to keep showing up. You're the guardian angel I never knew I had. To my brother, Alexander: during your own busy life, your MBA, and now your demanding career, you've never turned me away when I've asked for help, unreceptive though I was at times, and you always made the time to provide advice, perspective, an understanding ear, and, most importantly, to call me out on my lame excuses. Thank you for pushing me to be better. To my parents, Harry and Francine: *mille mercis* for your untiring support. I am grateful for Francine's timely personal interventions when I needed them most and Harry's motivating phone calls. The support I have received from both of you has, of course, spanned an entire lifetime and not merely the time it took for me to undertake my degree and write this dissertation. This is the culmination of not just your efforts and mine, but also of your unflagging devotion to my education, my well-being and the possibilities you envisioned for me from a young age. It is impossible to quantify the emotional, financial, psychological and physical support you have both provided to me over all these years, and provided without hesitation or condition. Thank you from the bottom of my heart.

Thank you all.

Abstract

The dichotomy of contrastive and allophonic phonological relationships has a long-standing tradition in phonology, but there is a growing body of research (see Hall, 2013, for a review) that points to phonological relationships that fall between contrastive and allophonic. The criteria most commonly used to define phonological relationships or resolve cases of ambiguous phonological relationships – namely (a) predictability of distribution, and (b) lexical distinction – are not always able to account for observed sound patterns. The main goal of this dissertation is to identify and apply quantitative measures (relative frequency and minimal pair counts) to the traditional criteria in order to better account for cases of intermediate phonological relationships or, in other words, to account for different strengths and degrees of contrast.

Twenty native speakers of Laurentian French (LF) participated in Experiment 1, an AX discrimination task, and Experiment 2, a four-interval AX (4IAX) task, which tested the broader relationships of allophony and contrast. It was hypothesized, based on previous experiments (Boomershine et al., 2008; Dupoux et al., 1997; Ettlinger & Johnson, 2009; Johnson & Babel, 2010; Kazanina et al., 2006; Peperkamp et al., 2003; Pruitt et al., 2006), that phones in an allophonic relationship would be more difficult to perceive than phones in a contrastive relationship. Results confirmed previous findings, with longer reaction times for allophonic pairs as compared to contrastive pairs in the AX task (p<.001), as well as in the 4IAX task (p = .004).

For Experiments 3, 4 and 5, thirty native speakers of LF participated in an AX, a 4IAX and a similarity rating task. Measures of functional load, frequency and acoustic similarity were applied to pairs of phones in allophonic and phonemic relationships in order to quantify the degree of contrast between pairs. If a gradient view of contrast was supported, it was hypothesized that High Contrast vowels [a-ɔ] would yield higher accuracy, faster reaction times and lower similarity ratings; Low Contrast vowels [y-ʏ] would yield lower accuracy, slower reaction times and higher

similarity ratings; and Mid Contrast vowels [o-ʊ] would yield results that fell between the two extremes. If, on the other hand, a strict binary interpretation of contrast was supported, High Contrast vowels and Mid Contrast vowels should yield similar results since these vowels are considered to be in a phonemic relationship, with higher accuracy, faster reaction times and lower similarity ratings, while Low Contrast vowels [y-ʏ], in an allophonic relationship, should yield lower accuracy, slower reaction times and higher similarity ratings.

The results from Experiments 3 (AX) and 4 (4IAX) show that the High Contrast pairs yielded significantly higher accuracy scores and faster reaction times than both Mid and Low Contrast pairs (Experiment 3: $p<.001$ for both High vs. Mid and High vs. Low comparisons; Experiment 4: $p = .039$ for High vs. Mid, $p = .055$ for High vs. Low comparisons). However, no significant differences were found between Mid and Low Contrast pairs in these two experiments. The results from Experiment 5 matched gradient predictions, showing significant differences between High, Mid and Low conditions, with similarity being judged highest for Low pairs, lowest for High pairs, and ratings for Mid pairs falling exactly between the other two levels ($p<.001$ for all comparisons).

While results do not perfectly match gradient predictions, the findings provide evidence counter to a strict binary interpretation of contrast since traditionally phonemic pairs (High [a-ɔ] and Mid [o-ʊ]) were significantly different from one another in all experiments. The lack of difference between Mid and Low Contrast pairs could be due to the measures of functional load and frequency for Mid pairs being closer to those of Low pairs, and thus did not reflect a level of contrast that was equidistant between High and Low Contrast. Nevertheless, taken together with the results from Experiment 5, the results appear to support a gradient view of phonological relationships rather than a strictly dichotomous view. Quantitative measures therefore show promise in accounting for cases of intermediate phonological relationships.

Résumé

La dichotomie entre la relation phonologique contrastive et la relation phonologique allophonique fait partie d'une tradition bien établie en phonologie, mais plusieurs recherches (voir Hall, 2013) indiquent qu'il existe des relations phonologiques qui se rangent entre ces deux extrêmes. Les critères les plus utilisés pour définir les relations phonologiques ou pour résoudre les cas de relations phonologiques ambigües – surtout (a) la probabilité de distribution, et (b) la distinction lexicale – ne sont pas toujours capables de rendre compte de l'ensemble des motifs sonores observés. Le but principal de cette thèse est d'identifier et d'appliquer des mesures quantitatives (la fréquence relative et les comptes de paires minimales) aux critères traditionnels afin de mieux rendre compte des cas de relations phonologiques intermédiaires ou, en d'autres mots, des différentes puissances et de degrés de contraste.

Vingt locuteurs natifs du français Laurentien (FL) ont participé à l'Expérience 1, une tâche de discrimination AX, et à l'Expérience 2, une tâche AX à quatre intervalles (4IAX), afin de tester les relations générales d'allophonie et contraste. L'hypothèse, basée sur les résultats d'expériences antécédentes (Boomershine et coll., 2008; Dupoux et coll., 1997; Ettlinger & Johnson, 2009; Johnson & Babel, 2010; Kazanina et coll., 2006; Peperkamp et coll., 2003; Pruitt et coll., 2006), est que les phones en relation allophonique seraient plus difficiles à percevoir que les phones en relation contrastive. Les résultats confirment ce qui a été trouvé dans les autres recherches, avec les temps de réaction plus longs pour les paires allophoniques comparées aux paires contrastives dans la tâche AX (p<.001), ainsi que dans la tâche 4IAX (p = .004).

Pour les Expériences 3, 4 et 5, trente locuteurs natifs du FL ont participé à une tâche AX, 4IAX et une tâche de jugement de similarité. Les mesures de compte de paires minimales et de fréquence relative ont été appliquées à des paires de phones en relation allophonique et phonémique afin de quantifier le degré de contraste entre les paires. Si une vision de contraste graduée était soutenue, l'hypothèse de

départ est que (1) les voyelles de Haut Contraste [a-ɔ] produiraient des taux d'erreurs plus bas, des temps de réaction plus rapides et des jugements de similarité plus basses; (2) les voyelles de Bas Contraste [y-ʏ] produiraient des taux d'erreurs plus hauts, des temps de réactions plus lentes et des classifications de similarité plus hautes; et (3) les voyelles de Mi-Contraste [o-ʊ] produiraient des résultats qui tomberaient entre les deux extrêmes. Si, par contre, une interprétation de contraste binaire était soutenue, les voyelles de Haut Contraste et les voyelles de Mi-Contraste devraient produire des résultats similaires les uns aux autres puisque ces voyelles sont considérées en relation phonémique, avec moins d'erreurs, des temps de réaction plus rapides et des classifications de similarité plus basses, tandis que les voyelles de Bas Contraste [y-ʏ], en relation allophonique, devraient produire plus d'erreurs, des temps de réaction plus lents et des jugements de similarité plus hautes.

Comme attendu, les résultats des Expériences 3 (AX) et 4 (4IAX) ont démontré que les paires de Haut Contraste produisaient des taux d'erreurs et des temps de réaction significativement plus bas et plus rapides que les paires de Mi- et Bas Contraste (Expérience 3 : $p<.001$ pour les comparaisons Haut vs. Mi-Contraste et pour Haut vs. Bas Contraste; Expérience 4 : $p = .039$ pour les comparaisons Haut vs. Mi-Contraste, et $p = .055$ pour Haut vs. Bas Contraste). Pourtant, aucune différence significative n'a été trouvée entre les paires de Mi- et Bas Contraste dans ces deux expériences. Les résultats de l'Expérience 5 ont correspondu avec les prédictions et ont indiqué une différence très significative entre les conditions de Haut, Mi- et Bas Contraste, avec les jugements de similarité les plus hauts pour les paires de Bas Contraste, les plus bas pour les paires de Haut Contraste, et les jugements pour les paires de Mi-Contraste se trouvant exactement entre les deux autres niveaux ($p<.001$ pour toutes les comparaisons).

Tandis que les résultats ne correspondent pas parfaitement aux prédictions principales, une interprétation de contraste strictement binaire n'est pas soutenue par les résultats puisque les paires traditionnellement considérées phonémiques

(Haut Contraste [a-ɔ] et Mi-Contraste [o-ʊ]) étaient significativement différentes les unes des autres dans toutes les expériences. Le manque de différence entre les paires de Mi- et Bas Contraste pourrait être dû au fait que les mesures de compte de paires minimales et de fréquence relative pour les paires Mi-Contraste étaient plus près de celles des paires de Bas Contraste, et donc n'ont pas reflété un niveau de contraste qui se trouvait à la même distance entre Haut et Bas Contraste. Pris ensemble avec les résultats de l'Expérience 5, pourtant, les résultats semblent soutenir un modèle gradué des relations phonologiques plutôt qu'un modèle strictement dichotomique. Les mesures quantitatives sont donc prometteuses pour rendre compte des cas de relations phonologiques intermédiaires.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

The concept of “contrast” is at the heart of phonological analysis (Avery, Dresher, & Rice, 2008). In phonological theory, the phonological relationship between speech sounds is what enables humans to differentiate between words, allowing us to attach meaning to sound, and is therefore a fundamental component of language. The difference between the initial consonants of ‘fast’ [fæst] and ‘vast’ [væst], for example, serves to signal a difference in meaning and thus distinguishes between lexical items. This distinctive function of speech sounds is one reason why a primary interest in phonology is to determine what sounds are contrastive in a given language: segments that distinguish between lexemes are considered to be in a contrastive relationship and are traditionally viewed as belonging to a stored inventory of underlying phonological representations. Segments that are in an allophonic relationship with each other are typically accounted for by rules or constraints.

However, there is a growing body of research (see Hall, 2013, for an extensive review) that suggests that the traditional dichotomy of contrastive vs. allophonic phonological relationships is more complex. Importantly, the criteria applied to determine whether two sounds contrast cannot account for phonological relationships that appear to fall between fully contrastive and fully allophonic. The cases that do not fit neatly into “fully contrastive” or “fully allophonic” relationships will be referred to as “intermediate” relationships. While the concept of gradient contrast is not new (see Cohn, 2006; Goldsmith, 1995; Ladd, 2006; Scobbie & Stuart-Smith, 2008), there are an increasing number of authors who employ terms to

describe intermediate relationships such as “quasi-contrastive”, “semi-allophonic”, and “mushy contrast” (Hall (2013) lists over one hundred separate articles). This has lead researchers to re-examine the way phonological relationships are determined (e.g. Dresher, 2008; Hall, 2009, 2013) and points to a need for further exploration into the different types of phonological relationships across languages. The lack of clarity between contrastive/allophonic phonological relationships may be due to either (a) the formulation of the criteria, (b) the way in which the criteria are applied, or (c) the expectation that there are only two types of phonological relationships.

The criteria most commonly used to determine phonological relationships are (Hall, 2013, pp. 223-225):

(a) lexical distinction: two sounds contrast if they serve to differentiate two lexemes or morphemes
(b) predictability of distribution: two sounds contrast if it is impossible to predict which segment will occur in a given environment
(c) alternations: two sounds are allophonic if they participate in allophonic alternations with each other
(d) native speaker judgment: two sounds are allophonic if speakers consider them to be the “same” sound
(e) phonological behaviour: for example, if a sound patterns similarly to other sounds that are contrastive, then that sound may also be considered to be contrastive
(f) phonetic similarity: two sounds in an allophonic relationship need to be somewhat phonetically similar to be considered allophonic
(g) orthography: sounds that are typically written with different graphemes may be contrastive while two sounds typically written with the same grapheme may be allophonic

Problems in determining phonological relationships arise when the criteria conflict. For example, using these criteria, a case of intermediate relationship is provided in Hall (2012) where Canadian Raising is discussed. Diphthongs [ɑɪ] and [ʌɪ] are generally predictably distributed and so the two sequences are most often presented

as allophones of a single phoneme. However, there are surface (i.e. not underlying) minimal pairs such as *writing* [rʌɪrɪŋ] and *riding* [rɑɪrɪŋ], where the diphthongs could be said to contrast before the flap [ɾ] and thus serve to distinguish between lexical items. It is therefore unclear whether such minimal pairs are sufficient to classify the relationship between the two diphthongs as "contrastive" even when they can be shown to derive from the same underlying representation.

Another example of an intermediate relationship can be found in cases of neutralization. In Bengali, nasal vowels contrast with oral vowels in some words, while in others, vocalic nasalization is predictable from the environment (Lahiri & Marslen-Wilson, 1991) where oral vowels become nasalized when they occur before nasal consonants. Languages such as German and Russian, for example, have obstruents that contrast for voicing in most positions but the contrast is neutralized word-finally (e.g. German singular "dog" *hund* [hun**t**] but plural "dogs" *hunden* [hun**d**ən]). Neutralization is an example of how two sounds may satisfy the criteria for contrast in one context, but for allophony in another context. This begs the question of whether this diminishes the strength of the contrast between the two sounds or whether once a contrast exists in any context, the sounds should always be considered "fully" contrastive (see also Dresher, 2009, and Hall, 2007).

Yet another example is found in Hualde (2005) who explores "quasi-phonemic" contrasts in Spanish. Hualde (2005) explains that glides are typically considered allophonic variants of high vowels since they are conditioned by their environment. However, preservation of high vowels may occur in the same environment that conditions glides, creating near-minimal pairs such as "duet" /dueto/ [d**u.'e**.to] with a

hiatus and “duel” /duelo/ ['d**we**.lo] with a glide. It is not clear whether this constitutes a contrastive relationship between [w] and [u] or not since the distribution of glides and high vowels is usually predictable but not always.

The above examples illustrate that the criteria typically used to determine phonological relationships may be applied in different ways and may yield different outcomes depending on the language. The examples also illustrate how contrastive and allophonic relationships are not mutually exclusive. An issue arises when contrastiveness is viewed in binary terms, i.e. when two sounds are said to contrast or not. This dissertation will provide support for the view that there is a continuum of contrast from allophonic to contrastive, with a range of values between each end of the spectrum.

Contrast here is quantified mainly by functional load as measured in minimal pair counts. Functional load is a measure of the frequency with which two sounds contrast in all possible environments (see Brown, 1988; Surendran & Niyogi, 2003; Wedel, Kaplan, & Jackson, 2013). Using vowel pairs in Laurentian French, three pairs of vowels were selected as representative of three different levels of contrast: vowels that participate in many minimal pairs represent the High Contrast condition; vowels that have no minimal pairs *and* are in an allophonic relationship represent the Low Contrast condition; and vowels between these values represent the Mid Contrast condition. To explore the differences in perception of these conditions, three experimental paradigms will be used: an AX discrimination task, a four-interval AX (4IAX) task, and a similarity-rating task. No studies on contrast use all three experimental paradigms to explore the topic of gradient contrast, nor do they report

reaction times for the 4IAX task. Using these three paradigms will allow for a comparison of results across task types and also contribute to the existing knowledge base by using a novel methodology in the 4IAX task, as well as adding to our understanding of the perception of contrast. It will be argued that the evidence from the experimental results points to a gradient nature of contrast.

Laurentian French (LF) is the language under examination in this study. The term “Laurentian” is used to refer to dialects of Canadian French whose roots go back to the St. Lawrence colony, now prevalent across French-speaking Canada (in Quebec, Ontario and elsewhere) but excluding Acadian French which has different phonetic and phonological properties as well as distinct historical origins (Côté, 2012). It is more a general term than “Quebec French” since LF is not restricted to the provincial borders of Quebec, and while there are differences between French dialects spoken in Ontario, Quebec and other (non-Acadian) areas in Canada, certain properties remain constant across provincial boundaries. The phonological inventory and alternations outlined in this study are common to all speakers of LF, and as such, the fact that the participants in the experiments described below come from different locations in Ontario and Quebec should not impact results. The phonological vocalic inventory of LF is provided in Figure 1.1, reproduced from Côté (2012: 239):

Table 2. Oral and nasalized vowels and rising diphthongs

i	*frise* 'curl'	y	*amuse* 'amuse'	u	*cool* 'cool'
ɪ	*quiz* 'quiz'	ʏ	*tube* 'tube'	ʊ	*coule* 'flow'
e	*game* 'game'	ø	*jeûne* 'fast'	o	*paume* 'palm'
ɛ	*faites* 'do.2PL'	œ	*jeune* 'young'	ɔ	*pomme* 'apple'
ɜ	*fête* 'party'				
a	*patte* 'leg'			ɒ	*pâte* 'dough'
ɛ̃	*crainte* 'fear'	œ̃	*jungle* 'jungle'	ɔ̃	*honte* 'shame'
				ɒ̃	*lente* 'slow.FEM'
ɥi	*tuile* 'tile'				
wɛ̃	*conjointe* 'partner'				
wa	*boite* 'limp'			wɒ	*boîte* 'box'

Figure 1.1. Phonological Vowel Inventory of Laurentian French.

The alternations of interest are: (a) affrication, whereby /t/ and /d/ are produced as [t^{s}] and [d^{z}] before high front vowels /i/ and /y/ and their lax counterparts, [ɪ] and [ʏ]; for example, "your.fem.sg." /ta/ produced [ta], versus "you.sg" /ty/ produced [t^{s}y]; and (b) vowel laxing in closed syllables that do not end in [(ʁ), v, z, ʒ][1] whereby high vowels /i/, /y/ and /u/ surface as [ɪ], [ʏ] and [ʊ]; for example, "small.masc." *petit* [pət^{s}i] versus "small.fem." *petite* [pət^{s}ɪt]. In this dissertation, it will be shown that it is easier for participants to perceive contrastive segments such as [t] and [d] than allophonic segments such as [t]-[t^{s}] and [y]-[ʏ].

This dissertation is organized as follows: Chapter 2 reviews the theoretical literature regarding contrast, the criteria for determining phonological relationships, and issues that arise when criteria conflict and lead to contradictory conclusions; Chapter 3 covers the background regarding the mechanics of the experimental methodologies to be employed in this dissertation; Chapter 4 covers the

[1] There is some variation in vowel quality before /R/ where vowels can be either tense or lax, as in [d^{z}i:ʁ] vs. [d^{z}ɪ:ʁ] (Côté, 2012).

experimental literature that supports the psychological reality of allophony and contrast based on psycholinguistic experiments; Chapter 5 details Experiments 1 and 2 which test the perception of the allophony and contrast as broad binary relationships; Chapter 6 addresses the quantification of the most common criteria to determine phonological relationships (i.e. lexical distinction by way of functional load, and predictability of distribution) as a means of calculating gradience in phonological relationships, and also presents the design of experiments carried out in Experiments 3 through 5; Chapter 7 presents the procedure for Experiment 3 (AX task), results and brief discussion; Chapter 8 presents the procedure for Experiment 4 (4IAX task), results and brief discussion; Chapter 9 presents the procedure for Experiment 5 (similarity-rating task), results and brief discussion; Chapter 10 discusses the overall results, their interpretation and conclusion of the dissertation. It will be argued that the findings in this dissertation support a gradient view of contrast and provide evidence counter to a binary, all-or-nothing view of contrastive relationships.

Chapter 2 Theories of Contrast

This chapter discusses how contrast has traditionally been defined, followed by a presentation of the different criteria used to determine phonological relationships. Issues arise when criteria conflict and when other factors condition the criteria, resulting in incompatible conclusions.

2.1. Definitions of Contrast

The most common use of the term “contrast” refers to the property of two units (e.g. a phoneme or a feature) serving to differentiate two higher-level categories. For example, a phonemic contrast refers to a difference that is used to distinguish between lexemes (Gussenhoven & Jacobs, 1998: 49) such as /b/ and /p/ distinguishing /bæd/ and /pæd/; a feature contrast differentiates between two otherwise identical phonemes, such as the feature [voice] that distinguishes the phonemes /b/ and /p/. The function of a contrast essentially defines what “contrast” is. Beyond this generalization, the specifics of how contrastive relationships are determined depend on the phonological framework, for instance, whether it uses features and phonemes (primarily stemming from Chomsky and Halle’s *Sound Pattern of English* (1968)) or exemplars and phonetic categories (Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2001, 2002, 2006). Exemplar theories are of particular interest here because they have different assumptions of what constitutes a category of speech sound and how the relationships between these categories are expressed and evaluated, which is of central importance in this thesis. Other authors have discussed the history of contrast in phonology (see, for example, Avery, Dresher, & Rice, 2008; Dresher, Piggott, & Rice, 1994; Hall, 2007) and this section

is not intended to duplicate this work; however, some of this history will be discussed in the following sections insofar as it is useful in clarifying how phonologists determine what units are in a contrastive relationship and of what these units consist.

2.2. Approaches to Determining Contrast

A typical phonological analysis begins by determining the relationships between categories of speech sounds, whether they are whole segments (i.e. phonemes) or specific features. In generative frameworks, contrast is often approached with an all-or-nothing view: two sounds either contrast, or they do not (Cohn, 2006). Saying that two segments "contrast" is a way of indicating that they participate in a specific type of phonological relationship, and is usually taken to entail that the two sounds are members of a phonological inventory and have underlying representations except when it can be shown that what appears to be a contrast on the surface is derived from other phonemes. For example, Boomershine, Hall, Hume, & Johnson (2008) present pairs of sounds in Spanish and English that exhibit different underlying contrasts and surface contrasts. In English, "leather" [lɛðɻ] and "letter" [lɛrɻ] create a surface contrast, where [ɾ] derives from /t/ (in this case) occurring intervocalically when the first vowel is stressed. However, because the realization of /t/ as [ɾ] can be explained, [ɾ] is not considered to be part of the phonemic inventory of English, even though it can be argued that two lexemes are distinguished by these sounds. Such an example is termed a "surface contrast" to indicate that an apparent contrast exists but that one of the sounds involved is not part of the phonemic inventory of the language. Similarly, in Spanish, *cada* [kaða]

"each", where the [ð] derives from intervocalic /d/, contrasts on the surface with *cara* [kara] "face", but [ð] is not included in the phonemic inventory of Spanish. Table 2.1 reproduces the summary of different relationships between [ð], [d] and [ɾ] in English and Spanish provided in Boomershine et al. (2008: 5):

Table 2.1.
Summary of Phonological Relations among [d], [ð] and [ɾ] in English and Spanish

Pair	[d] - [ɾ]		[d] - [ð]		[ɾ] - [ð]	
Language	English	Spanish	English	Spanish	English	Spanish
phonemic (underlying) contrast	-	+	+	-	+	+
surface contrast	-	-	+	-	+	+

Because segments can contrast in some positions and not in others, creating "surface" contrasts, the question has been raised by a growing number of authors (see Hall (2013) for an extensive list) as to whether or not contrast is indeed all or nothing or whether it is perhaps gradient, with intermediate relationships between purely contrastive and purely allophonic (Hall, 2013) or with gradient realizations of contrast (Cohn, 2006). The question of whether contrast is gradient or binary in nature directly impacts the way in which contrasts and underlying representations are arrived at (for example, by comparing minimal pairs) and can have a significant effect on the outcome of any given analysis, as well as on the implications of what is assumed to be stored in an underlying representation. The criteria most commonly used to determine contrast are discussed in the following section.

2.2.1. Criteria for Determining Phonological Relationships

Hall (2013) extensively explores the criteria most commonly used to determine phonological relationships. As mentioned in Chapter 1, these are (Hall, 2013, pp. 223-225):

(a) lexical distinction: two sounds serve to differentiate two lexemes or morphemes
(b) predictability of distribution: two sounds contrast if it is impossible to predict which segment will occur in a given phonological environment
(c) alternations: two sounds are allophonic if they participate in allophonic alternations with each other
(d) native speaker judgment: two sounds are allophonic if speakers consider them to be the "same" sound
(e) phonological behaviour: for example, if a sound patterns similarly to other sounds that are contrastive, then that sound may also be considered to be contrastive
(f) phonetic similarity: two sounds in an allophonic relationship need to be somewhat phonetically similar to be considered allophonic
(g) orthography: sounds that are typically written with different graphemes may be contrastive while two sounds typically written with the same grapheme may be allophonic

The two most important and commonly-used criteria are *lexical distinction* (also called the "distinctive function") and *predictability of distribution*. These criteria will form the basis of the experiments included in this thesis and so it is worth discussing each of these in turn in greater detail.

2.2.1.1. The Distinctive Function

If two sounds serve a distinctive function – i.e. they are used as a means of distinguishing between two lexemes or morphemes – they are considered to be in a contrastive relationship. The comparison of minimal pairs is the most common method of establishing the distinctive function of two sounds considered contrastive in a language, whether in terms of features or segments. For example, /p/ and /b/ differ in voicing so that in English, they are considered two separate categories because they serve to differentiate lexemes (e.g. /bɪt/ and /pɪt/). However, no English lexemes or morphemes are distinguished by the same degree of aspiration as in Hindy, for example, such that /p^h/ and /p/ contrast underlyingly. Therefore,

while these sounds are contrastive in a language like Hindi, they are allophonic in English (Werker & Tees, 1984). In Hindi, aspirated and unaspirated /p/ contrast as in “to take care of” [pal] and “knife blade” [p^{h}al] (Ladefoged, 1975). In English, voiceless stops are conditioned by their environment: they are aspirated before stressed syllables as in [**p^{h}**ɪt], but will be unaspirated after a sibilant, as in “spit” [s**p**ɪt], or in an unstressed syllable, as in “tepid” ['tɛ.**p**ɪd].

The distinctive function, or the property of differentiating two otherwise indistinguishable units, was central to the structuralist approach to contrast where phonological units could be defined solely in terms of differences and oppositions to other units, i.e. in terms of what a segment is *not* (e.g. Saussure et al., 1916; Twaddell, 1935). Trubetzkoy (1939) provides a good example in his description of the German /r/ in which he states “it is not a vowel, not a specific obstruent, not a nasal, nor an *l*” (p. 73). Although this quotation does not refer to lexical contrast, this example illustrates oppositions being used to establish the phonemicity of a sound. On the other hand, a generativist approach to contrast (e.g. Chomsky and Halle, 1968) allows segments to be defined by the substance of their universal features: for example, a segment *x* will be comprised of features *α* and *β*, regardless of the language in which it is found. Feature matrices may then be compared one against another: for example, *x* contains features *α* and *β* while *z* contains features *α* and *γ*, therefore, features *β* and *γ* contrast because they serve to distinguish between phonemes *x* and *z*. In both types of approach, it is common practice to focus on minimal pairs – that is, pairs of units that differ minimally by one element – because

this comparative method is thought to “reveal the contrasting features in the purest way” (Dresher, 2009: 12).

2.2.1.2. Issues in Determining Contrastive Features

The method of minimal comparison is not always as straight-forward as it may appear. The way in which minimal comparisons are carried out can yield different conclusions about what is to be included in an underlying representation or a phoneme inventory. Indeed, for such a common criterion, there are few attested formalizations of how this minimal comparison should be carried out. Contrastive Specification (Steriade, 1995) and Radical Underspecification (Archangeli, 1988) are two examples of feature theories that use contrast-driven comparisons to determine underlying representations. The two theories differ, however, in what they include in an underlying representation. Contrastive Specification posits that all and only contrastive features are specified underlyingly, and it eliminates predictable feature values (Steriade, 1995), while Radical Underspecification claims that all and only unpredictable features are specified (Archangeli, 1988). In Contrastive Specification, contrastive features can only be determined to be present if they serve to contrast two *fully specified* segments that are compared pairwise. This method of determining underlying features has been termed by Dresher et al. (1994) and Dresher (2009) as the “Pairwise Algorithm” (detailed in Archangeli, 1988: 192):

a. Fully specify all segments.
b. Isolate all pairs of segments.
c. Determine which segment pairs differ by a single feature specification.
d. Designate such feature specifications as ‘contrastive’ on the members of that pair.
e. Once all pairs have been examined and appropriate feature specifications have been marked ‘contrastive’, delete all unmarked feature specifications on each segment.

This algorithm is unique in that it explicitly sets out a process by which to determine what is contrastive and what is not. While this assumes fully-specified underlying representations and is more concerned with features that with phonemes as a unit of comparison, the algorithm is unique in that it attempts to formalize this selection process. Where this algorithm fails is that it may not be able to distinguish members of an inventory either (a) when there are too many features relative to the number of phonemes in an inventory because it is unable to determine which redundant features to discard (since it begins with fully-specified segments); or (b) when two phonemes are minimally contrastive with two other phonemes along a single dimension, and therefore have the same feature specifications. For example, after applying the algorithm to fully-specified representations of the five-vowel inventory of Maranungku for [high], [low] and [back] when only minimally different pairs are used to determine underlying feature specifications, /i/ and /ʊ/ are distinguished by [back] only, as are /æ/ and /ɑ/, making /i/ and /æ/ the same underlyingly (Dresher, 2009: 18):

Specifications for Maranungku according to the Pairwise Algorithm:

	i	æ	ɑ	ə	ʊ	Contrasts
high				–	+	{ə, ʊ}
low			+	–		{ɑ, ə}
back	–	–	+		+	{i, ʊ}; {æ, ɑ}

(Note that lack of a feature specification does not contrast with presence of a feature specification.) Dresher (2009) thus proposes the Successive Division Algorithm (SDA) as a modified way of carrying out feature comparisons which has the advantage of not depending on only those features that are members of minimal

pairs and does not require auxiliary mechanisms for multiple logical redundancies. I will not enter into a detailed comparison of Contrastive Specification versus Radical Underspecification here; the Successive Division Algorithm is presented here as an example of a formalized algorithm for determining contrastive features, and to point to some of the issues that arise from determining contrastive features using a method based on minimal comparisons (when assuming fully-specified underlying representations). This algorithm is also notable in its rarity since few have made explicit the process by which they propose features for underlying representations. As Hall (2013) points out, dividing sounds and features by what is contrastive and what is not is usually the first step in modern phonological analysis and not an end of phonological analysis in and of itself, which can explain why so few describe this process.

Even with the aid of an algorithm, the minimal pair test for contrastiveness by itself is likely insufficient to determine what is part of an underlying representation. One example of this is found in analyses pertaining to a phenomenon in second language (L2) acquisition known as “differential substitution” where speakers of different first languages (L1) will produce different phones for the same foreign phone in a target language. For example, Hungarian learners of English substitute [t] for [θ] while European French learners substitute [s] for [θ] (Weinberger, 1990). This is frequently explained by proposing different underlying representations being transferred or “mapped” onto the new target phone based on the L1 inventories. This is not too problematic when the inventories of two languages being compared contain different consonants, and thus different feature contrasts could be argued to

be active in each language based on the rest of the phonological inventory. In other words, different underlying representations for Hungarian [t] and French [t] may be posited depending on other contrasts in the language, explaining why [s] might be a better substitute for French learners. A problem arises, however, when learners speak different dialects of the same L1, such as European French and Laurentian French (LF). Both dialects have the same phonological consonantal inventory so that a comparative analysis should yield the same underlying features for the same consonants. However, European French speakers substitute [s] and [z] for English [θ] and [ð], while LF speakers substitute [t] and [d] for the same English segments (Lombardi, 2003). If the assumption that L1 feature matrices are being transferred onto a novel L2 phone is correct, then European French speakers and LF speakers should substitute the same consonants. Using any kind of algorithm would yield the same features and underlying representations for both dialects of French, and would not be able to account for the differences in the L2 substituted phone.

The solution to this particular case may lie in dialect-specific active phonological processes. Jesney (2005) proposes that the redundant feature [strident] is active in Laurentian French due to this feature participating in the affrication process by which /t, d/ become [t^s, d^z] before high front vowels. This feature is not active in European French since it serves no distinctive function, nor do any alternations involving this feature exist, and so should not influence European French speakers' choice of substitute in the same way. This then begs the question of whether or not features that are only active in allophonic processes are stored and accessed in the same way as features that comprise underlying

representations. In Jesney's analysis, she suggests that [strident] is part of the feature matrix that comprises a representation for the affricates [t^{s}] and [d^{z}], which goes against the notion that surface forms do not have underlying representations that are faithful to the surface form without being considered a phoneme. This example illustrates why something as seemingly simple as a comparison of minimal pairs does not always yield clear analyses, nor does it always allow phonologists to arrive at a clear resolution of what is to be included in an underlying representation.

2.2.1.3. Issues in Determining Contrastive Phonemes

With regards to phonemes, the minimal pair test has led to disagreements on the members of a phonemic inventory. For example, in Japanese, [ɸ] and [t^{s}] only occur before [u]; however, in foreign words, [ɸ] and [t^{s}] may occur before other vowels. Vance (1987) therefore includes /ɸ/ and /t^{s}/ as phonemes while Ito & Mester (1995) and Brown (1997) do not since they consider the alternation before high vowels to be a core sound pattern of Japanese and the contrasting loanword lexical items to be marginal.

Another example may be found again in LF. High tense vowels [i, y, u] become lax in closed syllables that do not end with [(ʁ),v,z,ʒ] (though this is somewhat simplified; there is variation in the tense or lax quality of the vowel before [ʁ] (see Côté, 2010)). For example, "petit" [pət^{s}i] 'small.masc.' has [i] in the final open syllable but "petite" [pət^{s}ɪt] 'small.fem.' in the final closed syllable. In morphologically unrelated words, one finds "lu" [ly] 'read.past.part.' and "lune" [lʏn] 'moon.' However, loanwords from English create minimal pairs such as "coule", *flow.3p.sg* [kʊl] and "cool" [kul], although these are relatively few in number. It is not

clear whether low-frequency contrasts such as in Laurentian French and Japanese make these sounds contrastive for the speakers of these languages, or perhaps less contrastive than high-frequency contrasts. These examples illustrate that a disagreement about what legitimizes a contrastive relationship can lead to different phonemic inventories. These examples also bring into question whether a single minimal pair is sufficient to classify the relationship between two phones as contrastive, and whether or not loan words should be among the lexical items being compared. It is important, therefore, to determine which criteria to use to establish phonological relationships, as well as the way in which the criteria should be applied, at the same time recognizing that this might differ across languages.

One way to carry out the minimal pair comparison in a more meaningful and quantifiable way is to calculate the functional load of contrasts in a given language. Functional load measures the frequencies of two given contrastive sounds and the degree to which those two sounds contrast in all possible environments in order to evaluate how much "work" that contrast does as compared to other contrasts (Brown, 1988; King, 1967; Wedel et al., 2013). In terms of the distinctive function, functional load is able to take the simple "yes" or "no" answer of whether or not two sounds contrast and place the relative importance of that contrast on a scale as compared to other contrasts in a given language. The exact method by which this was quantified will be discussed in greater detail in Chapter 6.

To summarize, the traditional method of comparing minimal pairs as evidence of the distinctive function of two sounds does not always lead to the same conclusions about what is included in a phonemic inventory of a language (such as

for Japanese and LF), nor what is included in a feature matrix of a particular segment in an inventory (such as the comparison of European and Laurentian French). One way the distinctive function has been quantified has been by way of functional load, which allows for a more objective assessment of the contribution of a contrast to the overall phonological system of a language. The distinctive function is not, however, the only criterion to be taken into consideration. The criterion of predictability of distribution also plays an important role in determining contrast.

2.2.1.4. Predictability of Distribution

Besides the distinctive function, there is another main criterion that is used to define contrast: the predictability of segmental distributions in a given language. Hall (2009: 2) defines this as the following:

> "Two segments X and Y are traditionally considered to be contrastive if, in at least one phonological environment in the language, it is impossible to tell which segment will occur. If in every phonological environment where at least one of the segments can occur, it is possible to predict which of the two segments will occur, then X and Y are allophonic."

The predictability criterion can also be satisfied by the minimal pair test so that the presence of a single minimal pair could arguably make the distribution of two phones completely unpredictable. Rather than being an all-or-nothing criterion, however, there are degrees of predictability. Hall (2009) quantifies predictability by three probabilistic measures: bias, environment-specific contrastiveness, and systemic contrastiveness. Bias and environment-specific contrastiveness reflect the likelihood of one sound or another occurring in a given phonological environment, while systemic contrastiveness reflects how much uncertainty there is when

choosing one sound or another across all environments. Using type and token frequencies, Hall (2009) devises algorithms to calculate the uncertainty (or *entropy*) of the distribution of segments, allowing for a gradient comparison of the effect that individual words have on the phonological relationship between two sounds across a phonological system. I return to this issue in more detail in the explanation of experimental criteria for the experiments in this study (Chapter 6). Due to the fact that some treat the predictability of distribution as an all-or-nothing criterion while others acknowledge its gradient nature, issues also arise with the application of this criterion.

2.2.1.5. Issues with Applying the Criterion of Predictability of Distribution

Determining phonological relationships becomes more complicated when the criterion of distinctiveness conflicts with the criterion of predictability of distribution. For example, in LF, the lexemes *saute* [sot] 'jump.3p.sg.pres.' and *sotte* [sɔt] 'stupid.fem.' are differentiated solely by their vowels. This satisfies the criterion of distinctiveness. Furthermore, in this example, their distribution is also unpredictable since the environment does not condition one or the other vowel. With these two criteria taken into account, these vowels would traditionally be viewed as contrastive sounds. However, there are other words where the distribution of [o] and [ɔ] is predictable, such as in *sot* [so] 'stupid.masc.' and *sotte* [sɔt] 'stupid.fem.', where the open variant occurs in a closed syllable and the closed variant in an open syllable in morphologically related words, and never occurs in open final syllables. Since their distribution is sometimes unpredictable (associated with contrast) and sometimes predictable (associated with allomorphy), it is not clear whether the relationship

between these sounds should be classified as contrastive, allophonic or as something between the two. The examples provided in Chapter 1 from Spanish, Bengali, and German are other cases where the distinctive criterion is satisfied in some contexts while the distribution of the same segments is predictable in others. When the indicators for typically contrastive relationships contradict each other in this way, it is not clear whether one criterion should override the other, or whether the segments involved should be classified as sharing a relationship that is intermediary to contrastive and allophonic. Rather than forcing a classification of these relationships as fully contrastive or fully allophonic, such cases may be indicative of intermediate levels of contrast. While the gradient nature of contrast may be difficult or even impossible to express in feature-based frameworks of representation, models that use exemplars are able to encode quantifiable measures into representations, as will be shown in the following section.

To summarize, the two main criteria of distinctiveness and predictability of distribution do not always provide a clear analysis of which segments may be considered contrastive, nor what features may be considered to be present in an underlying representation. The problem may lie in how these criteria are applied in different analyses, or it may lie in the assumption that contrast is binary in nature. Feature-based theories are not well equipped to represent a gradient view of contrastiveness, but other frameworks (such as exemplar theory) are better able to accommodate such a view.

2.2.2. Phonological Representations in Exemplar Theory

The criteria of predictability of distribution and lexical distinction are also valid for an exemplar-theoretic approach to defining contrast, though the categories participating in a contrastive relationship are not viewed in the same way. This section provides details on the view of categories in exemplar theory (Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2001, 2002, 2003, 2006).

2.2.2.1.Phonetic Categories

Before we can enter into a discussion of phonological relationships, it is necessary to clarify what the relationship would be between, or in other words, what it is that participates in a phonological relationship. The analogue to the phoneme in exemplar theory is the "phonetic category," except that the phonetic category as a construct is less abstract than are phonemes in that they may include redundant information (Pierrehumbert, 2003). Pierrehumbert (2003) states that they can be viewed as "peaks in the total phonetic distribution of the language" across a mental phonetic map, parametrized with acoustic, articulatory and perceptual information.[2] In exemplar theory, experience both creates and reinforces phonetic categories and the relationships between categories. The contents of the exemplar-theoretic category include multi-modal stored experiences such as (in the case of speech) acoustic, articulatory, semantic, pragmatic and social information (Bybee, 2001, 2006; Goldinger, 1996; Pierrehumbert, 2000). Categories consist of remembered tokens of experience organized in a cognitive or mental phonetic map. When similar

[2] Pierrehumbert (2003, 2006) also recognizes other levels of representation (e.g. lexical) but since the focus of this dissertation is primarily concerned with phonetic categories, I will not delve into other levels.

remembered tokens reach a large enough number, this group is generalized and a category is formed (Bybee, 2006; Pierrehumbert, 2000, 2001).

A category is more robust when its associated exemplar tokens are more frequent since every new token mapped to an existing category strengthens that category by grouping more and more similar exemplars together (Bybee, 2006; Pierrehumbert, 2001). For example, high-frequency exemplars are more resistant to change (Bybee, 2006) suggesting that frequency contributes to category robustness. Frequent recent experiences of exemplars will also have higher resting activation levels than infrequent exemplars (Pierrehumbert, 2001), and as such, frequency effects will directly impact the processing of a sound. Frequently-encountered categories will also be favoured in speech perception because this involves resolving a competition between possible alternative classifications; the cumulative force of the more frequent exemplars will steer the resolution of that competition in one direction or another (Pierrehumbert, 2006). The frequency of speech sounds in their various environments therefore influences categories and speech processing. Similar tokens are organized in terms of members that are more or less central to the category rather than in terms of features (Bybee, 2006). These facts are not easily captured by more abstract conceptions of underlying representations.

Token frequency alone is not enough to constitute a category – it must be coupled with the ability to judge the similarity of different sounds (Bybee, 2006; Pierrehumbert, 2002). When a listener hears a new token, its similarity to other exemplars is computed and classified according to other similar tokens on the existing mental map of categories (Pierrehumbert, 2002). Being able to judge

similarity is important not only for initial category creation, since a learner must know how to compare exemplars, but it is also integral to the subsequent perception and categorization of new tokens.

This brings up the important question of *how* similarity is judged. If different tokens need to be grouped by their similarity to one another, then at some point a learner must determine which cues are more important for classification. This act of paying attention to particular cues that are more relevant to the categorization process is called “attentional weighting” or “cue weighting” (c.f. Goldstone, 1998; Johnson, 1997; McGuire 2007). Since some phonetic cues signal a phonological contrast and others do not, a speaker must learn which variants of a phonetic category are considered to be within that same category, and how much a phone may deviate before it signals a category change, or becomes unrecognizable as a category of that language. Some evidence that frequency plays a role in category formation comes from research looking at statistical learning with infants between 6 to 8 months of age (Maye, Werker, & Gerken, 2002). Infants show different preferences of stimuli from a [da]-[ta] voice onset time continuum depending on whether they were first exposed to a bimodal distribution or unimodal distribution of the contiunum. While infants in both groups heard examples of each token along the continuum, the bimodal group had more frequent tokens from the endpoints of the continuum, while the unimodal group heard stimuli from the centre of the continuum more frequently. Only infants exposed to the bimodal distribution were able to discriminate between stimuli, suggesting that frequency influences the formation of speech categories.

2.2.2.2. Relationships Between Categories

Relationships between categories, according to Hall (2012), are defined by the number of links shared between other levels of representation (e.g. phonetic, lexical, semantic, orthographic, etc.) of those categories. When one category is "linked to" another, it means that a group of exemplars has something in common with another group: categories in an allophonic relationship have more shared links via other levels of representation, while categories in a contrastive relationship have fewer shared links. For example, in English, [d] and [ɾ] are in allophonic relationship because they share phonetic properties, are linked to lexical categories such as "ride" and "rider" that share the semantics of "ride," and orthography; [h] and [ŋ] are not in an allophonic relationship in English even though they are in complementary distribution because they do not share enough links on other levels of representation (Hall, 2008: 6).

Hall (2008) uses type and token frequencies in order to calculate the predictability (i.e. probability) of distribution of sounds in Canadian English as a measure of contrast. She found that diphthongs implicated in Canadian Raising ([ɑɪ] and [ʌɪ], [ɑʊ] and [ʌʊ],) were less contrastive according to token frequency calculations than type frequency calculations. For example, it may be possible for two sounds to occur in a given environment (type frequency) but one sound may occur twice as often as the other sound in the same environment (token frequency), rendering the relationship less contrastive than one where the occurrence of both sounds in the same environment is equally frequent. This begs the question of how much these less-frequent co-occurrences affect contrast. While frequency is central

to category creation, it is also important when defining the relationships between categories and it can be used to quantify levels of phonological relationship. For example, Bybee (2001) found that in adults, the perception of /t/ deletion is more prevalent in high-frequency words than in low-frequency words. It would not be possible to represent variation in an alternation in a feature-based model, nor the effect that frequency has on establishing relationships between categories.

To summarize, frequency directly impacts the strength of representations as well as the strength and nature of the relationships between categories. As a result, frequency calculations will be used as a measure of different levels of phonological relationship or strength of contrast between the vowels examined in the experiments presented in this work. Furthermore, the role of the minimal pair comparison in determining contrastive segments is not to be dismissed due to its sometimes problematic nature, but rather, using quantitative measures of minimal pair counts will yield a gradient spectrum of contrast. Following exemplar theories, there is assumed to be a representation for each speech sound and that sounds may be more or less contrastive to one another along a continuum depending on the connections shared on other levels of representation.

Chapter 3 Experimental Mechanics: Task Types

This chapter presents the experimental tasks used in this dissertation. Besides explaining the mechanics of each task, this will also set the stage for the discussion of the background literature pertaining to the perception of contrast and allophony. This chapter begins with the AX discrimination task, followed by the four-interval AX task, and then the similarity-rating task.

3.1. AX Discrimination

AX tasks have been used to explore a wide variety of questions and issues in speech perception. These include: the perception of allophones (Boomershine et al., 2008; Werker & Tees, 1984); multiple levels of speech processing (Pisoni, 1973; Werker & Logan, 1985); the effect of time elapsed between presentation of stimuli (known as "interstimulus interval"; Pisoni, 1973; Werker & Logan, 1985; Werker & Tees, 1984), support for exemplar-based speech processing (Ettlinger & Johnson, 2009); the effect of training and task type on perception (Aliaga-García & Mora, 2009; Curtin, Goad, & Pater, 1998), the ability to discriminate phonological contrast by children (Graham & House, 1970), and the effect of high- versus low-probability phonotactics on perception (Vitevitch & Luce, 1998).[3]

In an AX task, participants hear two stimuli and are instructed to determine whether the two stimuli are the same or different. Three main types of stimuli are used: "physically identical," where the "same" stimuli consist of exactly the same

[3] Much more has been published on each of these topics, but these studies incorporate AX tasks specifically. For example, the related AXB task has also been adapted to use with children, such as the Dinosaur Task (e.g., Bishop, Adams, Nation, & Rosen, 2005; Noordenbos, Segers, Serniclaes, Mitterer, & Verhoeven, 2012a and 2012b).

token; “name identical,” where the “same” stimuli consist of the same sounds but are acoustically different; and “different,” where the stimuli differ along a given dimension so as not to be in the same phonetic category (Werker & Logan, 1985). The decision around which type of stimuli to use is related to the type of task and questions the researcher is aiming to answer: if one expects participants to be performing a low-level acoustic perception task, one would include name identical stimuli to see whether participants were sensitive to the minute differences between tokens of phonologically same sounds. If results from physically-identical stimuli are compared to results from name-identical stimuli and one finds that the results are not statistically different between stimulus types, it would indicate that participants are using a higher-level mode of processing and ignoring fine acoustic differences (i.e. treating physically different tokens as members of the same category). In this study, name-identical stimuli were used in order to be able to verify whether or not participants were using a low-level acoustic mode of perception or a high-level phonological mode of perception.

In addition to presenting different types of stimuli, the AX task can be fixed or roving, depending on whether the stimuli varies within a block. In a fixed AX task, only one stimulus pair is presented within a block, so while the stimuli may be same or different, or presented in a different order (AX or XA), it is always the same two stimuli that are being presented. In a roving AX task, both A and X may change in each trial. A roving design is used in the below experiments in order to test multiple conditions and maximize the number of trials in the experiment. The AX task has the advantage of being simple to explain to participants, but the disadvantages include

not knowing what strategies participants use to judge how two sounds are same or different. For example, AX tasks can have a very strong response bias where participants may only choose “different” if they are very sure of their response (Gerrits & Schouten, 2004), and will tend to choose “same” if unsure.

3.2. Effect of Interstimulus Interval

One important aspect of several experimental methodologies, including the AX task, is the period of time between stimuli – or the “interstimulus interval” (ISI). It has been argued that the ISI may determine what mode of perception participants use when performing the task (Pisoni, 1973, 1975; Studdert-Kennedy, Shankweiler, & Pisoni, 1972; Werker & Logan, 1985). In other words, this task, and indeed all discrimination tasks, may tap into different modes of processing depending on the ISI, with longer ISIs encouraging a phonological mode of processing, obscuring finer acoustic differences, and shorter ISIs encouraging a more phonetic/auditory mode of processing (discussed in greater detail in Chapter 4).

However, some researchers who have examined the effect of ISI have found contradictory results. In some studies that control for ISI as a result of the Werker & Logan (1985) study such as Brannen (2002) (though using an AXB task) no effect of ISI was found, while in others such as Gerrits (2001), using a two-interval forced choice task and a 4IAX task (described below) it was found that longer ISI improved discrimination, which is not consistent with Werker and Logan’s (1985) results since they found that longer ISI obscured acoustic differences. It is therefore difficult to determine whether the effect of ISI is contingent on the task or is a generally-applicable rule.

Extremely short ISI (250 ms or less) is reported to tap into auditory processing only. However, using a short ISI does not seem to necessarily guarantee the absence of phonological effects. Boomershine et al. (2008) tested whether allophones are perceived as less distinct than a pair of contrastive sounds within an L1 (English and Spanish) using a speeded AX discrimination task with an ISI of 200 ms. Listeners heard two VCV syllables, where the consonant was either [d], [ɾ] or [ð], and had to decide if they were the same of different. Reaction times, reported as normalized Z scores, were interpreted as representative of similarity, where slower reaction times were associated with greater similarity and faster reaction times were associated with less similarity. Results indicated that English speakers perceived [d]/[ɾ] (allophonic relationship) as more similar than [d]/[ð] (phonemic relationship). In other words, participants had slower RTs for allophonic pairs and faster RTs for phonemic pairs. Spanish speakers perceived [d]/[ð] (allophonic relationship) as more similar than [d]/[ɾ] (phonemic relationship), mirroring the results from English participants who also perceived phones in an allophonic relationship as more similar. The RT results closely matched similarity rating results from the same study on the same stimuli, although using an ISI of 1000 ms which was designed to encourage a phonemic mode of processing by participants. These results lead the authors to conclude that it may be impossible to separate "phonetic responding" from "phonological structure" (p. 17). In other words, the length of ISI does not appear to guarantee either phonemic or phonetic processing: language specific phonological effects were still found with short ISI, and these results matched the similarity-rating task with longer ISI. In the present study, a long ISI of 1500 ms is

used as a precaution, but it is uncertain whether this will guarantee a phonological mode of processing.

3.3. Four-Interval AX Task (4IAX)

The 4IAX task is less common in research on speech perception. Topics explored with this methodology include: the perception of allophones (Peperkamp et al., 2004); multiple levels of speech processing (Pisoni, 1975); the effect of task type on discrimination (Gerrits, 2001; Gerrits & Schouten, 2004); and the state of second language acquisition tested by way of the perception of allophones (Larson-Hall, 2004).

This task can take several different forms. In the basic 4IAX task, two pairs of stimuli are presented, where one pair contains identical stimuli and one pair contains different stimuli (e.g. AA-BA, AB-BB, etc.). The listener must decide which pair of stimuli is the same: the first or the second (Gerrits & Schouten, 2004; Larson-Hall, 2004). In the basic version, the intervening pause between the first and second pair is a bit longer to emphasize the fact that there are two separate pairs of stimuli, rather than have a shorter pause giving the impression of a string of four stimuli.

In another form of the 4IAX task, also known as the “four-interval oddity” task or the “four-interval, two-alternative forced choice” task (4IAXFC), two outside “flanking stimuli” are kept constant to provide a reference frame while two medial stimuli change, e.g. AB-AA, AA-BA (Gerrits & Schouten, 2004). The ISI is kept constant. It is a “forced choice” because participants only have a fixed number of possible responses and are forced to choose between one answer or the other. Each

stimulus presentation is known as an “interval” and so each trial contains four intervals, or one quad (i.e. two pairs) of stimuli. The listener must decide if the odd stimulus out is in the second or third interval, though this is essentially the same as deciding which pair contains identical stimuli (Gerrits & Schouten, 2004). In the 4IAXFC task, the flanking stimuli facilitate auditory comparison since there is a ready point of reference equidistant from the second and third stimuli (Gerrits & Schouten, 2004).

One advantage of the 4IAX task is a reduction of the bias present in an AX task because both “same” and “different” pairs are presented in each trial and therefore a participant cannot default to a single response when unsure (McGuire, 2010). For example, in an AX task, participants hear two stimuli and do not know whether the stimuli are the same or not, and when unsure, are more likely to choose “same” (Gerrits & Schouten, 2004). In the 4IAX task, participants know one pair must be “same” and the other must be “different.” Being unsure should not bias responses to either the first or second pair. According to Larson-Hall (2004), a further advantage of the 4IAX task is that it avoids a response bias when discrimination is difficult and discourages listeners from making judgments based on non-linguistic factors such as minor acoustic differences between stimuli.

One disadvantage of the 4IAX task is that reaction times can be difficult to measure (McGuire, 2010). Participants are able to respond after the second stimulus is heard, but may choose to respond after the third or fourth stimulus for reassurance that they have made the correct decision (McGuire, 2010). In the present study, participants will be forced to listen to all four stimuli by introducing two

additional response options: in addition to “first pair same” or “second pair same,” it is possible that both pairs contain same stimuli, or that both pairs contain different stimuli, requiring participants to listen until the very end of the stimulus before making a judgment. In this way, reaction times may be more accurately measured.

3.4. Similarity-Rating Task

Similarity-rating tasks have been used to look at the perceived similarity of allophones (Boomershine et al., 2008), the correlation between entropy and perceived similarity (Hall, 2009); the perception of non-native phonemes (Johnson & Babel, 2010); and the acquisition of non-native allophony (Hall, 2008).

This task may be comprised of the same kind of stimuli as an AX task, but the instructions differ so that instead of deciding whether two stimuli are the same or different, listeners must quantify how much they find the stimuli similar or different. To do this, participants are provided a scale from very similar to very different, known as a “Likert scale,” usually on a scale of 5, 6 or 7 points. There are different advantages and disadvantages to each of these scales. With odd-numbered scales, the mid-point is typically selected when participants are unsure of their response, while if the scale is too broad, participants will tend to avoid the endpoints (Colman, Norris, & Preston, 1997; Dawes, 2008). An even-numbered scale forces participants to classify what they hear either on the side of “more similar” or “more different” since there is no mid-point on the scale.

To compensate for various participant strategies, results should be normalized to equally represent the differences between conditions. For example, a

participant who avoids using endpoints on a 5-point scale may predominantly respond with 2s, 3s and 4s while another participant may use 1s, 3s and 5s for the same conditions. Z-score transformations are used (Boomershine et al., 2008; Hall, 2008) to normalize responses across participants. The mean per condition is then represented by a value of 0, and any value above and below this represents how many standard deviations away from the mean a particular response was, which in turn represents the extent to which the other conditions are perceived as more similar or more different from one another.

3.5. Summary

Three task types will be used in the experiments in this thesis: AX, 4IAX and similarity-rating. Each has different advantages and disadvantages, but across experimental paradigms, ISI could play a crucial role in determining whether stimuli are processed in a phonological manner or simply acoustic manner, with longer ISI encouraging a phonological mode of perception. Furthermore, each of these experimental paradigms has been used independently to gather evidence of phonological contrast influencing speech perception; however, they have never all been used in a single study. Finally, while a similarity rating task has been used to show evidence of the effect of phonological relationships on perception, by itself, such a task cannot give insight into why participants deemed one pair of sounds as more or less similar than another. Using the AX and 4IAX tasks in conjunction with a similarity rating task will provide insight on the reasons motivating speakers' judgments (i.e. how they judge similarity). This dissertation will use all three

experimental paradigms and in doing so contribute to determining how results from these tasks can be compared across paradigms.

Chapter 4 Previous Experimental Investigations of Contrast

Many aspects of speech perception have been tested using different types of psycholinguistic experiments. This chapter will examine a subset of these – namely, discrimination experiments – in terms of the levels of language processing they purportedly tap into, the aspects of a particular linguistic theory they support, and the pros and cons of different experimental paradigms. The discussion will centre on the three types of experimental tasks used in the present study: the AX task, the 4-interval AX task and the similarity-rating task.[4] Aspects of speech perception that are relevant to the perception of allophony and contrast are also discussed.

Generally speaking, discrimination experiments are designed around our ability to differentiate between stimuli (be they sounds, colours, or other stimuli) and to make a judgment based on a comparison between these stimuli. Such experiments come in various formats, but one of the most common is the AX task where "A" designates one stimulus and "X" designate the stimulus being compared to A. X is normally the stimulus of interest to an experimenter being compared with A, or in the case of a three-way comparison, X may be compared with another stimulus B: AX, ABX, AXB, and the 4-interval AX or "dual pair" task, though this is not an exhaustive list.[5] The similarity-rating task is also a form of discrimination task since stimuli are compared and participants are forced to quantify how similar they find the stimuli. The literature using experiments designed to test the perception of phonological

[4] While similarity-rating tasks are not typically counted among discrimination tasks, discrimination between stimuli is an essential step that must be completed before deciding how similar or how different two stimuli are. I therefore consider it part of the same task family as discrimination tasks.

[5] The "four-interval AX" task is so-called because stimuli are presented at four intervals.

relationships is not very extensive, but in general, one form or another of discrimination experiment is used. This chapter explores how various discrimination experiments have been used to test phonological perception and the differences that have been found with regards to the perception of contrast and allophony.

4.1. Different Modes of Processing: Phonetic vs. Phonological Perception

To briefly reiterate what was discussed in Chapter 3, in an AX task, participants hear two stimuli and are instructed to determine whether the two stimuli are the same or different. It has the advantage of being simple to explain to participants, but the disadvantages include not knowing what strategies or parameters participants use to judge how two sounds are same or different. Furthermore, this task, and indeed essentially all the discrimination tasks, may rely on differences of interstimulus interval (ISI; i.e. the time between stimuli) to distinguish between what level of processing is being tested (Werker & Logan, 1985). If the ISI is short, it is said that participants use a low-level auditory mode of processing, whereas if the ISI is long, it is said that participants use a higher-level phonological mode of processing. This is important when interpreting claims made by various researchers because many experimental designs rely on this parameter, even resulting in a sub-set of AX task known as the “speeded AX task” which uses an especially short ISI to purposefully tap into low-level processing (Boomershine et al., 2008).

Knowing the mode of processing is crucial to interpreting experimental results. For research aiming to explore the psychological reality of phonology, it is necessary to establish whether participants are simply processing stimuli in a phonetic manner,

as they might with non-speech sounds. For example, Best, McRoberts and Sithole (1988) found that English-speaking adults were capable of discriminating Zulu clicks, even though these were novel sounds to participants, because they treated them as non-speech sounds.

How do we know that perception involves multiple levels of processing, and moreover, for researchers interested in phonology, how can we ensure that a particular experimental design is in fact able to tap into the phonological mode of processing specifically? With regards to the first of these two questions, evidence for multiple levels of processing were first put forward in studies by Pisoni (1973, 1975) and colleagues (Studdert-Kennedy, Shankweiler, & Pisoni, 1972). These authors posited a two-level speech perception process: auditory and phonetic. The "auditory" stage refers to transforming the acoustic signal into different psychological dimensions (e.g. pitch, loudness, timbre, duration) roughly corresponding to measurable dimensions of a spectrogram (Studdert-Kennedy et al., 1972). The "phonetic" stage refers to the "transformation of psychological (auditory) dimensions into phonetic features" (Studdert-Kennedy et al., 1972: 456) such as Place and Voice. It may be argued that these "phonetic features" are in fact part of phonological structure, so this "phonetic" stage could be interpreted as analogous to a phonological mode of processing.

Werker and Logan (1985) went a step further and provided evidence for more than two levels of processing. This study examined the possibility of different levels of processing using three AX experiments and posited that differences in ISI determine whether listeners use an acoustic, phonetic or phonemic mode of

perception, corresponding to 250 ms, 500 ms and 1500 ms respectively. If the ISI was short, it was predicted that participants would use a low-level auditory mode of processing, whereas if the ISI was long, it was predicted that participants would use a higher-level phonological mode of processing. Their study looking at English listeners' perception of Hindi place contrasts using unaspirated retroflex and dental consonants. In the first experiment, no effect of ISI was found in the proportion of "same" responses by type of pairing, i.e. physically identical pairs, "name identical" pairs (i.e. the same stimuli but acoustically different tokens) or different pairs. In their second experiment, they modified the initial within-subjects design to a between-subjects design and also recorded reaction times (RTs). This time, a main effect of ISI was found on the proportion of Same responses: as ISI progressed from 250 ms to 500 ms to 1500 ms, differences between pair types were reduced, suggesting decreased sensitivity to acoustic differences between stimuli with longer ISI. With regards to RT results, while there were differences in response time among the three pair types, these were greatly reduced in the 1500 ms condition echoing the results for the proportion of "same" responses. The proportion of Same responses combined with the RT results led the authors to conclude that there was evidence of three levels of processing, with the shortest ISI corresponding to "auditory" perception, longer ISI corresponding to "phonetic" perception, and the longest ISI at 1500 ms resulting in a "phonemic" mode of perception. As discussed in Chapter 3, there is contradictory evidence with regards to ISI, with extremely short ISIs not being able to completely mask phonological effects (Boomershine et al., 2008), but since it appears to be possible to manipulate the extent to which sounds will be perceived phonetically or phonologically, there seems to be enough evidence to

support multiple modes of processing. In order to encourage a phonological mode of processing, ISI was set to 1500 ms for the experiments in this dissertation.

4.2. Perception of Contrastive and Allophonic Relationships

Many studies have tested the perception of phones in contrastive and allophonic relationships, though not specifically to test phonological relationships. For example, it has long been established that humans perceive non-contrastive phones with some difficulty as in the case of the perception of voicing contrasts by English speakers, where non-aspirated and aspirated phones, for example [t] and [t^h], are allophones of /t/ but distinct phonemes in other languages, such as Thai and Hindi (see, for example, Lisker & Abramson, 1970; Polka, 1991; Pruitt, Jenkins, & Strange, 2006; Werker et al., 1981). It has also been shown that the "difficulty" in perceiving gradience, such as on a scale of voice onset time, depends on the listeners' established categories (e.g. Lisker & Abramson, 1970; Pisoni & Tash, 1974). This suggests that phones are perceived according to their established phonologically contrastive categories, which is known as "categorical perception."

It has also been found that consonants and vowels may be processed differently. Nespor and colleagues (e.g., Nespor, Peña, & Mehler, 2003; Mehler, Peña, Nespor, & Bonatti, 2006) suggest that consonants play a greater role than vowels with regards to specifying lexical entries and for parsing speech streams. Neurologically-speaking, Caramazza, Chialant, Capasso, & Miceli (2000) found that consonants and vowels may be processed in different parts of the brain by looking at the performance of aphasics whose ability to process speech was restricted to either vowels or consonants, but not both. However, with regards to the perception

of contrastive segments, evidence remains consistent across consonants and vowels.

4.2.1. Perception of Contrastive and Allophonic Relationships: 4IAX Task

No studies have specifically explored the perception of contrastive vs. allophonic relationships by way of a 4IAX task, but the methodology employed in Peperkamp et al. (2004) is similar in that two pairs of stimuli were presented to participants and they had to compare stimuli within and across pairs. Peperkamp et al. (2004) tested European French speakers on their perception of the uvular voiced fricative [ʁ] and its voiceless allophone [χ] which only occurs next to voiceless segments. In one task (500 ms ISI), participants were told they would hear two monosyllabic "words" (called the "isolation" condition) and had to choose whether the words were the same or different. The second task was reminiscent of a 4IAX task, where participants heard two pairs of two syllables and were asked to judge similarity across the pairs. It consisted of two pairs of VC.CV "sentences" (called the "context" condition) where participants were instructed to indicate whether the first words of both sentences were identical or different. For the "allophonic" condition, half of the trials contained sentences with a consonant cluster that did not agree in terms of voicing, and was thus phonotactically illegal (e.g. [iχ.be]-[iχ.zu]), and the other half of trials agreed in voicing, and were thus phonotactically legal (e.g. [aʁ.do]-[aχ.sa]). For the "phonemic" condition, which contained the consonants [m] and [n] instead of [ʁ] and [χ], all sequences were phonotactically legal. The authors found that discrimination of the allophonic condition yielded significantly lower accuracy scores when the sounds occurred in the a phonotactically legal context for

the allophones. They conclude that allophones are only difficult to discriminate when embedded within their trigger phonological context. Context is therefore an important element to control for in further experimental research since the difficulty in discriminating allophones may be contingent on this factor.

4.2.2. Perception of Contrastive and Allophonic Relationships: Similarity-Rating Task

Other studies that explore the perception of contrast make use of similarity rating tasks. Allophonic alternations entail a change in phonetic category that does not signal a change in phonological category. In a similarity rating task, it is therefore expected that sounds that do not cue a contrast should be difficult to perceive, and they should therefore be judged as being more similar. A similarity rating task is thought to be able to show subtleties in the range of "belonging" to a category; for example, if a listener judges [t] and [t^h] as being very different, this is believed to reflect a phonological relationship and two separate categories, whereas if a listener judges them as being very similar, this reflects an allophonic relationship, or belonging to the same phonetic category (see, for example, Boomershine et al., 2008).

One study that uses a similarity rating task to explore contrastive and allophonic relationships is Hall (2008). In her study, she looks at similarity ratings of Spanish and English bilinguals at various levels of fluency to explore the evolution of phonological relationships between sounds as learners integrate new contrasts in their target language. Looking at [d], [ɾ] and [ð], she predicts that [d] and [ɾ] will be judged as more similar for beginner English learners of Spanish and for advanced

Spanish learners of English since these phones are in an allophonic relationship in English; conversely, she predicts that advanced English learners of Spanish and beginner Spanish learners of English should perceive [d] and [ð] as more similar since these phones are in allophonic relationship in Spanish. [d] and [ɾ] do not contrast on the surface in Spanish since, in environments were [ɾ] can occur (e.g. intervocalically), [ð] is the surface realization of /d/. While the results did not reach statistical significance, these predictions were borne out as a trend: [d] vs. [ɾ] (allophonic in English) was judged as being more similar by beginner learners of Spanish and more different by intermediate and advanced learners, reflecting the change from native English phonological relationships to target Spanish relationships, while [d] vs. [ð] was judged as more similar by advanced learners of Spanish than by beginner learners. Hall concludes that these results indicate that experience informs and changes the representations of categories.

Another study that demonstrates that similarity rating responses reflect the allophonic or contrastive relationship between two phones is Boomershine et al. (2008). Boomershine et al. (2008) tested whether allophones are perceived as less distinct than a pair of contrastive sounds within an L1, again testing Spanish and English speakers on [d], [ɾ] and [ð]. They use similarity ratings as well as reaction times (both reported as normalized Z scores) from a speeded AX discrimination task where slower reaction times were associated with greater similarity and faster reaction times were associated with less similarity. Results indicated that English speakers perceived [d]/[ɾ] (allophonic relationship) as more similar than [d]/[ð] (phonemic relationship) and Spanish speakers perceived [d]/[ð] (allophonic

relationship) as more similar than [d]/[ɾ] (phonemic relationship). The results for the pair [ð]/[ɾ] patterned like the contrastive pairs for the respective language groups, which is likely due to them being allophones of different phonemes in each language.

4.2.3. Difficulties in Comparing Across Studies

Because of the variety of tasks used to test contrast and allophony, comparing across studies can be challenging. In Larson-Hall's (2004) study of Japanese learners of Russian using a 4IAX task, participants circled their answer on a sheet of paper, and so RTs were naturally not recorded and analyzed. Reaction times were also not recorded in the Gerrits and Schouten (2004) study, though this may be due to reasons that are inherent to the methodology. For the 4IAX design, McGuire (2010) lists difficulty in interpreting RTs among its disadvantages. However, the reasons for this depend on the possible responses. McGuire (2010: 5) states that it is possible for participants to make a decision upon hearing the second stimulus in a four-stimulus trial, but that they may wait until the third or even fourth stimulus for "reassurance" in their decision. This being the case, if participants were to be given more than two options such that instead of searching for the "same" pair, where participants would know as soon as they hear the second stimulus what the answer was, they would be forced to listen all the way to the final stimulus in each trial. Then, reaction times would become comparable across conditions. In the present study, participants were instructed that *both* pairs in a quad may consist of same stimuli and *both* may consist of different stimuli, making for four possible responses and forcing participants to listen to the end of each trial. Furthermore, as with

Werker & Logan (1985), reaction times may indicate differences where none can be seen in accuracy or the proportion of “same” responses. This modification should allow for meaningful interpretation of reaction times for this task. It is challenging to compare 4IAX results to results from other paradigms because some researchers have measured only accuracy (Gerrits & Schouten, 2004; Larson-Hall, 2004) while others find differences among RTs in another experimental paradigm (e.g. Boomershine et al. (2008) with an AX task).

4.3. Summary

This chapter has looked at some of the experimental methodologies used to explore the perception of phones in allophonic and contrastive relationships. Generally speaking, phones that are exemplars of a single category are difficult to discriminate while phones of different categories are easier to discriminate. More specifically, allophones have been found to be difficult to perceive in their trigger context (Peperkamp et al., 2004), and have also been found to be perceived as more similar to one another (Boomershine et al., 2008) than phonemes. Furthermore, experimental factors such as ISI may play an important role in determining whether participants are processing stimuli in a phonological way or in a more auditory way (Werker & Logan, 1985) and so will be controlled for in the following experiments.

In this study, in addition to testing the extremes on the scale of contrastive to allophonic relationships, stimuli will also be included to test intermediate relationships of contrast. Vowels that exemplify High, Mid and Low degrees of contrast will be tested in an AX, 4IAX and similarity rating task to facilitate

comparison across experimental paradigms and comparison with results from previous research. Different results are expected depending on whether phonological relationships are binary in nature (i.e. fully contrastive or fully allophonic) or gradient in nature, where intermediate relationships between contrastive and allophonic exist. The following chapter explains the experimental conditions and predictions for Experiments 1 and 2 which test the broader relationships of allophony and contrast. In subsequent chapters, Experiments 3, 4 and 5 will explore intermediate relationships between the two extremes of the scale of phonological relationships.

Chapter 5 Experiments 1 and 2: Allophony and Contrast

This chapter presents experiments conducted to explore the coarser phonological relationships of allophony and contrast. The findings from these experiments led to methodological choices for the subsequent experiments in this dissertation.

5.1. Experimental 1 and 2 Goals

The main goal of these experiments was to replicate and extend previous research indicating that segments in an allophonic relationship are more difficult to perceive than segments in a contrastive relationship (e.g. Boomershine et al., 2008; Peperkamp et al., 2003). To explore these relationships, the consonants [t], [d], [t^S] and [d^Z] in Laurentian French (LF) were studied by way of an AX task and a 4IAX task. In LF, /t/ and /d/ affricate to [t^S] and [d^Z], respectively, before high, front vowels [i] and [y] and their lax counterparts, [ɪ] and [ʏ]. For example, “petit”, ‘small.masc’ [pəti].

According to the criteria of lexical distinction and predictability of distribution, [t]-[t^S] and [d]-[d^Z] are in an allophonic relationship since (a) no words are distinguished by these pairs of sounds,[6] and (b) their distribution is highly predictable. In addition to allophonic relationships, pairs of phones were also chosen to test contrastive relationships, containing [t], [d], [s] and [z]. Lastly, sibilance is a salient auditory cue (see, for example, Hura et al., 1992) that could have made contrastive pairs such as [t]-[s] too easy to perceive for reasons not related to

[6] Note that there may be a few exceptions that are very low in frequency; for example, “tip” [tɪp] vs. “type” [t^Sɪp] (although this pronunciation does not apply to all speakers of LF). Some other words are also exempt from this process, such as “building” [bɪldɪŋ] and “meeting” [mitɪŋ].

phonological relationships. Control stimuli containing other contrastive pairs of consonants were also included in order to verify whether results obtained for contrastive test pairs would match those of contrastive control pairs. Control stimuli included various combinations of the following consonants: [b], [v], [g], [p], [f] and [k] whose status is also phonemic in LF. (See Table 5.2 below for syllable pairings used in this study.)

5.2. Predictions

Since previous research found that it is difficult to distinguish between pairs of allophones, it was predicted that stimuli containing [t]-[t^{s}] pairs would also be difficult to distinguish, yielding lower accuracy scores and slower reaction times (RTs). Conversely, contrastive phones should be easier to distinguish, and so contrastive [t]-[s] and [d]-[z] pairs are predicted to yield higher accuracy scores and faster RTs than allophonic pairs (Prediction 1). Control items were also included to verify that results obtained for contrastive test pairs (i.e. [t]-[s] and [d]-[z]) were not specific to the acoustic properties of these two particular sounds, but were representative of a contrastive relationship: if results for [t]-[s] pairs are different from those for allophonic [t]-[t^{s}] pairs as well as stimuli such as contrastive [b]-[v], then this would indicate that results are not related to relationship and acoustic differences play a greater role in determining how stimuli are perceived (Prediction 2).

The specific hierarchy of anticipated results in Prediction 2 (below in Table 5.1) was based on acoustic analyses carried out in Mielke (2012). Mielke (2012) developed a phonetically based metric by which to judge the similarity of phones using acoustic and articulatory data, including nasal and oral airflow, vocal fold

activity, larynx height, and ultrasound video of the tongue and lips.[7] Spectral information and vocal tract shape was also used to calculate phonetic distances between phones. Following an analysis using principal components to compare phones, where the first principal component accounts for the most variance between tokens, the second principal component the next most variance, etc., it was also found that different consonants are most similar to one another in terms of place and manner of articulation: [d^{Z}] with [t^{S}], [s] with [z], and [d] with [t]. In terms of acoustic principal components, [t] and [d] pattern together in terms of what may be interpreted as sibilance, while [d^{Z}] and [t^{S}] resemble each other in this same dimension, as do [s] and [z]. There is also a pattern within each set of consonants by voicing, such that [t^{S}] and [s] are slightly more similar to each other than [t^{S}] to [t] due to sibilance, but overall, the three types of segments (stop, affricate and fricative) exhibit clear differences (the same goes for the voiced equivalents); see Figure 5.1.

[7] This study will be discussed further in Chapter 6.

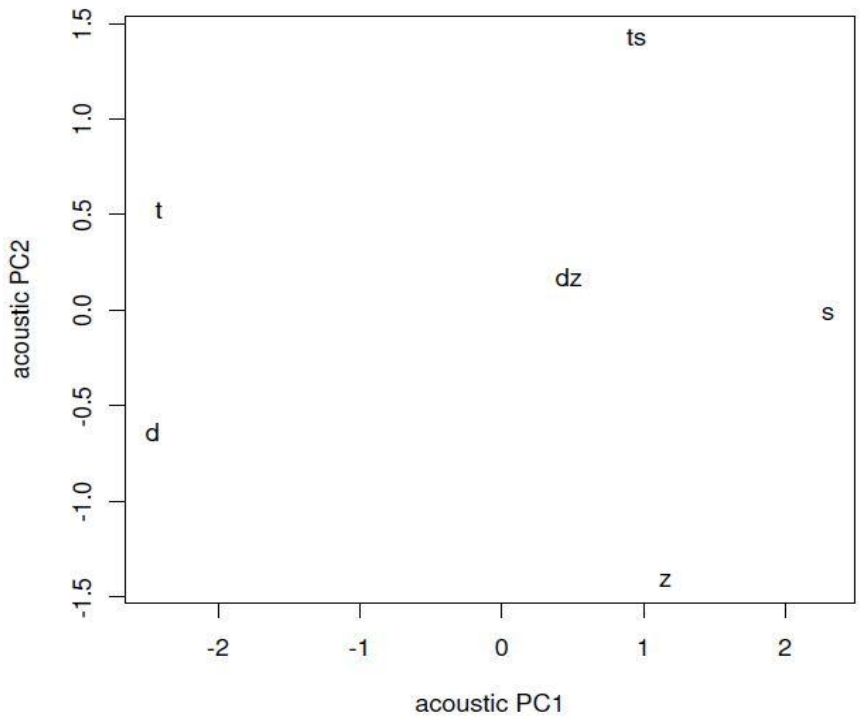


Figure 5.1. Experiments 1 & 2: Acoustic Similarity by Principal Components.

Principal component 1 may be interpreted as sibilance or frication, and principal component 2 as voicing. Consonants shown are based on data from Mielke (2012) and not the present stimuli, but this metric still provides a solid starting point from which to base acoustic predictions if acoustic differences drive the way phones are judged to be similar to one another in Experiments 1 and 2. If acoustic differences are paramount, or if they are favoured by the experimental design, it should indeed be the case that [s] be more difficult to perceive when paired with [t^{s}] than [t] with [t^{s}] or [t] with [s]. This means that [t^{s} -s] would be the most difficult to perceive, followed by [t^{s} -t], and finally [t-s] would be the easiest to perceive. With regards to control stimuli containing a variety of consonants, these were not measured in the principal components analysis, but it would be expected that due to differences between the various Control pairs (see the shaded squares in Table 5.2

for all combinations included in these experiments), results from these stimuli would resemble those of [t]-[s].

Predictions may be summarized as in Table 5.1, where “A” indicates “Allophonic,” “C” indicates “Contrastive” and “Ctrl” indicates “Control Items”; “≠” means “is different from” and “>” means the condition to the left of this symbol is expected to yield higher accuracy or faster RTs than the condition to the right of this symbol. Different predictions are obtained depending on whether results match the acoustics of the stimuli or the phonological relationship between stimuli.

Table 5.1
Predictions for Experiments 1 (AX) and 2 (4IAX): [t^{s}] vs. [t] (Allophonic) and [t] vs. [s] (Contrastive)

Prediction	Accuracy	RTs
1. Relationship	C=Ctrl>A	C=Ctrl>A
2. Acoustics	C≠Ctrl>A	C≠Ctrl>A

It is also expected that Trial Type (Same, Different) will have an effect on RTs due to the response bias discussed in Chapter 3, with Same responses yielding faster RTs since participants are typically more certain of Same stimuli. The stimuli used in Experiments 1 and 2 are provided in Table 5.2, coded for condition.

Table 5.2

Syllable Pairs Coded by Condition

1st σ \ 2nd σ	**ti**	**tsi**	**si**	**fi**	**pi**	**ki**	**di**	**dzi**	**zi**	**vi**	**bi**	**gi**
ti	SC	DA	DC									
tsi	DA	SA	DI									
si	DC	DI	SC									
fi				SC	DC	DC						
pi				DC	SC	DC						
ki				DC	DC	SC						
di							SC	DA	DC			
dzi							DA	SA	DI			
zi							DC	DI	SC			
vi										SC	DC	DC
bi										DC	SC	DC
gi										DC	DC	SC

Note. Shaded squares = control items; blank squares = pairings not used; S = Same; D = Different; A = Allophonic; C = Contrastive; I = Indirect.

These experiments also paired [t^{S}] and [s], labelled "I" for an "Indirect" relationship, namely to see whether [ts] would act as a contrastive segment with [s] since the two are underlyingly contrastive (/t/ and /s/). However, it is possible that [t^{S}] and [d^{Z}] were often perceived as [s] or [z] due to the stop often being misperceived as empty space, and so these stimuli were not included in the present analysis. Stimuli containing consonants in allophonic relationship such as [t^{S}i]-[ti] were labelled "Different-Allophonic" (DA). Pairs containing [di]-[zi] were labelled "Different-Contrastive" (DC) since /t/ and /s/ contrast in LF. For the "same" pairs, there are two possibilities: "Same-Allophonic" (SA), such as [t^{S}i]-[t^{S}i] or [d^{Z}i]-[d^{Z}i], and "Same-Contrastive" (SC), such as [ti]-[ti] and [si]-[si]. Of course, when a pair consists of two of the same token, they do not have a relationship with each other, but it was necessary to see whether there were any differences among Same pairs with [t^{S}] and [d^{Z}] versus [t] and [s] for two related reasons: when no differences are expected, then the presence of differences could signal either (a) a problem with the

methodology, or (b) perhaps a fundamental difference between sounds whose surface representation matches their underlying phonemic content and sounds whose surface representation does not match their underlying phonemic content. While this was not expected, the results for Same stimuli were explored to rule out these two possibilities. The two control conditions are CDC (Control-Different-Contrastive) and CSC (Control-Same-Contrastive) containing control (non-test) stimuli such as [bi-gi]. Coding for "Relationship" in the AX task is as in Table 5.3; coding for "Relationship" in the 4IAX task is as in Table 5.4 (a)-(c):

Table 5.3
Coding of Pairs by Relationship with Example Stimuli – AX Task

Pair Type	**Test**	**Control**
DA	dzi-di	--
DC	zi-di	gi-bi
SA	dzi-dzi	--
SC	di-di	gi-gi

Table 5.4 (a)-(c)
Coding of Quads by Relationship with Example Stimuli – 4IAX Task

(a) Coding of "**Different-First**" Quads

Quad Type	**Test**		**Control**	
DA+SA	tsi - ti	tsi - tsi	--	--
DA+SC	tsi - ti	ti - ti	--	--
DC+SA	ti - si	tsi - tsi	--	--
DC+SC	ti - si	ti - ti	pi - ki	pi - pi

(b) Coding for "**Same-First**" Quads

Quad Type	**Test**		**Control**	
SA+DA	tsi - tsi	ti - tsi	--	--
SC+DA	ti - ti	ti - tsi	--	--
SA+DC	tsi - tsi	si - ti	--	--
SC+DC	ti - ti	si - ti	pi - pi	pi - ki

(c) Coding of "**Same-Only**" Quads

Quad Type	Test		Control	
SA+SC	tsi - tsi	si - si	--	--
SC+SA	ti - ti	tsi - tsi	--	--
SA+SA	tsi - tsi	ti - ti	--	--
SC+SC	ti - ti	tsi - tsi	pi - pi	ki - ki

5.3. Participants for Experiments 1 & 2

Participants were all native speakers of Laurentian French (N=20; 6 men), either from Ontario (Ottawa: N=9; Kapuskasing: N=1) or from Quebec (Hull/Gatineau region: N=8; Montreal region: N=2), ranging in age from 20 to 43 years (average=27, SD=5.9). Most participants also had a good command of English as determined by a self-assessment questionnaire, but none spoke an additional language in which [ts] was contrastive. No participants self-reported any hearing difficulties or learning disabilities. All participants did both experiments with Experiment 1 followed by Experiment 2 so as to proceed in order of increasing task difficulty. None of the participants were compensated beyond juice and cookies.

5.4. Stimuli for Experiments 1 & 2

All stimuli were created from French words as read by a female with training in phonetics and a native speaker of Quebec French from Quebec City, who had been living for several years in the Hull/Gatineau region. She was also fluent in English, although English was acquired later in life.[8] Recordings were made on a Marantz digital recorder with a Sennheiser microphone, and all splicing and acoustic

[8] Research has shown that voice onset time (VOT) values for bilinguals tend to fall between those of monolinguals of both languages concerned (Lee & Iverson, 2012).

analyses were subsequently done with Praat (version 5.2.29) software. Words were chosen in order to elicit the affricates [ts] and [dz], thus all consonants occurred before before [i], as well as control stimuli with common stops and fricatives. The speaker also read words such as “tea” couched in a French carrier sentence in order to elicit syllables such as [ti] without affrication; however, since the speaker knew the goal was to produce an unaffricated [t] for French participants, the speaker’s vowels were not English-like. Acoustic analyses were done on the French vowels and consonants in order to determine: (a) formant values for vowels, (b) stop duration times, and (c) fricative duration times. The most representative and easily identifiable token of each vowel as identified by French-speaking judges was then spliced onto the ideal token of the various consonants so that participants would be unable to use vowel differences to distinguish one syllable from another, thus forcing participants to discriminate based on consonantal differences alone. All syllables were normalized for intensity, set to 70 dB.[9]

These syllables were then coupled with other syllables to create syllable pairs, which became the stimuli in the AX task. These pairs were then coupled into quads for the 4IAX task (as above in Tables 5.3 and 5.4). All trials consisted of either both “test” or both “control” syllables; in other words, a pairing such as [ti] followed by [ki] did not occur; only pairings such as [ti]-[t^{S}i] and [pi]-[ki] were used. All Same pairs were acoustically different tokens (for the consonant portion) and balanced for order so that, for example, participants heard $[ti]_1$-$[ti]_2$ as well as $[ti]_2$-$[ti]_1$. All Different pairs were also balanced for order so that participants heard both [ti]-[t^{S}i] and [t^{S}i]-[ti]. In

[9] Participants were told they could change the volume to what they felt was a comfortable level if they found the stimuli too quiet or too loud. However, none did.

these cases, it was always the same token of a given syllable (i.e. $[ti]_1$-$[t^{s}i]_1$ and $[t^{s}i]_1$-$[ti]_1$). For the Experiment 1 (the AX discrimination task), participants heard 50% voiced consonants and 50% voiceless consonants.

5.5. Procedure: Experiment 1 – AX Discrimination Task

The experiment was presented on a Mac computer (OSX platform) using PsyScope software (build 57) which recorded accuracy and reaction times. Participants were instructed to listen to two syllables and then to press [/] on the keyboard if they thought the syllables were the same, or [z] if they thought the syllables were different. The syllables were separated by a 1500 ms ISI, and the next trial began once they had responded to the previous trial. If participants did not respond within 5 seconds from the beginning of the stimulus, the experiment would automatically advance to the next trial and that trial would be recorded as timed-out.

The experiment began with 6 practice trials where the answers did not count towards accuracy rates or reaction times, followed by two blocks of 48 trials each, making the total number of trials 96, with a self-timed pause in between the two blocks and following the practice trials. 48 trials consisted of “test” syllables, 48 of “control” syllables, equally divided between Same and Different answers. All participants heard each possible stimulus pairing twice, once in each randomized block. This task took approximately 15 minutes to complete.

5.6. Results

5.6.1. Experiment 1: AX Task

For all figures, error bars depict +/- 1 standard error. A *p* value of 0.05 or less is considered statistically significant, while a *p* value greater than 0.05 and less than 0.075 is considered a trend.

5.6.1.1. Experiment 1: Accuracy Results

The mean accuracy scores by Relationship and Trial Type is shown in Figure 5.2. The closer the score is to 1, the more correct answers were given. Figure 5.2 shows that overall, accuracy scores were relatively high with scores for the Different Allophonic stimuli being the lowest.

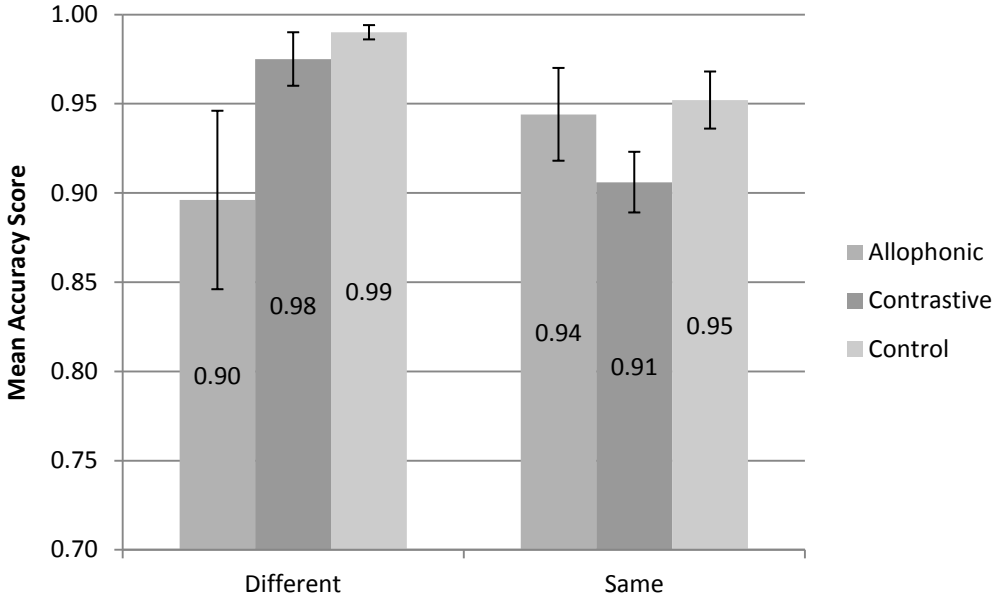


Figure 5.2. Experiment 1: Mean Accuracy Scores by Relationship and Trial Type.

A repeated measures 2x3 ANOVA was done to test the differences between the conditions of Trial Type (Different, Same) and Relationship (Allophonic, Contrastive, Control). The assumption of sphericity was violated for Relationship and so

Greenhouse-Geisser adjusted values are used. No main effect of Relationship was found (F(1.3, 24.8) = 2.65, p = 0.108), nor of Trial Type (F(1, 19) = .47, p = .501), nor was there a significant interaction between factors (F(1.3, 25.0) = 2.98, p = .087). In short, there were no significant differences in accuracy between allophonic or contrastive stimuli.

5.6.1.2. Experiment 1: Accuracy Scores – Brief Discussion

According to Prediction 1, namely that results would pattern according to phonological relationships, the Allophonic condition would elicit lower accuracy scores overall. Among Different Allophonic pairs, this was borne out in raw measures but was not statistically significant due to the high degree of variance among these pairs. Analyses indicated that there were no significant differences in terms of accuracy between sounds in an allophonic relationship and sounds in a contrastive relationship, nor between contrastive and control items.

5.6.1.3. Experiment 1: Reaction Time Results

For the AX task, reaction times were recorded from the onset of the stimulus rather than from the offset of the stimulus. Thus, the duration of the stimulus was subtracted from the recorded reaction time in order to arrive at a more realistic reaction time score for each trial. The following RT analyses include only correct responses. In total, there were 126 total incorrect responses and 8 trials that timed-out (i.e. the response time was greater than 5 seconds) that were removed from the analyses. Responses that fell more than three standard deviations above or below the condition mean were discarded. Of 1786 correct responses, 95 tokens were

discarded (or 5.32%). The mean RTs by Relationship and Trial Type are shown in Figure 5.3.

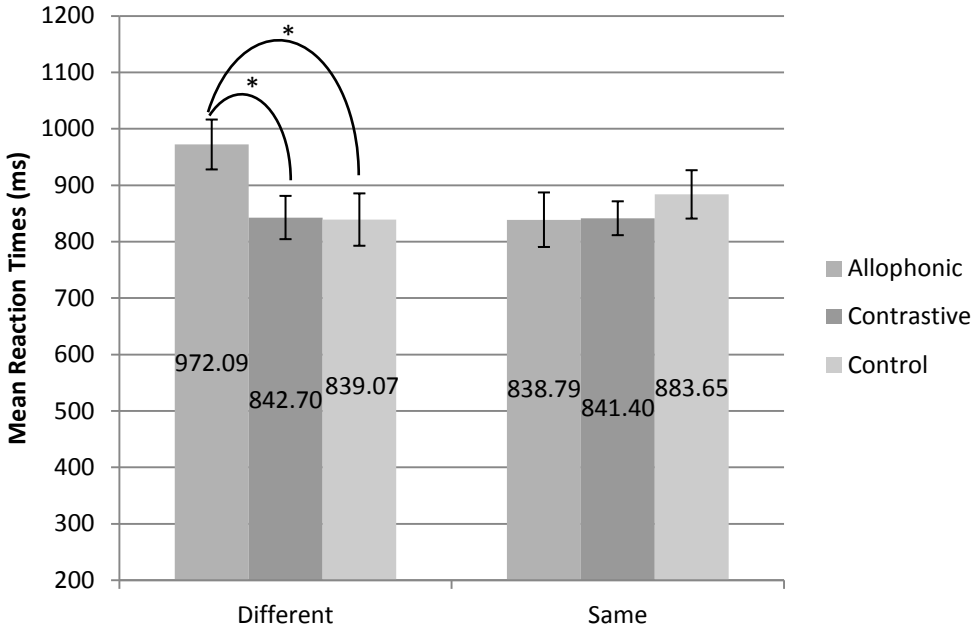


Figure 5.3. Experiment 1: Mean Reaction Times by Relationship and Trial Type. *Note.* A "*" indicates a significant difference.

A 2x3 repeated measures ANOVA indicated a main effect of Relationship ($F(2, 38) = 6.02$, $p = .005$), and a significant interaction between Relationship and Trial Type ($F(2, 38) = 11.03$, $p < .001$). Due to this interaction, Different and Same pairs were subsequently analyzed separately.

Among Different pairs, a main effect of Relationship was found ($F(2, 38) = 15.5$, $p < .001$), with pairs in the Allophonic condition yielding significantly slower RTs, but with no significant difference between Contrastive pairs and Control pairs. Pairwise comparisons are provided in Table 5.5. Among the same pairs, there were no significant differences ($F(2,38) = 1.73$, $p = .191$).

Table 5.5
Experiment 1: Effect of Relationship on Reaction Times

Trial Type	Relationship		Mean Difference	Sig.
Different	Allophonic	Contrastive	129.39*	< .001
		Control	133.02*	.002
	Contrastive	Control	3.62	1.00
Same	Allophonic	Contrastive	-2.61	1.00
		Control	-44.86	.315
	Contrastive	Control	-42.25	.307

Note. A "*" indicates a statistically significant result.

5.6.1.4. Experiment 1: Reaction Times – Brief Discussion

In Prediction 1, it was hypothesized that if the nature of the relationship between stimuli determined how they were perceived, the Allophonic stimuli would yield slower RTs than either the Contrastive or Control stimuli, and that Contrastive and Control stimuli should yield roughly equivalent results since both consist of stimuli in contrastive relationships (the difference being that the stimuli in the Contrastive condition contain consonants that also appear in the Allophonic condition). However, if there were differences between Allophonic, Contrastive and Control conditions (Prediction 2), this would suggest that participants were likely relying more on a lower-level acoustic mode of perception, since any differences in acoustics would yield different results. Prediction 1 was borne out with the (Different) Allophonic stimuli (i.e. [ts]-[t]) yielding significantly slower RTs compared to the other conditions, and with (Different) Contrastive and Control stimuli yielding almost exactly identical results.

5.7. Experiment 2 – 4IAX Discrimination Task

5.7.1. Procedure: Experiment 2 – 4IAX Discrimination Task

Participants were presented one quad at a time and instructed to find the "same" pair(s) in the quad. The stimuli were set up so that within the quad, either (a) the first pair contained syllables that were the same (e.g. AA-BA), (b) the second pair contained syllables that were the same (e.g. BA-BB), or (c) both pairs contained syllables that were the same (e.g. AA-BB) (see Table 5.4 for all possible permutations of the three response options). ISI within a pair was set to 1500 ms and between pairs was set to 2000 ms. Participants pressed [z], [/] and [spacebar] respectively, according to what they heard. A standard 4IAX task is designed so that only the first two options are available; in this case, however, the third option was incorporated in the hopes of preventing participants from simply focusing on a single pair to choose their response by process of elimination. For example, if the first pair has different stimuli, the second pair would necessarily contain same stimuli, even if participants did not listen to it, and *vice versa*. Thus, participants would only have to pay attention to the first pair, though some participants would potentially not respond until later in order to be sure of their answer, creating discrepancies in RTs. Unfortunately, upon examining the results with the three-option design used here, it became clear the fourth logical possibility (i.e. both pairs containing different syllables) should have been included. As discussed below, participants had inordinately fast reaction times for all quads that began with a Different pair because participants knew the answer as soon as the first pair was different, thus creating a response bias. This experimental bias was corrected in Experiment 4 (Chapter 8).

Experiment 2 began with 6 practice trials where the answers did not count towards accuracy rates or reaction times. This was followed by three blocks of 60 trials each, making the total number of trials 180, with a self-timed pause between each of the blocks and following the practice questions. Unlike the AX task, the stimuli for the 4IAX task were not randomized by the PsyScope program itself, but a list was compiled of all stimuli and then assigned a random number in Excel. This was done so that a control pair would not be paired with a test pair by PsyScope. Thus, all participants heard a quasi-randomized list. There were two groups: one half of participants heard a voiced version of the experiment, and the other, a voiceless version, since mixing voiced and voiceless trials would have made the experiment too long to do in one sitting. 90 trials consisted of “test” quads, 90 of “control” quads. Possible answers, however, were not equally divided: there were 72 “first pair same” answers, 72 “second pair same” answers, and 36 “both pairs same” answers. This was done so that all participants heard each possible stimulus quad only once. This task took approximately 40 minutes to complete.

5.7.1.2 Experiment 2: Accuracy Results

Figure 5.4 provides the mean accuracy scores for the 4IAX task. The closer the accuracy score is to 1, the more correct responses were given.

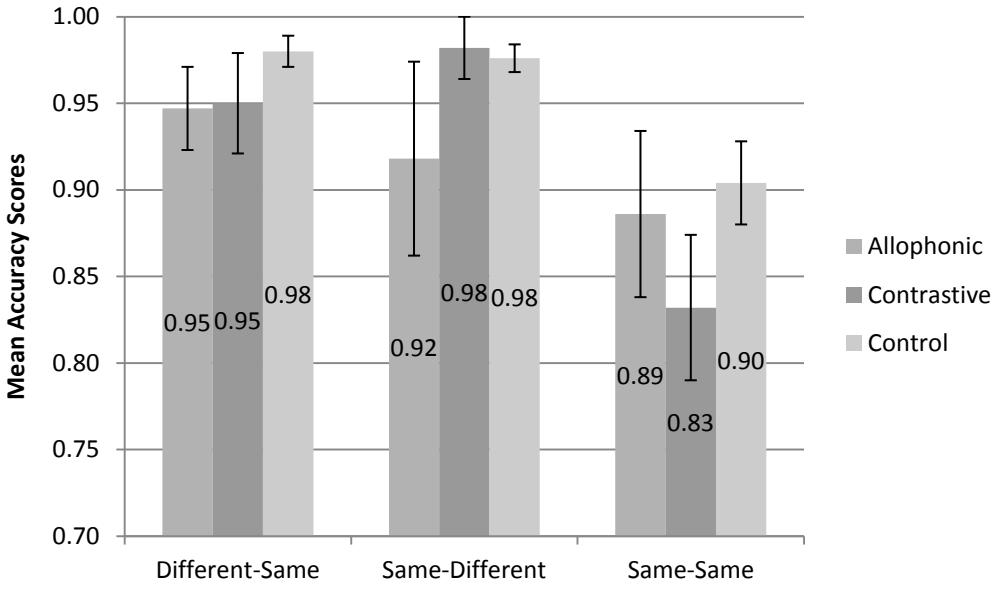


Figure 5.4. Experiment 2: Mean Accuracy Scores by Trial Type and Relationship.

A 3 x 3 repeated measures ANOVA was done on Relationship (Allophonic, Contrastive, Control) and Trial Type (Different-Same, Same-Different, Same-Same). The assumption of sphericity was violated for Trial Type and the interaction of Trial Type and Relationship, and so Greenhouse-Geisser adjusted values were used. A main effect of Trial Type was found ($F(1.18, 21.16) = 6.98$, $p = .012$), with Same-Same trials yielding significantly lower accuracy scores that Different-Same trials ($p=.026$) and a trend for lower accuracy scores than Same-Different trials ($p=.06$). There was no significant effect of Relationship ($F(2, 36) = 1.02$, $p = .37$) and no interaction between factors ($F(2.43, 43.72) = 1.37$, $p = .27$).

5.7.1.2 Experiment 2: Accuracy – Brief Discussion

The lack of significant differences between Allophonic and Contrastive conditions suggests that, although the Allophonic condition yielded lower absolute scores in the Same-Different trial type, a 4IAX task may not be sensitive enough to

elicit differences in accuracy; however, as discussed below, this does not mean that listeners are treating all conditions as equal. As compared to the results from the AX task (Experiment 1), the results match, with lower accuracy scores in Different trials for the Allophonic condition and slightly lower accuracy scores for Contrastive stimuli in Same stimuli.

The reason for Same-Same trials is mostly due to two participants obtaining particularly low accuracy scores for the Same-Same Allophonic condition (50% and 33%). These participants were not eliminated from the overall results because their average scores were within the normal range, but due to the lower power of this experiment, these two scores had a greater effect on the overall results.

5.7.1.3 Experiment 2: Reaction Time Results

There were 40 time-outs and 42 incorrect responses not included in the analysis for RTs. Of the 1372 correctly-answered responses, the data was trimmed by any result that fell 3 standard deviations above or below the mean for each condition (N=77, or 5.95%).

Due to a bias created by the response options (whereby participants were asked to identify whether it was the first pair of a quad that consisted of same stimuli, the second pair, or both pairs), Different-Same quads yielded much faster RTs than the other trial types. As soon as the first pair was heard by participants, they were able to answer without listening to the rest of the trial, while for the Same-Different and Same-Same quads, participants had to listen all the way to the end of

the trial. This is visible in Figure 5.5 which provides the mean reaction times by Trial Type and Relationship.

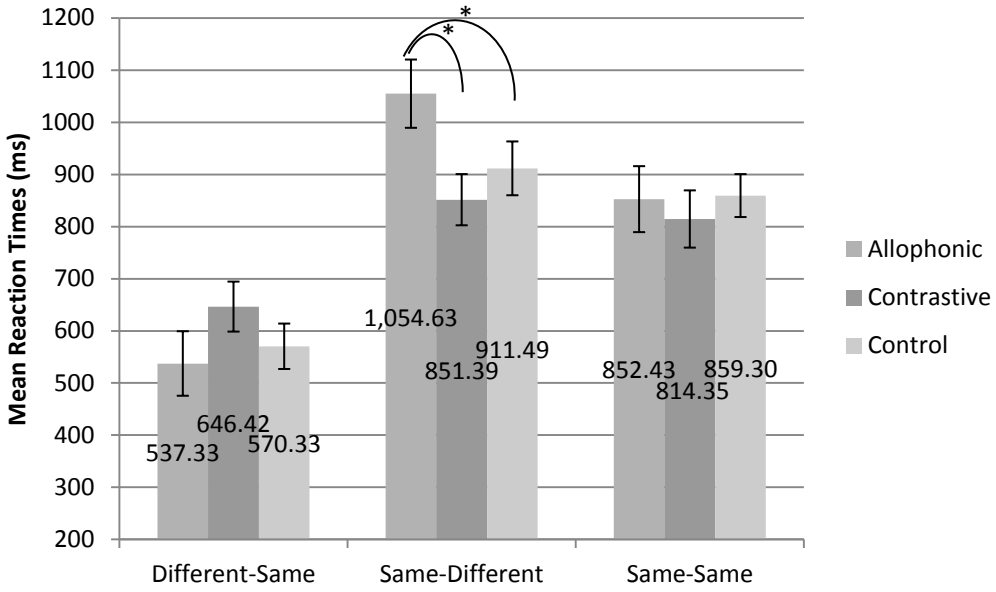


Figure 5.5. Experiment 2: Reaction Times by Trial Type and Relationship.
Note: An arc with a "*" symbol indicates a significant difference. An arc without the "*" indicates a trend.

A 3 x 3 repeated measures ANOVA was done on Trial Type and Relationship. The assumption of sphericity was violated for the interaction of Trial Type and Relationship and so Greenhouse-Geisser adjusted values were used. As expected, a significant interaction of Trial Type and Relationship was found ($F(2.91, 52.39) = 6.25$, $p = .001$) and so each trial type was subsequently analyzed separately.

For Different-Same quads, a main effect of Relationship was found ($F(2,35) = 3.87$, $p = .03$). The pairwise comparisons for Different-Same quads are provided in Table 5.6. None of the differences between conditions reached significance, but a trend was found between Allophonic and Contrastive quads, with RTs for Allophonic quads being faster than for Contrastive quads.

Table 5.6
Experiment 2: Pairwise Comparisons by Relationship for Different-Same Quads

Relationship (a)	Relationship (b)	Mean Difference	Sig.
Allophonic	Contrastive	-109.09	0.075
	Control	-33.01	1.00
Contrastive	Control	76.08	0.197

For Same-Different quads, the assumption of sphericity was violated and so Greenhouse-Geisser adjusted values were used. A main effect of Relationship was found (F(1.54, 27.72) = 11.46, p = .001). Pairwise comparisons, provided in Table 5.7, show that the Allophonic condition yielded significantly slower RTs and both the Contrastive and the Control quads. There was no significant difference between Contrastive and Control conditions.

Table 5.7
Experiment 2: Pairwise Comparisons by Relationship for Same-Different Quads

Relationship (a)	Relationship (b)	Mean Difference	Sig.
Allophonic	Contrastive	203.24*	0.004
	Control	143.15*	0.003
Contrastive	Control	-60.10	0.400

For Same-Same quads, no main effect of Relationship was found (F(2, 36) = .455, p = .64).

5.7.1.4 Experiment 2: Reaction Times – Brief Discussion

For this 4IAX task, it was predicted that if phonological relationship determined participants' results, stimuli in an allophonic relationship would yield lower accuracy scores and slower RTs than Contrastive stimuli, which in turn should yield comparable results to the Control condition since both are comprised of contrastive

phones (Prediction 1); if acoustics determined results, then stimuli that was more acoustically similar (e.g. allophones [t] and [t^{s}]) would yield lower accuracy scores and slower RTs, while there should be differences between the Contrastive and Control stimuli since they are comprised of different types of stimuli (Prediction 2).

It was found that with regard to accuracy scores, there were no significant differences. This may be interpreted either as the task not being sensitive enough to elicit differences with regards to this measure, or simply that participants had no significant difficulties completing the task. Neither prediction 1 or 2 were borne out in this respect.

With regard to RTs, no differences were found when both quads consisted of Same stimuli. This is expected since when all stimuli are the same, there is in essence an absence of relationship between stimuli. One would expect RTs to be the same across conditions. For Different-Same stimuli, there was a trend for Allophonic quads to yield faster RTs than Contrastive stimuli, which was not expected since allophonic stimuli in Experiment 1 were shown to yield slower RTs, and there was no difference between Allophonic quads and Control quads, also contrary to the findings in Experiment 1. Since RTs were measured as of the 2nd stimulus, and given the response bias introduced for Different-Same stimuli, the variability could be random in that some participants may simply have waited longer to be sure of their responses some of the time but not all of the time, since they had the potential to respond immediately after the first pair in a quad. These results are therefore likely not representative of the influence of any particular factor on perception.

For Same-Different pairs, Prediction 1 was borne out, with there being no significant difference between Contrastive and Control quads, and Allophonic quads yielding significantly slower RTs than the other two conditions. This appears to provide support for Relationship determining how participants perceived the stimuli.

5.8 Experiments 1 and 2: Discussion

When results from experiments 1 and 2 are taken together, the results consistently show that stimuli in an allophonic relationship are more difficult to process than stimuli in a contrastive relationship. Results for these experiments can be summarized as in Table 5.8:

Table 5.8
Experiments 1 and 2: Summary of Results

Experiment	Results	Prediction 1: Relationship C=Ctrl>A	Prediction 2: Acoustics C≠Ctrl>A
1. AX	Accuracy	x	x
	RTs	✓	x
2. 4IAX	Accuracy	x	x
	RTs (Same-Diff.)	✓	x

Note. "C" = "Contrastive", "Ctrl" = "Control", "A" = "Allophonic."

These results are not simply due to the acoustic properties of the stimuli chosen for the Contrastive condition since the same results are obtained for the Control condition which also contains contrastive stimuli, and there are no differences between the two conditions. There is strong evidence for the broader phonological relationships of allophony and contrast. However, it is possible that these are simply two extremes of a single continuum, and that intermediate levels of

contrast will yield differences in results. Experiments 1 and 2 confirm what has been found in previous literature, namely that there is behavioural evidence for a binary view of contrast if the experiments are designed as such. Other factors need to be taken into account if intermediate levels of phonological relationship are to be tested. The following chapters outline Experiments 3, 4 and 5 which were designed so as to test an intermediate level of contrast between purely allophonic and purely contrastive. The next chapter provides details regarding how intermediate levels of contrast may be quantified and tested, and how this was incorporated into the experimental design for Experiments 3, 4 and 5. Results for these experiments are discussed in detail in chapters 7, 8 and 9.

Chapter 6 Experimental Criteria

Given the theoretical approaches to determining phonological relationships discussed in Chapter 2 and the methodological considerations discussed in Chapters 3 and 4, this chapter aims to show how a gradient view of contrast can be translated into concrete measures and experimental factors. Criteria for the selection of experimental stimuli are discussed here, namely: the acoustic similarity between two sounds, the number of minimal pairs that two sounds participate in (i.e. the functional load), and the predictability of distribution of pairs of sounds.

6.1. Quantifying Phonological Criteria for Contrast

As discussed in Chapter 2, the main criteria for determining the phonological relationship between two sounds are (1) whether the two sounds serve to differentiate between lexemes, and (2) the predictability of their distributions. These are not, however, static criteria to be viewed in binary all-or-nothing terms, but rather as gradient and quantifiable criteria. Due to the gradient nature of these criteria, the relationships they define are also expected to be gradient, which will be tested in a series of experiments.

There are only a few attempts in the literature to quantify and test the above criteria experimentally. With regards to the second criterion, Hall (2009) noted that the probability of distribution of two sounds is a quantifiable, scalar measure, rather than ‘predictable’ or ‘not predictable.’ She defines relationships between segments by the amount of uncertainty, or “entropy,” between them, and defines high entropy sequences as being less predictable (contrastive) and low-entropy sequences as

being more predictable (allophonic). The highest certainty/predictability is found when one sound cannot occur in a particular environment (probability is zero), and the least certainty is found when two sounds are allowed to occur in a particular environment (as with segments in minimal pairs). While she focuses on predictability of distribution as the central criterion of her model, she also acknowledges the existence of other criteria such as lexical distinction, though she does not quantify or incorporate this into her model.

One of Hall's (2009) main predictions is that high entropy pairs of sounds (i.e. more uncertain and with a higher degree of contrast) will be perceived as less similar. To test this, she used a similarity-rating task to explore the perception of German consonantal allophones in CV and VC syllables. Hall predicted that segments that are allophones of each other – and therefore predictably distributed and low entropy – would be perceived as more similar than two contrastive sounds. She tested the correlation between the averaged normalized similarity rating within each pair of segments, and type and token entropy across vowel and syllable environments. Results showed that there was no indication that entropy (as defined by Hall) was a predictor of perceived similarity. The lack of a result could be due to reasons related to: (a) the experimental methodology and/or design[10]; (b) the possibility that there is no relationship between entropy and perceived similarity between segments and, by extension of the interpretation of the similarity rating task, no relationship between entropy and phonological contrast; or (c) that the effect of entropy was obscured by another factor. Despite the lack of statistically

[10] Hall (2009) provides a list in critical detail to benefit future research.

significant results, in order to compare predictions based on the criterion of predictability of distribution to those based on lexical distinction, the predictability of distribution of the vowel pairs used in Experiments 3 to 5 were calculated, discussed below.

As discussed in Chapter 2, functional load is a measure by which the work of particular contrasts may be quantified, and this is the means by which the other most common criterion for contrast – lexical distinction – was situated on a scale from Low to High contrast. Both measures are discussed in more detail in the following section, as well as the role of frequency in more general terms.

6.2. Functional Load: Minimal Pair Counts

While the existence of a minimal pair has traditionally been accepted as sufficient evidence of contrast between two phonemes (as discussed in Chapter 2), the number of minimal pairs shared by two sounds can make a difference to the robustness of the contrast.[11] The number of minimal pairs distinguished by a given contrast is a measure of the functional load of a contrast: phonemes that distinguish many lexemes have a high functional load, and conversely, phonemes that distinguish few lexical items have a low functional load. (Wedel, Kaplan, & Jackson, 2012). Wedel et al. (2012) looked at multiple corpora and phonological systems of several languages, including English (Received Pronunciation and American),

[11] The term "minimal pair" is very close in meaning to "word neighbour," which is defined as a word that differs from another word by one phoneme substitution, deletion or addition in any position (Vitevitch & Luce, 1999). In this thesis, the difference between "word neighbours" and "minimal pairs" is that in my calculations, deletions and additions were not counted as minimal pairs. Furthermore, the difference between words in a minimal pair may be that of one phone, not necessarily one phoneme, since counts were calculated based on phonetic transcriptions.

German, Dutch, French, Spanish, Slovak, Korean, and Cantonese, and found that phonemes with a high functional load are less likely to merge[12] over time, or in other words, are more likely to maintain their contrast. In their study, the number of minimal pairs was found to be a significant predictive factor of merger (i.e. loss of contrast) and where two phonemes shared only few minimal pairs, relative phoneme frequency became a significant predictor of merger.[13] The fact that these two measures are correlated with maintenance or loss of contrast suggests that both the number of minimal pairs and relative phoneme frequency are reliable indicators of the robustness of contrast, and also suggests that not all contrasts are of equivalent value to a given phonological system. For example, see work that shows that the number of minimal pair contrasts also varies by position in a word (Zamuner, 2009; Zamuner, Morin-Lessard & Bouchat-Laird, 2014; de Bree, Zamuner & Wijnen, 2013). Since minimal pair count and relative phoneme frequency were found to be significantly correlated to contrast, these measures in particular will be used as measures of contrast.

6.2.1. Multiple Minimal Pair Counts

The fact that a high functional load is correlated with maintaining contrasts as found in Wedel et al. (2012) suggests that functional load is also a reliable measure of contrast between two phonemes. Since, in the strictest interpretation of allophony, two allophones should never occur in the same environment, they should never

[12] A phonological merger is when two contrastive phonemes become one; for example, "cot" and "caught" were historically (and synchronically in certain dialects) produced with two different vowels but are now homophonous in many regions of the US and Canada (Wedel et al., 2012).

[13] As part of their definition of "merger," they also included cases of neutralization since this is a context-sensitive merger that eliminates contrast in specific phonological environments.

participate in a minimal pair with each other; however, it is quite common for one of the two allophones to participate in multiple minimal pairs – just perhaps not with its related allophone(s). For example, in LF, [y] and [ʏ] are in an allophonic relationship, with [y] occurring in open syllables and closed syllables before [r/ʁ], [v], [z,] and [ʒ], and [ʏ] occurring in word-final closed syllables ending with other consonants.[14] There are also no minimal pairs between [y] and [ʏ]. However, [y] contrasts with other vowels (e.g. [ly], [li], [la], [lu], etc.). These other minimal pairs contribute to the strength of [y]'s functional load), as do other words that constitute a minimal pair with [ʏ]. In the present study, two types of minimal pair count were taken into account when selecting stimuli to be tested: one count for a pair of vowels, such as [y] and [ʏ], and another count to account for how many minimal pairs a given vowel partakes in with other sounds in the inventory.

6.2.2. Calculating Minimal Pair Counts

The OMNILEX database (Desrochers, 2006) was used to establish a word list of French one-syllable words of CV (N=1032), VC (N=376), CVC (N=2910) and CCV (N=230) syllable structures (total words = 4548). The database includes approximately 102,000 lexical entries originating from multiple French dictionaries and the Lexique corpus (New, Pallier, Ferrand, & Matos, 2004). Since the phonetic transcriptions were done based on European French, the word lists were re-transcribed to reflect a standard LF pronunciation by a native speaker and expert in LF phonetics and phonology, paying particular attention to vowels [ɪ], [ʏ], [ʊ] and

[14] However, the quality of the vowel can be either tense or lax before [r/ʁ] (Côté, 2012).

[ɜ][15] because these vowels do not occur in European French. Due to the size of the corpus, it was not feasible re-transcribe the entire database, but future studies could be enlarged to include more syllable types.

While minimal pair counts and phonological neighbourhood density values were provided in the OMNILEX database, these could not be used for LF due to the EF transcriptions. The number of minimal pairs for each of 16 vowels[16] was therefore calculated anew with the LF transcriptions by counting, for example, how many times a given vowel occurred after [b] in a CV syllable, then how many times that vowel occurred after [d], and so on, for every consonant and vowel combination in CV, VC, CVC and CCV syllables. Vowels were compared across consonantal contexts in order to determine how many vowels occurred in exactly the same phonetic context (in other words, to determine how many minimal pairs existed implicating a single consonantal context for all the vowels). From this, it was calculated (a) how many minimal pairs a *single vowel* participated in with all other vowels (referred to as "individual count" below[17]), and (b) how many minimal pairs existed between *two specific vowels* (referred to as "shared count"). This method of calculating minimal pairs was developed based on Brown (1988). Table 6.1 below summarizes these results. The vowels are in IPA. I have chosen to transcribe the [ə] in the OMNILEX corpus as [ə]. However, Martin (1998) and Séguin (2010) have shown that these two vowels are systematically produced and perceived as [œ]. For reasons of consistency, I have chosen not to merge the schwa with [œ].

[15] [ɜ] was not used in the experimental stimuli, but for the sake of calculating minimal pairs, this vowel was transcribed for words such as "party" *fête* [fɜt] vs. "done.fem." *faite* [fɛt].
[16] Nasal vowels and diphthongs were not included. No consonants were excluded.
[17] This is referred to as "relative contrastiveness" in Hume et al., 2013.

Table 6.1

Minimal Pair Counts by Vowel in Laurentian French for CV, VC, CVC and CCV Syllables

Vs	a	ɒ	e	ə	ɛ	ø	o	ɔ	œ	i	u	y	ɜ	ɪ	ʊ	ʏ
a	X	32	16	9	70	21	55	86	20	36	32	29	24	83	62	41
ɒ	32	X	11	6	25	11	29	21	3	20	15	12	10	25	19	11
e	16	11	X	9	18	15	21	2	1	19	18	16	0	3	1	2
ə	9	6	9	X	8	7	8	1	0	9	9	9	0	0	0	0
ɛ	70	25	18	8	X	16	54	52	8	25	21	17	8	58	36	26
ø	21	11	15	7	16	X	19	8	2	17	16	15	2	8	2	3
o	55	29	21	8	54	19	X	34	8	25	23	20	13	35	22	12
ɔ	86	21	2	1	52	8	34	X	17	16	14	9	20	68	55	38
œ	20	3	1	0	8	2	8	17	X	9	8	7	7	15	17	8
i	36	20	19	9	25	17	25	16	9	X	31	29	14	1	8	7
u	32	15	18	9	21	16	23	14	8	31	X	24	10	12	3	8
y	29	12	16	9	17	15	20	9	7	29	24	X	8	6	6	0
ɜ	24	10	0	0	8	2	13	20	7	14	10	8	X	21	12	13
ɪ	83	25	3	0	58	8	35	68	15	1	12	6	21	X	58	44
ʊ	62	19	1	0	36	2	22	55	17	8	3	6	12	58	X	32
ʏ	41	11	2	0	26	3	12	38	8	7	8	0	13	44	32	X
Total	**616**	**250**	**152**	**75**	**442**	**162**	**378**	**441**	**130**	**266**	**244**	**207**	**162**	**437**	**333**	**245**

Note. "Shared counts" are numbers in single cells (e.g. [a] with [o] = 55 minimal pairs); "individual counts" are the totals of each column in the bottom line (e.g. [a] participates in 616 one-syllable minimal pairs).

A cell in Table 6.1. can be read as the "shared count" of minimal pairs for the vowels in a given column and row; for example, [a] and [ɒ] share 32 minimal pairs (among one-syllable words in OMNILEX); [a] and [ɔ] share 86, etc. The "Total" at the bottom of each column indicates the "individual count," which is how many minimal pairs a given vowel participates in with the 15 other vowels. For example, [o] participates in 378 minimal pairs in total with all other vowels, while [ə] only participates in 75.

In order to represent a scale of contrast, vowels were chosen from the high-end, middle, and low-end range of minimal pair counts, both in terms of individual vowel counts as well as shared counts. The final selection of vowels used in stimuli and their corresponding minimal pair counts are provided in Table. 6.2. "Individual

Count" indicates how many minimal pairs a given vowel participates in with all other vowels in LF; "Shared Count" indicates how many minimal pairs exist that differ only by the two vowels in a given contrast level.

Table 6.2.
Individual and Paired Vowels - Minimal Pair Counts

Contrast Level	**Individual Count**		**Shared Count**
High	a	616	86
	ɔ	441	
Mid	o	378	22
	ʊ	333	
Low	y	207	0
	ʏ	245	

The designation of "High," "Mid," and "Low" Contrast were roughly based on the upper, mid and bottom third of the numbers of minimal pairs for vowels; however, the difference between "High" and "Mid" in terms of minimal pairs is not proportionate to the difference between "Mid" and "Low". There was a necessary trade-off between choosing tokens that matched equally well for number of minimal pairs and acoustic similarity (described in section 6.6.1), since some vowels did match better for number of minimal pairs but were so acoustically different from one another that there would be no challenge in a discrimination task with such stimuli. In addition, the Low contrast pair [y]-[ʏ] is allophonic in LF – which also happen to share no minimal pairs – while the other two pairs contrast in some contexts; for example, "gamme" [gam] 'scale' versus "gomme" [gɔm] 'eraser'; "role" [ʁol] 'role' versus "roule" [ʁʊl] 'roll.3p.sg.' This will be important when examining the binary view of contrast, discussed in the predictions below (§6.7).

In order to verify the above calculations of functional load, the same sub-set of the OMNILEX corpus was entered into the *Phonological Corpus Tools* (PCT) software developed by Hall, Blake, Allen, Fry, Mackie, & McAuliffe (2015). In terms of what I have called the "shared count" (i.e. the raw number of minimal pairs for two given sounds), the results were comparable, with 0 minimal pairs for [y]-[ʏ]; 19 minimal pairs for [o]-[ʊ]; and 84 minimal pairs for [a-ɔ].[18] Since the sub-set of the OMNILEX corpus contains 1262 open syllables (CV (N=1032) + CCV (N=230)) and 3286 closed syllables (VC (N=376) + CVC (N=2910)), one might expect there to be a bias towards lax vowels since they do not occur in open syllables; however, when CCVC syllables are added to the calculation by PCT, the number of minimal pairs for [y]-[ʏ] remains at 0; the number of minimal pairs for [o]-[ʊ] increases to 22 from 19; and the number of minimal pairs for [a-ɔ] increases to 112 from 84. The effect on functional load for including CCVC syllables would therefore not change the relative ranking of Low, Mid and High pairs, except perhaps to further separate the High pair from the other pairs.

6.3. Predictability of Distribution

Predictability of distribution of the above vowel pairs was calculated using the *Phonological Corpus Tools* software developed by Hall et al. (2015). Using the algorithms from Hall (2009), predictability of distribution was calculated in terms of entropy (uncertainty), where a value of '0' indicates complementary distribution and '1' indicates perfect contrast. The frequency of the vowels occurring in a particular environment weight their contribution to the distribution in the calculation so that

[18] The difference in minimal pairs between my calculations and the calculations done by PCT are likely due to human error since some of the calculations and counts were done manually in Excel.

vowels occurring frequently in a particular environment contribute more to the entropy value. The environments included in the calculation were: before a consonant that was [ʁ, v, z] or [ʒ]; before a consonant that was not [ʁ, v, z] or [ʒ]; and word-finally. The entropy for the stimuli pairs is provided in Table 6.3 for the vowels in question (a) before a consonant that is either [ʁ, v, z] or [ʒ], (b) before a consonant that is not [ʁ, v, z] or [ʒ], and (c) before the end of a word (i.e. in an open syllable). The average entropy for each vowel pair is provided in the right-most column. (The number of minimal pairs ("shared count") for each vowel pair as calculated by PCT is included for ease of comparison.)

Table 6.3
Predictability of Distribution (ProD) for Stimuli Pairs

Contrast Level	Vowels	Shared Count	ProD _C, C= [ʁ, v, z, ʒ]	ProD _C, C ≠ [ʁ, v, z, ʒ]	ProD _#	Avg. ProD
High	a ɔ	84	0.985	0.986	0.000	0.915
Mid	o ʊ	19	0.000	0.996	0.000	0.654
Low	y ʏ	0	0.000	0.000	0.000	0.000

Note that the predictability of distribution for the Low pair is 0 because the frequency of one of the two sounds is always 0 for each of the listed environments. Overall, the entropy values for predictability of distribution place the vowel pairs in the same categories of High, Mid and Low contrast, but with the slight difference that by the shared minimal pair count, Mid pairs pattern more closely with Low pairs, while by predictability of distribution, Mid pairs are closer to High pairs. It is not possible to evaluate if these differences are significant, but it is possible that, since the categories of contrast according to both measures align the same way, this could be a potential confound: it will not necessarily be possible to determine whether results would reflect the functional load or the predictability of distribution of the chosen

vowels, nor would it necessarily be possible to tease apart an interaction of these criteria since they would seem to classify the selected vowel pairs in the same way.

6.4. Frequency

Phonotactic probability, which takes multiple frequency measures into account (such as, for example, single phoneme frequency, bi-phone frequency, transitional frequency, etc.), has been found to influence processing in a number of ways. For example, real words with dense similarity neighbourhoods (which are also words with many minimal pairs) have been shown to have slower processing times due to lexical competition, while in non-words, phonotactic probability plays a greater role, with high phonotactic probability non-words yielding faster reaction times (Vitevitch & Luce, 1998; 1999). There is also a correlation between frequency and the neighbourhood density such that high-probability segments are found in high-density neighbourhoods and low-probability segments are found in low-density neighbourhoods, because the more frequent a sound is, the more chances it has to occur in a minimal pair. The relative frequency of a sound is therefore a factor that was controlled for.

Given the findings by Vitevitch and Luce (1998,1999) the type frequency of the vowels in LF were taken into account based on the same selections from the OMNILEX database as described above (i.e. from CV, VC, CCV and CVC syllables), calculating the number of times a vowel occurred in the various syllables types. Since there is no corpus with lexical frequencies for LF phones and it was not feasible to re-transcribe the entire OMNILEX database, frequency was only calculated within the aforementioned syllables types, such that the maximum

number of times [a] could occur includes [ba], [pa], [ta], [la], etc., for all consonants that occur in CV syllables in LF; then added to this was the number of times [a] could occur in VC syllables, in CCV syllables, and finally, in CVC syllables. Table 6.4 presents the frequency of vowels totalled across all one-syllable words with CV, VC, CCV and CVC syllable structures.

Table 6.4
Least to Most Frequent Vowels in One-Syllable Words in LF (IPA)

Min-Max	**ɘ**	**œ**	**ø**	**e**	**y**	**ɒ**	**ʏ**	**ɜ**	**u**	**i**	**ʊ**	**o**	**ɔ**	**ɪ**	**ɛ**	**a**
	13	40	41	52	63	76	76	79	84	106	110	154	202	206	238	305

The vowels chosen according to minimal pair counts (Table 6.2) are again provided in Table 6.5 with their corresponding type frequency in the OMNILEX corpus in the same syllable types as above.

Table 6.5
Vowels by Minimal Pair Counts and Frequency

Contrast Level	**Individual Count**		**Shared Count**	**Frequency**
High	a	616	86	305
	ɔ	441		202
Mid	o	378	22	154
	ʊ	333		110
Low	y	207	0	63
	ʏ	245		76

The high-contrast vowel pair [a-ɔ] consists of more minimal pairs and high type-frequency vowels; the low-contrast vowel pair [y-ʏ] consists of no minimal pairs and low type-frequency vowels; the mid-level contrast vowel pair [o-ʊ] falls between the other two levels in both minimal pair count and type frequency. Vowels were matched as best as possible for these counts while taking acoustic similarity into consideration. This was done based on knowledge of the typical phonetic properties of the chosen vowels, and verified with acoustic analyses, provided below in §6.5.1.

In summary, the number of minimal pairs and the relative type frequency have been shown to be correlated to robustness of contrast and speed of processing (Wedel et al., 2012, Vitevitch & Luce, 1999). These were therefore calculated for the vowels of LF for CV, VC, CVC and CCV words. Pairs of vowels that best matched each other for these values were classified as “High,” “Mid” or “Low” on a scale of contrast. An estimate of acoustic similarity was also taken into account when choosing the stimuli so that the vowels roughly matched in this aspect so as not introduce a confound with the other measures and thus inadvertently favour one condition over the other. In other words, this was done to avoid having the High pair of vowels be maximally acoustically different compared to the pair of Low vowels. The next section details how the stimuli were created, and section 6.5 provides detailed acoustic analyses for the stimuli recorded and used in the subsequent experiments.

6.5. Stimuli Creation Process

Stimuli were produced by a trained male phonetician who is a native speaker of LF from Quebec City. This dialect in particular was chosen so that participants from the Ottawa-Gatineau region would be, as much as possible, equally biased for any dialectal differences, since choosing a speaker from either Gatineau or Ottawa would favour listeners of those dialects. Vowels were produced in CVC frames consisting of [l, b, f] and [ʃ] so that all were non-words of French, except for the proper name [bɔb] “Bob” and the word [lɔl] which has been borrowed from the English acronym meaning “laughing out loud” (or “lol”) in cyberspeak.

Stimuli were recorded in a sound-attenuated booth with a Shure “microflex” omnidirectional condenser boundary microphone (model MX392/0) on a Marantz digital recorder and subsequently normalized in Praat (version 5.3.79; Boersma & Weenik, 2014) for amplitude (70 dB) and intonational curve so that participants would not be able to use these cues to distinguish between stimuli. Vowel and consonant length were not manipulated since this could affect the recognisability of the vowels and perceived naturalness. Tokens in Same pairs were always acoustically different and, as much as possible, tokens in both Same and Different pairs were matched for intonation curve. In a few cases, the intonation curve was manipulated synthetically.

6.6. Acoustic Measures

Figure 6.1 shows first and second formant (F1 and F2) measurements for the vowels in the stimuli used in this study. Measurements were taken in Praat from a steady-state portion of the vowel as close as possible to the mid-point.

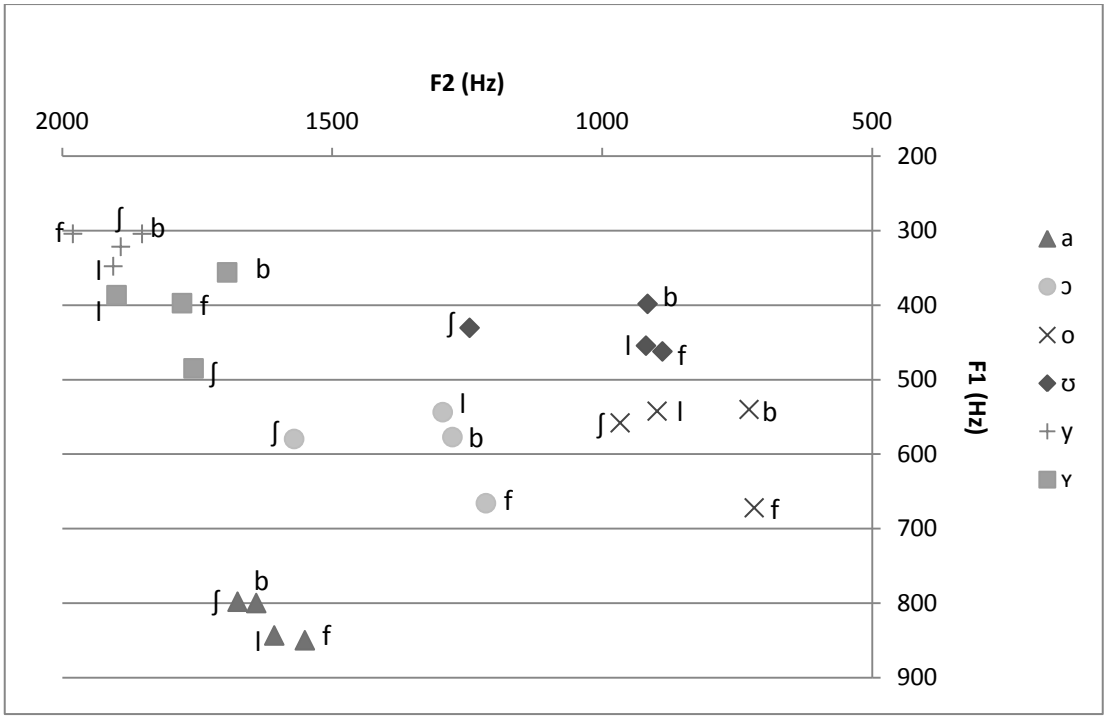


Figure 6.1. First and Second Formant Values of Vowels in CVC Stimuli Frames

Table 6.6 provides the F1, F2, F3 and duration (in seconds) values plotted in Figure 6.1 by stimulus.

Table 6.6
F1, F2, F3 (Hz) and Duration (seconds) Values by Stimulus

Stim.	F1	F2	F3	Dur.	Stim.	F1	F2	F3	Dur.
ʃʊʃ	430.50	1245.57	2768.30	0.205	bʊb	398.36	915.94	2504.40	0.128
ʃʏʃ	484.95	1756.74	2540.34	0.176	bʏb	356.09	1694.55	2315.90	0.120
ʃaʃ	798.15	1675.32	2468.15	0.177	bab	799.92	1640.71	2573.01	0.090
ʃoʃ	558.06	966.77	2464.89	0.198	bob	540.05	728.03	2743.03	0.132
ʃyʃ	321.60	1891.62	2535.44	0.173	byb	304.34	1852.19	2586.67	0.139
ʃɔʃ	579.72	1570.53	2461.16	0.153	bɔb	577.18	1277.35	2546.67	0.130
Stim.	**F1**	**F2**	**F3**	**Dur.**	**Stim.**	**F1**	**F2**	**F3**	**Dur.**
fʊf	462.18	888.45	2679.88	0.193	lʊl	454.47	918.91	2421.90	0.137
fʏf	397.39	1778.16	2143.17	0.197	lʏl	386.68	1899.38	2199.56	0.140
faf	849.92	1550.68	2267.14	0.179	lal	843.46	1607.32	2454.92	0.119
fof	672.06	718.46	2497.94	0.162	lol	542.35	898.33	2373.87	0.145
fyf	304.25	1980.27	2316.06	0.213	lyl	347.86	1905.57	2481.78	0.147
fɔf	665.94	1215.33	2343.49	0.189	lɔl	543.72	1295.33	2402.16	0.144

F1 roughly corresponds to vowel/tongue height; F2 roughly corresponds to the anteriority of the tongue in the oral cavity; F3 roughly corresponds to rounding. Looking at Figure 6.1, there are a few cases which appear to be outliers for [ɔ], [o] and [ʊ]. These tokens produced after [ʃ] or [l], causing them to be produced more in front in the mouth, resulting in higher F2 values.

Overall, the values seen in Figure 6.1 are comparable to Martin's (2002: 84) vowel space for male speakers of the Quebec dialect in terms of relative position in the vowel space, reproduced here in Figure 6.2:

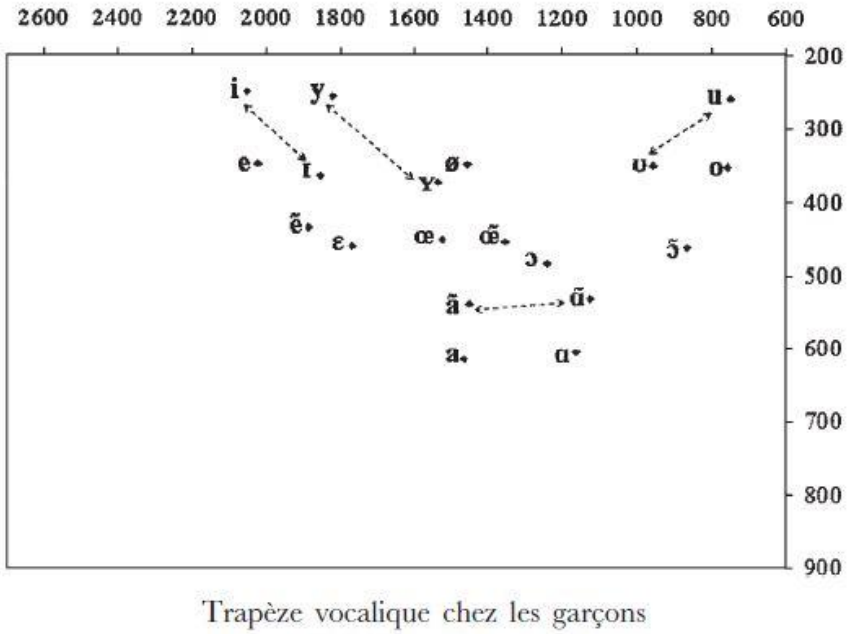


Figure 6.2. Martin's (2002) Vowel Space for Male Speakers from Quebec City

The fact that the relative position of the categories roughly match between this study and Martin's means that the stimuli participants heard are representative of the LF vowels selected, and no vowels in the stimuli are anomalies of their phonetic category or unrecognizable as a member of their category.

6.6.1. Acoustic Similarity

While stimuli were initially chosen based on estimated levels of contrast by minimal pair count and relative frequency of vowels, they were also chosen because they are roughly equally acoustically similar. If they were too different acoustically, it would be difficult to compare results across conditions since facility in perception might be attributed to greater acoustic difference rather than level of contrast. On the other hand, they should not be so similar that they are overly difficult to distinguish.

6.6.1.1. Acoustic Similarity Based on F1 and F2 values

There is more than one way of evaluating acoustic similarity. If listeners rely more on F1 and F2 values, then participants may perceive stimuli according to differences between these measures. Table 6.7 provides the differences between the F1 and F2 values averaged across consonantal frames for the recorded stimuli by vowel pair.

Table 6.7
Differences Between Mean F1, F2 (in Hz) and Duration (seconds) by Stimulus

Contrast	**Vowels**	**Mean F1**		**Difference**
High	[a-ɔ]	822.86	591.64	231.22
Mid	[o-ʊ]	578.13	436.38	141.75
Low	[ʏ-y]	406.28	319.51	86.77
Contrast	**Vowels**	**Mean F2**		**Difference**
High	[a-ɔ]	1618.508	1339.635	278.87
Mid	[o-ʊ]	827.8983	992.2199	164.32
Low	[ʏ-y]	1782.207	1907.413	125.21
Contrast	**Vowels**	**Mean Duration**		**Difference**
High	[a-ɔ]	0.166	0.148	0.0172
Mid	[o-ʊ]	0.221	0.140	0.0807
Low	[ʏ-y]	0.150	0.133	0.0166

Table 6.7 shows that the largest difference between vowels in terms of F1 and F2 is between the High Contrast pair [a-ɔ], followed by the Mid Contrast [o-ʊ] pair, followed by the Low Contrast [ʏ-y] pair. On the other hand, by duration, [o-ʊ] exhibits the largest difference while [a-ɔ] and [ʏ-y] are relatively similar. It is possible, then, that participants will perceive these vowels according to F1-F2 differences rather than other acoustic parameters such as those quantified in the following section using weighted acoustic differences, or may rely on other cues such as duration when responding in Experiments 3-5. In the case where formant values are relied upon, it would not be possible to draw a line between judgments based on F1-F2

values and strength of contrast as calculated by minimal pair counts and predictability of distribution. However, if participants maximize all information available, they may rely perform relatively well on stimuli containing [o-ʊ] due to the large difference in duration between these vowels. There are, however, other methods of determining acoustic similarity that yield different predictions.

6.6.1.2. Machine-Assisted Calculation of Acoustic Similarity

Another way of measuring acoustic similarity was developed by Mielke (2012), who used a phonetically-based metric by which to assess the similarity of sounds combining multiple sources of acoustic and articulatory data, including nasal and oral airflow, vocal fold activity, larynx height, and ultrasound video of the tongue and lips. Spectral information and vocal tract shape was also used to calculate phonetic distances between phones. In the present study, the acoustic distance was measured between the six vowels selected as stimuli ([a], [ɔ], [o], [ʊ], [y], [ʏ]) using the same methods as in Mielke (2012) developed for acoustic comparisons, and based on the stimuli recordings that participants listened to in this study. To do this the waveforms of the stimuli were converted into matrices of 12 Mel-frequency cepstral coefficients (MFCCs) in Praat, and then a dynamic time warping technique (DTW) was used to quantify acoustic similarity between vowels. Table 6.8 provides the weighted acoustic distance measures between vowels in the stimuli used in this study, where a lower number indicates less acoustic distance (i.e. greater similarity) and a higher number indicates greater acoustic distance (i.e. less similarity). For example, [ʏ] and [y] have an acoustic distance of approximately 130.8, while [a] and [o] have an acoustic distance of approximately 227.9, so [y] and [ʏ] are more similar

to each other than [a] is to [o]. While only specific vowel pairs were used as stimuli, distances between other vowels were calculated in order to provide a more complete picture of the relative distances between all vowels.

Table 6.8
Acoustic Distance by Vowel Pair

	[a]	[ɔ]	[o]	[ʊ]	[ʏ]	[y]
[a]	0	111.3945	227.8852	185.8103	178.209	252.6404
[ɔ]		0	150.2568	133.5706	148.3132	235.2551
[o]			0	94.16523	198.4187	226.4718
[ʊ]				0	153.6567	200.1113
[ʏ]					0	130.838
[y]						0

Figure 6.3, below, provides a graphic interpretation of Table 6.8 ordered from most similar to least similar vowel pair. This figure shows that the three vowel pairs [a-ɔ, o-ʊ, y-ʏ] used in this study are the three most acoustically similar of the possible combinations of the 6 vowels. Somewhat surprisingly, [ʏ] and [y] are not the most similar, which was expected based on the F1-F2 vowel plot above.

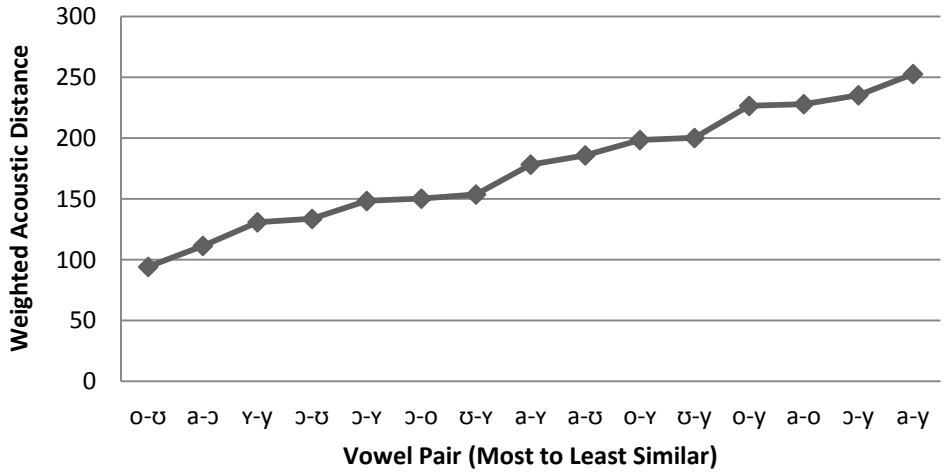


Figure 6.3. Weighted Acoustic Differences by Vowel Pair

It is important to recall that as Praat calculates the weighted distances between vowels, all spectral information is used, regardless of how salient the frequencies analyzed are to human speech perception. These results are therefore perhaps less surprising considering that [y] and [ʏ] exhibit greater differences in the higher frequencies (F3 and above) than the other vowels. It is not clear, however, whether the distance between [y] and [ʏ] – about 130 – is significantly different from the distance between [a] and [ɔ] – about 111. This analysis simply shows which vowel pairs are the most similar relative to other vowel pairs.

A further caveat to interpreting these results is that humans do not perceive all acoustic differences in proportion; for example, absolute differences in pitch are more difficult to perceive in the higher frequencies than in the lower frequencies (Yip, 2002). Therefore, it cannot be expected that participants will perceive the vowels according to the absolute acoustic differences provided above, so it cannot be predicted from this analysis that, for example, participants would perceive [o]-[ʊ] as the most similar pair, followed by [a]-[ɔ] as the second most similar pair, followed by [y]-[ʏ], because their phonological system will still play a role in how these vowels are perceived. The pattern found in the above results would be predicted, however, if participants were processing stimuli as if they were non-speech stimuli, which results in different predictions than if participants distinguish stimuli based on functional load, predictability of distribution and F1-F2 differences. If this obtains in the results, it would be reasonable to suspect that participants were performing the experimental tasks as if with non-speech stimuli; if this does not obtain in the results, this would likely be indicative of phonological structure being imposed on the

acoustic information, or else that the cues that are perceptually salient to participants are other than the cues measured in the weighted acoustic distances.

6.6.1.3. Principal Component Analysis: Non-Speech Perception

To further explore the acoustics of the stimuli used in Experiments 3 through 5, and to verify what components the weighted acoustic differences were based on, a principal component analysis was performed in R (statistical software, using the "princomp" and "predict" commands), again based on the methodology described in Mielke (2012). The acoustic distance measures above show to what extent two vowels are similar or not; the principal component analysis below shows what component contributed the most to that similarity or difference based on the weighted acoustic distances inputted into the analysis. The first principal component accounts for the largest amount of variance in the data; the second principal component accounts for the largest amount of variance that is left once the first component is accounted for, and so on for subsequent components. The principal component analysis shows how these vowels are similar to one another and how much each dimension contributes to their similarity. The acoustic dimensions that contributed the most can then be inferred based on the way in which the sounds analyzed pattern together. Figure 6.4 shows the vowels in this study plotted by the first two principal components ("Comp. 1" and "Comp. 2") that contributed the most to differences between vowels.

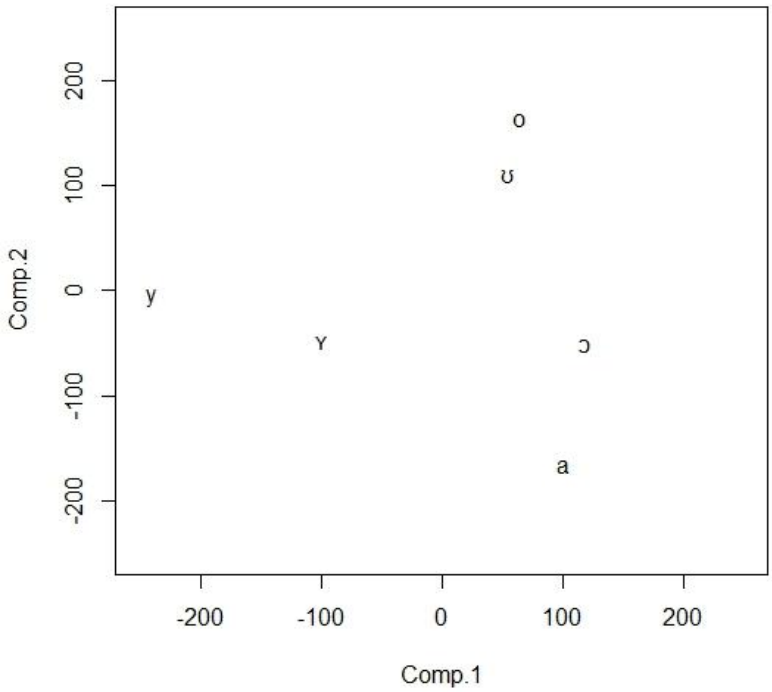


Figure 6.4. Principal Component Analysis

Figure 6.4 is a visualization of the contribution of each component to the acoustic distances between vowels, so that the distances between vowels can be considered a visualization of how similar the vowels are to one another.[19] As mentioned above, it was surprising that [y]-[ʏ] was not the most similar vowel pair, but perhaps higher frequencies that are less relevant to human speech perception played a role in this measure. To verify this effect, the stimuli were downsampled to 11000 Hz to eliminate periodicity above 5500 Hz and were then re-analyzed. Figure 6.5 shows the outcome of the re-analysis. Note that there is very little difference between Figures 6.4 and 6.5.

[19] Recall that these are not statements about these vowels in general or as prototypes but refer to specific tokens produced by one single speaker.

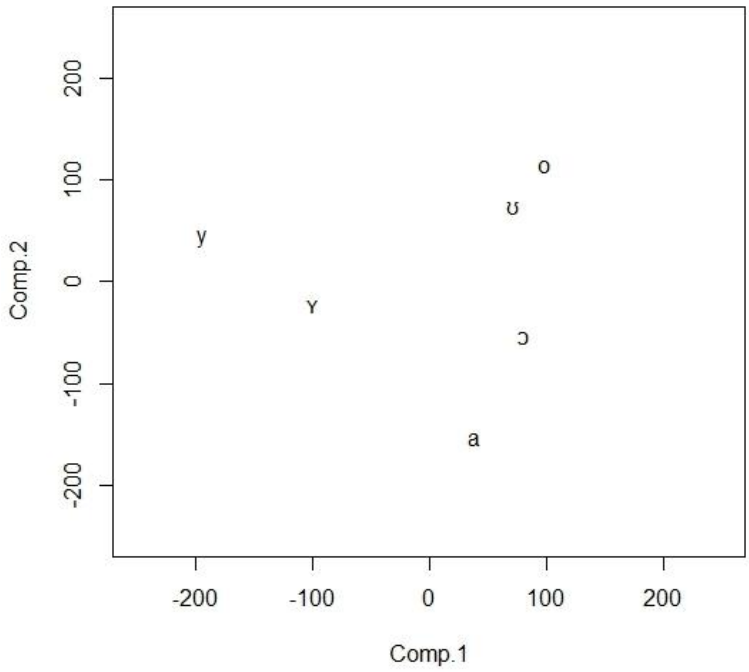


Figure 6.5. Principal Component Analysis with Downsampled Stimuli

Based on the differences between the vowels included in this analysis, and the fact that that they do not match the differences based on F1 and F2, it is possible that the weighted acoustic differences were skewed by parameters in the acoustic signal that are less salient to human speech perception such as in the higher frequencies (e.g. F3 and above). While such cues are less crucial to vowel perception in humans, a machine-based metric such as the one used by Praat would include all available information.

If participants perceive stimuli according to their weighted acoustic differences, this would suggest that they would be performing the task as a non-speech task (i.e. in a purely acoustic/auditory manner) evidenced by paying attention to cues that are not normally salient. Results would pattern according to these analyses in this

section, with [o]-[ʊ] perceived as most similar, [y]-[ʏ] being perceived as least similar, and results for [a]-[ɔ] falling between the other two pairs (L>H>M).

6.7. Predictions

There are different possible outcomes depending on what mode of processing participants use. Both accuracy scores and reaction times (RTs) were recorded in Experiments 3 and 4, and similarity-ratings were recorded in Experiment 5. There are three possible predictions that apply across Experiments 3 through 5:

1. Gradient Contrast. If participants do indeed employ a phonological mode of perception, and if the gradient view of contrast is correct, it is predicted that for the variable of Contrast (High, Mid, Low), the High Contrast [a]-[ɔ] vowel pair should be the easiest to discriminate, resulting in high accuracy, fastest reaction times and lowest similarity rating; the Mid Contrast [o]-[ʊ] vowel pair should result in lower accuracy scores, slower reaction times and a similarity rating between High and Low; the Low Contrast [y]-[ʏ] vowel pair should result in the lowest accuracy scores, slowest reaction times, and highest similarity rating (H>M>L).
2. Binary Contrast. On the other hand, if a binary view of contrast is correct, it is predicted that [a]-[ɔ] and [o]-[ʊ] would both yield similar results of faster RTs, higher accuracy and less similarity, while [y]-[ʏ] would not since they are in an allophonic relationship (H=M>L).
3. Acoustics: Non Speech. A final possibility is that participants may complete the task as if it were a non-speech task, responding in accordance with the weighted acoustic distances and global measures provided above, including

information from all possible frequencies available in the signal. In this case, results may pattern such that [o]-[ʊ] would yield the slowest RTs, lowest accuracy and highest similarity, followed by [a]-[ɔ], with [y]-[ʏ] yielding the fastest RTs, highest accuracy and least similarity, since this final pair was calculated to be the most different vowel pair, likely due to differences in formant values in the higher frequency ranges (L>H>M).

In summary, what the experimental results will show is (Prediction 1) whether a gradient or (Prediction 2) binary view of contrast is able to account for the results, or (Prediction 3) if the task was being processed in a purely auditory fashion. Table 6.9 compiles these predictions by dependent variable.

Table 6.9
Predictions Across Experiments by Result[20]

Prediction	Accuracy	RTs	Ratings
1. Gradient Contrast	H>M>L	H>M>L	H>M>L
2. Binary Contrast	H=M>L	H=M>L	H=M>L
3. Acoustics: Non-Speech	L>H>M	L>H>M	L>H>M

6.8. Experiments 3, 4 and 5

Three different experimental paradigms will be used to explore these predictions: an AX task (Experiment 3), a 4IAX task (Experiment 4), and a similarity-rating task (Experiment 5). No previous study has incorporated all three experimental paradigms, and for the 4IAX task, no study has reported RTs. Incorporating all three tasks will allow for comparisons with previous research where

[20] There is also a possibility that results will pattern according to the null hypothesis of H=M=L.

differences in the perception of contrastive and allophonic speech sounds were found with a single paradigm or with two paradigms, and will allow for a comparison of results across task type. For example, if one task yielded results that patterned as in Prediction (1) while another yielded results that patterned as in Prediction (2), it could indicate that different tasks are better suited to tapping into different modes of perception.

6.9. Summary

In summary, the phonological criteria of lexical distinction and predictability of distribution are represented in this study by minimal pair count and relative phoneme frequency. These were calculated for vowels in LF and an acoustically-similar sub-set of pairs of vowels were chosen as representative of a range of these values: vowels [a]-[ɔ] represent a “High Contrast” pair; [o]-[ʊ] represent a “Mid Contrast” pair; and [y]-[ʏ] represent a “Low Contrast” pair. Furthermore, the Low Contrast pair consists of allophones, which in a binary view of contrast would suggest that this pair will be perceived as being more similar than the other two.

The measured acoustic distances based on the phonetically-based metric indicate that, acoustically speaking, the vowel pair [o]-[ʊ] is the most similar pair, followed by [a]-[ɔ], followed by [y]-[ʏ]. Out of the possible vowel combinations of the six vowels chosen based on minimal pair count and phoneme frequency, these three vowel pairs are the most similar acoustically which should reduce the possibility that experimental results will only reflect acoustic differences. The weighted acoustic differences may eventually provide a kind of diagnostic so that if participants treat stimuli as non-speech in the subsequent experiments, then it is

expected that results would pattern according to absolute acoustic differences. These predictions will be tested by way of an AX task, a 4IAX task and a similarity-rating task, described in detail in the next three chapters.

Chapter 7 Experiment 3: AX Discrimination Task

This chapter outlines the experimental protocol for the AX task, including participant details, procedures, predictions, and results. Information regarding participants applies to the experiments in Chapters 7 and 8.

7.1. Predictions

The predictions for the AX task are as in §6.7: (1) according to a gradient structure of contrast, High Contrast pairs should yield higher accuracy and faster RTs, Low Contrast pairs should yield the lowest accuracy and slowest RTs, and Mid Contrast pairs yielding results between the two other pairs; (2) according to a binary structure of contrast, then High and Mid Contrast pairs should pattern similarly to one another with higher accuracy and faster RTs, and Low Contrast pairs should yield lower accuracy and slower RTs; (3) if stimuli are processed in a purely acoustic, non-speech manner, then, according to the weighted acoustic differences in Chapter 5, Mid Contrast pairs will yield the lowest accuracy and slowest RTs, followed by High Contrast pairs, followed by Low Contrast pairs. These are summarized in Table 7.1.

Table 7.1
Predictions for Experiment 3 by Result

	Prediction	Accuracy	RTs
1.	Gradient Contrast	H>M>L	H>M>L
2.	Binary Contrast	H=M>L	H=M>L
3.	Acoustics: Non-Speech	L>H>M	L>H>M

7.2. Participants

A total of 31 participants were tested. An additional 10 were tested but not included in the analysis for several reasons: being speakers of a dialect other than Laurentian French (N=5), not being a native speaker of French (N=1), keyboard problems (N=1), falling asleep during the experiment (N=1), being older than the age range cut-off which was 40 years (N=1), and not understanding the task (N=1).[21] The same 10 participants were removed from all three experiments to make results more comparable across experiments.

Of the 31 participants, the mean age was 27.06 years (median=24; SD=7.93), ranging from 19 to 38; 5 males. 14 participants were born and raised in Ottawa, ON; 3 in Gatineau, QC; the other participants were from various other places in Quebec and Ontario but had all been living in either Ottawa or Gatineau for the past 5 years. Twelve were recruited through the "Integrated System of Participation in Research" (ISPR) recruitment system of the Department of Psychology at the University of Ottawa and were all undergraduate students in Psychology receiving course credit for participating in the study. The project description, consent form, pre-screening questionnaire, and language background questionnaire in Appendices B through E. The remaining participants were of various backgrounds and were not compensated for their participation beyond an offer of thanks and cookies. These were recruited predominantly through word-of-mouth. None of the participants self-reported any hearing difficulties or learning disabilities.

[21] Accuracy scores suggest the participant guessed or did not understand the task.

7.3. Procedure

Participants were spoken to only in French prior to the experiments. Written instructions were also provided in French and presented on the screen at the beginning of each experiment. All participants heard all conditions in each of the three experiments, and they completed the three experiments in the same order (the AX task, followed by the 4IAX task, and then the similarity-rating task). The response keys and blocks within each experiment were mixed, described in more detail below.

In the AX task, there were two conditions: Trial Type (Different, Same) and Contrast (High, Mid, Low). There were 48 total trials, with 8 trials in each condition (Different: High, Mid, and Low Contrast; Same: High, Mid and Low Contrast; 6 x 8 = 48 trials). Stimuli were quasi-randomized by assigning pairs a random number in Excel, sorting the stimuli and then ensuring that there were no more than three consecutive trials of a condition. Stimuli were divided into two blocks of 24 trials with a self-timed pause between the blocks.

For the AX task, participants were told they would hear one syllable followed by another syllable. They were instructed to press one key if they thought the two syllables they heard were the same, or another key if they thought the two syllables were different. Response keys were labelled so participants would not have to remember which button was which. Of the 31 participants, 16 had the “Same” response correspond to their left hand (the < z > key) and “Different” correspond to their right hand (the < / > key); 15 had “Different” correspond to their left hand and “Same” correspond to their right hand. There were 2 experimental lists,

counterbalanced across participants. The beginning of each trial was indicated by a tone to draw the participants' attention.

7.3.1. Equipment

Most participants used a MacBook Pro, and four (of the retained participants) used a Mac Pro. Specifics for each computer are provided in Appendix F. Stimuli were presented using PsyScope experimental presentation software.

7.3.2. Stimuli

Four consonant C_C frames consisting of [l], [b, [f] and [ʃ] were combined with the six vowels [a], [ɔ], [o], [ʊ], [y] and [ʏ] making for 24 unique syllables. Syllable pairs in the Different condition were paired by consonant (i.e. all consonants in a given trial were the same) so that, as much as possible, participants would only use vowel differences to distinguish between stimuli. For example, they would hear [bob-bʊb] and never [bob-lʊl]. Stimuli in the Same condition always consisted of two acoustically different stimuli; for example, [bob_1-bob_2].

7.4. Results

Results will be presented first by accuracy and then by RTs, with a short discussion after each presentation of the results. The main discussion will be after all results have been presented, in Chapter 10.

7.4.1. Experiment 3: Accuracy

The mean accuracy score by Contrast is shown in Figure 7.1. The closer the score is to 1, the more correct answers were given. Figure 7.1 shows that accuracy scores were relatively high across conditions, with Different, Low Contrast pairs

exhibiting the lowest accuracy score. The error bars in this and subsequent figures represent the standard error as calculated by SPSS (version 21).

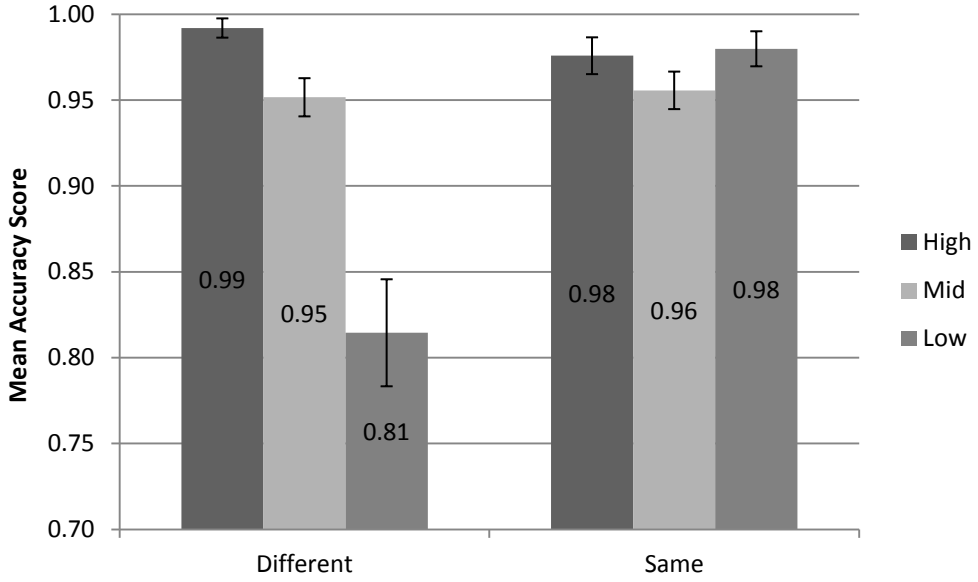


Figure 7.1. Mean Accuracy Scores by Contrast and Trial Type

To investigate possible reasons for the lower accuracy of Different Low pairs, the individual items were examined more closely and it was found that the majority of the errors were with [lyl-lʏl] and [lʏl-lyl] stimuli. Table 7.2 summarizes the total correct responses by stimuli pairs in the Different Low condition.

Table 7.2
Accuracy Scores for Different, Low Contrast Stimuli

Stimuli	Incorrect	Correct	%Correct	% of Total Errors
byb-bʏb	4	26	86.67	8.70
bʏb-byb	0	30	100.00	0.00
fyf-fʏf	3	27	90.00	6.52
fʏf-fyf	1	29	96.67	2.17
lʏl-lyl	17	13	43.33	37.00
lyl-lʏl	16	14	46.67	34.78
ʃʏʃ-ʃyʃ	1	29	96.67	2.17
ʃyʃ-ʃʏʃ	4	26	86.67	8.70

Note that [lʏl]-[lyl] and [lyl]-[lʏl] account for 37% and 34.78% of errors for the Different Low Contrast condition. Since the Low stimuli with the [l_l] frame in particular caused the majority of errors, and errors were not evenly spread over the rest of the condition to suggest that this was a property of the condition, these stimuli were removed. The vowels in these particular stimuli were the closest in terms of F1 and F2 values (see Figure 6.1) and this is likely the reason for their easy confusability. The equivalent Same stimuli (i.e. [lyl]-[lyl] and [lʏl]-[lʏl]) were also removed because high accuracy scores for these stimuli are difficult to interpret since participants often mistook Different Low stimuli with [l_l] as being the same, and so a high accuracy score for Same stimuli does not reveal whether they were accurately perceived or not. With these problematic stimuli removed from the analysis, the results found in Figure 7.2 are obtained.

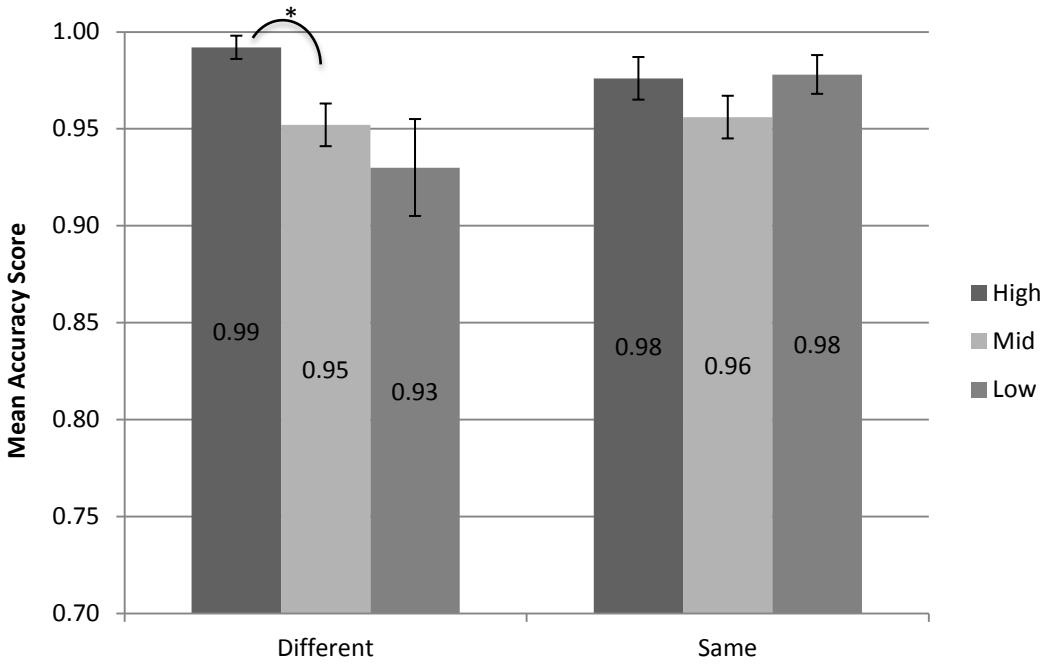


Figure 7.2. Mean Accuracy Scores Without [lyl-lʏl]/[lʏl-lʏl] Stimuli.
Note. A "*" indicates a statistically significant difference.

A repeated measures 2 x 3 ANOVA was done on accuracy scores with Trial Type (Same, Different) and Contrast (High, Mid, Low). For Contrast, the assumption of sphericity was violated and so Greenhouse-Geisser adjusted values were used. There were no significant main effects of Trial Type ($F(1, 30) = .92$, $p = .346$) or Contrast ($F(1.38, 41.39) = 2.73$, $p = .094$). There was, however, a significant interaction of Response Type x Contrast ($F(2, 60) = 3.84$, $p=.027$). To examine this interaction, Different and Same trials were analyzed separately with 1-way ANOVAs. For Different trials, the assumption of sphericity was violated and Greenhouse-Geisser-corrected values were used. There was a near significant effect of Contrast ($F(1.3, 39.1) = 3.63$, $p=.054$). Pairwise comparisons using Bonferroni corrections for multiple comparisons and are provided in Table 7.3, below.[22] Pairwise comparisons showed that participants were significantly more accurate on the High contrast pairs compared to the Mid Contrast pairs ($p = .007$), there was a trend of High Contrast pairs yielding higher accuracy compared to Low contrast pairs ($p =.08$), and no difference between Mid and Low Contrast pairs. For the 1-way ANOVA looking at Same trials, there was no effect of Contrast ($F(2,60) = 1.76$, $p = .181$).

[22] Bonferroni corrections have been used for all subsequent pairwise comparisons.

Table 7.3

Accuracy: Mean Differences by Contrast

Response Type	Contrast (a)	Contrast (b)	Mean Difference	Sig.
Different	High	Mid	0.040*	0.007
		Low	0.062	0.083
	Mid	Low	0.022	1.000
Same	High	Mid	0.020	0.402
		Low	-0.003	1.000
	Mid	Low	-0.023	0.343

Note: A "*" indicates a significant difference. Duplicate values were removed from the table (for example, High vs. Mid is included so Mid vs. High was not included).

7.4.2. Discussion: Accuracy

What first appeared to be poorer results for the Low Contrast condition among Different pairs proved to be due to responses to stimuli with the consonant [l]. Upon removing these stimuli, accuracy scores evened out across conditions and the overall accuracy rates were quite high, all in the high 90% range. Among the Different trials, participants were significantly more accurate on High contrast pairs than on Mid-Contrast pairs, and a trend of High Contrast pairs being more accurate than Low contrast pairs was found. While the accuracy among the High, Mid and Low Contrast pairs were not all statistically significant from one another, the scores decreased in accuracy from High to Mid to Low Contrast conditions, as per predictions according to Gradient Contrast (prediction 1). The predictions based on the binary view of contrast were not supported. This view predicted that H=M>L, but based on the accuracy results, High Contrast stimuli differed from Mid Contrast stimuli. It is clear from the fact that participants performed relatively well on Low vowel pairs that the vowels [y-ʏ] were not so similar to one another so as to be

impossible to discriminate, at least at an auditory level, but the pair was difficult enough to cause more errors than the other conditions.

It is unsurprising that differences among Same pairs were not statistically significant across conditions since it is less difficult to confirm two things are the same than to determine differences between stimuli and then identify them as different, resulting in the known response bias for AX tasks where participants tend to choose "Same" when unsure (Gerrits & Schouten, 2004).

7.4.3. Experiment 3: Reaction Times

Due to the known issue with stimuli containing [lyl] and [lʏl] in Low Contrast pairs, the problematic stimuli were removed from the RT analyses as well. The following analyses for RTs are based on correct responses only (46 incorrect responses and 2 time-outs[23] were removed from the analyses). Within the set of remaining correct responses (i.e. without incorrect responses and without stimuli containing [lyl] and [lʏl]), responses that were more than three standard deviations above or below each condition mean were removed (N=45). The total amount of data removed was 14.58% (217/1442 trials).

Once these trials were removed, there was one participant who had no correct responses for the Different Low condition. For the missing value for this participant, SPSS was used to impute a value that was statistically plausible for this participant given their other response times and the response times of all other participants.

[23] When participants took too long to answer, no reaction time was recorded by PsyScope. These cases are referred to as "time-outs."

This was the only imputed value. Figure 7.3 provides the mean reaction times per condition.

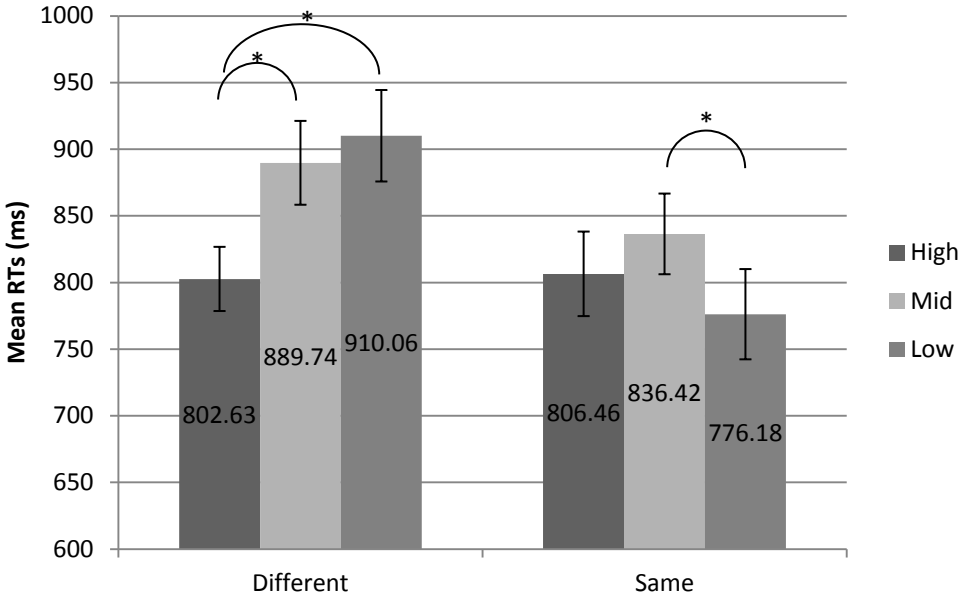


Figure 7.3. Mean Reaction Times by Contrast and Trial Type.
Note: A "*" indicates a significant difference.

A 2 x 3 repeated measures ANOVA showed main effects of Trial Type ($F(1, 30) = 7.82$, $p = .009$), Contrast ($F(2, 60) = 9.57$, $p < .001$) as well as a significant interaction between Trial Type and Contrast ($F(2,60) = 13.55$, $p < .001$). To further explore the interaction, 1-way ANOVAs were done on Same and Different trials separately. A significant effect of Contrast was found among Different pairs ($F(2,60) = 15.45$, $p < .001$), as well as among Same pairs ($F(2,60) = 6.01$, $p = .004$). Table 7.4 provides the pairwise comparisons of High, Mid and Low contrasts for both Different and Same trial types.

Table 7.4
Mean Differences in RTs by Contrast and Trial Type

Response Type	Contrast (a)	Contrast (b)	Mean Difference	Sig.
Different	High	Mid	-87.11*	<.001
		Low	-107.43*	<.001
	Mid	Low	-20.32	1.00
Same	High	Mid	-29.96	0.310
		Low	30.29	0.332
	Mid	Low	60.24*	0.002

Note: A "*" indicates significance. A negative value indicates that the condition in the first column was faster than the condition it is being compared to in the second column; a positive value indicates it was slower.

To summarize Table 7.4, among Different pairs, RTs for the High Contrast condition were significantly faster compared to the Mid and Low Contrast conditions. In the Same trials types, the Low Contrast condition was significantly faster compared to the Mid Contrast condition.

7.4.4. Discussion: Reaction Times

Based on a gradient notion of contrast, it was predicted in (1) that High Contrast stimuli ([a]-[ɔ]) would yield the fastest RTs, followed by Mid Contrast ([o]-[ʊ]), followed by Low Contrast ([y]-[ʏ]) (H>M>L). This is partly borne out for the Different pairs, though only the High condition is statistically significant from the other two contrast levels. However, if the contrastive pairs [a-ɔ] and [o-ʊ] yielded similar results and responses to allophones [y-ʏ] yielded different results (H=M>L), this would have provided support to the view that participants were utilizing a phonological mode of processing that was dependent on a binary type of phonological relationship. As the results statistically set the High Contrast condition apart from the Mid Contrast condition, this scenario does not appear to apply. For the Same pairs, the Low Contrast pairs ([y]-[ʏ]) yielded the fastest RTs compared to

the Mid Contrast pairs, which was not predicted. Despite the unexpected results, they still do not support a binary conception of phonological relationship because the Low Contrast pair would have been significantly different from both the High and Mid Contrast pairs.

7.5. Discussion: Overall Results for Experiment 3

A summary of results is provided in Table 7.5. An "x" indicates the results did not support the prediction; a "(✓)" symbol in parentheses indicates the prediction was partially supported by the results obtained.

Table 7.5
Experiment 3: Summary of Results for Different Pairs

Results	**Prediction 1: Gradient Contrast H>M>L**	**Prediction 2: Binary Contrast H=M>L**	**Prediction 3: Non-speech Acoustics L>H>M**
Accuracy	(✓)	x	x
RTs	(✓)	x	x

Based on a gradient view of contrast (Prediction 1), it was predicted that the level of Contrast (High, Mid, Low) would be reflected in both RTs and accuracy scores with the fastest RTs and highest accuracy for the High condition, followed by the Mid condition, with the slowest RTs and lowest accuracy for the Low condition. This prediction was partially borne out. While the only significant difference for accuracy scores was found between Different High and Mid pairs, results in terms of RTs showed High Contrast pairs were significantly different from both Mid and Low pairs: High Contrast pairs yielded the fastest RTs and highest accuracy, but Mid and Low pairs were not significantly different from one another. Overall, the results

demonstrate a preference and facility for High Contrast stimuli. If contrast was strictly binary in nature (Prediction 2), then it would be expected that High and Mid pairs would yield roughly identical results, and this was not the case, nor was there evidence of a non-speech mode of perception based on raw acoustic differences (Prediction 3). Overall, there was a trend for results to pattern according to a gradient view of contrast.

Chapter 8 Experiment 4: 4-Interval AX Discrimination Task

This chapter presents the predictions, experimental procedure, design, and results for the four-interval AX (4IAX) task.

8.1. Predictions

The predictions for the 4IAX task are as in §6.7 and the same as for the AX task: (1) if a gradient view of contrast is supported, High Contrast pairs should yield higher accuracy and faster RTs, Low Contrast pairs should yield the lowest accuracy and slowest RTs, and Mid Contrast pairs yielding results between the two other pairs; (2) if a binary view of contrast is supported, then High and Mid Contrast pairs should pattern similarly to one another with higher accuracy and faster RTs, and Low Contrast pairs should yield lower accuracy and slower RTs; (3) if stimuli are processed in a purely acoustic, non-speech manner, then, according to the weighted acoustic differences in Chapter 6, Mid Contrast pairs will yield the lowest accuracy and slowest RTs, followed by High Contrast pairs, followed by Low Contrast pairs. These are summarized in Table 8.1.

Table 8.1
Predictions for AX Task by Result

	Prediction	Accuracy	RTs
1.	Gradient Contrast	H>M>L	H>M>L
2.	Binary Contrast	H=M>L	H=M>L
3.	Acoustics: Non-Speech	L>H>M	L>H>M

8.2. Participants

Participants were the same as described in Chapter 6 (N=31). An analysis of the accuracy scores for the participants showed that one participant (P8F) was an outlier, and so this participant was removed from the analyses for the 4IAX task. (Appendix G provides a boxplot showing P8F as an outlier.)

8.3. Procedure

Participants were instructed in French to press one of four keys according to whether they thought they heard: (a) the first of two pairs containing same syllables, e.g. [bab-bab]+[bab-bɔb], (b) the second of two pairs containing same syllables, e.g. [bab-bɔb]+[bab-bab], (c) both pairs containing same syllables, e.g. [bab-bab]+[bab-bab], or (d) neither pair containing same syllables (i.e. two different pairs), e.g. [bab-bɔb]+[bab-bɔb]. Only vowels for a single condition were heard in a given trial, so that, for example, participants never heard one pair with High Contrast vowels (e.g. [bab-bɔb]) paired with a pair with Low Contrast vowels (e.g. [byb-bYb]); High Contrast vowels were always presented alongside other High Contrast vowels.

In the 4IAX task, there were 2 conditions, Trial Type (Different-Different, Mixed Different-Same, Same-Same) and Contrast (High, Mid, Low). There were 2 experimental lists, counterbalanced across participants. Stimuli were quasi-randomized as in Chapter 6. The response keys were not changed between participants (so that, for example, the response keys for “both same” and “both different” never differed), and a cardboard cut-out was used to cover all but the relevant response keys which were labelled so that participants did not have to

memorize which key was which. There was a short practice block of 4 trials. Each of the subsequent 4 blocks consisted of 24 trials each for a total of 96 trials. The beginning of each trial was indicated by a tone to draw the participants' attention. There was a self-timed pause between each block.

8.3.1. Equipment

The equipment was the same as described in Chapter 7.

8.3.2. Stimuli

The stimuli were the same as described in Chapter 7, with four consonant C_C frames consisting of [l, b, f] and [ʃ] were combined with the six vowels [a], [ɔ], [o], [ʊ], [y] and [ʏ] making for 24 unique syllables. However, the combination of syllables in the 4IAX task was different from the AX task described in Chapter 7.

Table 8.2 presents sample stimuli by Contrast and Quad Type. "H" corresponds to "High"; "M" to "Mid"; "L" to "Low"; "S" to "Same"; and "D" to "Different." Same-Different and Different-Same quads in the 4IAX task include one pair of each type were collapsed to form a "Mixed" condition for analysis.

Table 8.2
Stimuli by Condition and Response Type

Contrast Level	Quad Type	Condition	1st Stimuli Pair	2nd Stimuli Pair	Response Type
High	HS-HD	H-Mix	ʃaʃ-ʃaʃ	ʃaʃ-ʃɔʃ	1st Same
	HD-HS		ʃaʃ-ʃɔʃ	ʃaʃ-ʃaʃ	2nd Same
	HS-HS	H-Same	ʃaʃ-ʃaʃ	ʃaʃ-ʃaʃ	Both Same
	HD-HD	H-Diff.	ʃaʃ-ʃɔʃ	ʃaʃ-ʃɔʃ	0 Same
Mid	MS-MD	M-Mix	ʃoʃ-ʃoʃ	ʃoʃ-ʃʊʃ	1st Same
	MD-MS		ʃoʃ-ʃʊʃ	ʃoʃ-ʃoʃ	2nd Same
	MS-MS	M-Same	ʃoʃ-ʃoʃ	ʃoʃ-ʃoʃ	Both Same
	MD-MD	M-Diff.	ʃoʃ-ʃʊʃ	ʃoʃ-ʃʊʃ	0 Same
Low	LS-LD	L-Mix	ʃyʃ-ʃyʃ	ʃyʃ-ʃʏʃ	1st Same
	LD-LS		ʃyʃ-ʃʏʃ	ʃyʃ-ʃyʃ	2nd Same
	LS-LS	L-Same	ʃyʃ-ʃyʃ	ʃyʃ-ʃyʃ	Both Same
	LD-LD	L-Diff.	ʃyʃ-ʃʏʃ	ʃyʃ-ʃʏʃ	0 Same

8.4. Excluded Stimuli

The following items were removed from the present analysis:

-An extra MD-MD quad was included (ʃoʃ-ʃʊ̲ʃ + ʃʊ̲ʃ-ʃoʃ), rather than MS-MD (ʃoʃ-ʃo̲ʃ + ʃʊ̲ʃ-ʃoʃ). While this was corrected later in the testing phase, the extra MD-MD tokens and the corrected MS-MD tokens were removed. (Total removed for **MS-MD**: **31** tokens.)

-The responses for one token of HD-HS (87+37) (fɔf-faf fɔf-fɔf) were not recorded for 15 of 31 participants due to programming error. This token was removed for all participants so that means for this condition were based on equal numbers across participants. (Total removed from **HD-HS**: **31** tokens.)

-8 trials were programmed as MS-MD quads rather than MS-MS. While this was corrected later in the testing phase, these tokens were removed to keep conditions balanced. (Total removed from **MS-MS**: **31** tokens.)

-The incorrect audio file was used for [lol-lʊl + lo̲l-lo̲l] . Responses for these trials were removed. (Total removed from **MD-MS**: **31** tokens.)

-**2** tokens were removed as “misfires” in the **MS-MS** condition (participants hit a response key before they had heard the second pair in the quad).

-**19** tokens were removed as “timeouts” (i.e. participants did not respond) from multiple conditions.

-Items with [lyl] or [lʏl] were found to be problematic in the AX task. These again appeared to cause some issues for listeners in the 4IAX task and so were removed from this analysis. **244** total items were removed from the **L-Diff.**, **L-Mix** and **L-Same** conditions.

-The rest of the responses for P8 were removed (N=**84**).

The original experiment had 2976 tokens across conditions and all responses. After the exclusions described above, the total number of items (before trimming) was **2503** (2976-31-31-31-31-2-19-244-84=2503). The remaining data was trimmed by 3 standard deviations above and below the mean of correct responses for each condition, removing another 152 tokens, or 6.69%. The grand total of items included in this analysis for RTs is 2166 (correct responses only), and for accuracy is 2351.

8.5. Results

Results will be presented first by accuracy and then by RTs, with a short discussion after each presentation of the results. The main discussion will be after all results have been presented, in Chapter 10.

8.5.1. Experiment 4: Accuracy

Same-Different and Different-Same quad types in the 4IAX task include one pair of each type and since no differences between these two trial types were anticipated, they were collapsed into a single “Mixed” condition. There was therefore

a three-level factor of Trial Type included in the subsequent analyses: Different-Different quads, Mixed (i.e. Same-Different and Different-Same) quads, and Same-Same quads.

Figure 8.1 provides the mean accuracy score by condition. The closer the accuracy score is to 1, the more correct responses were given. Error bars in all subsequent graphs represent +/- 1 standard error.

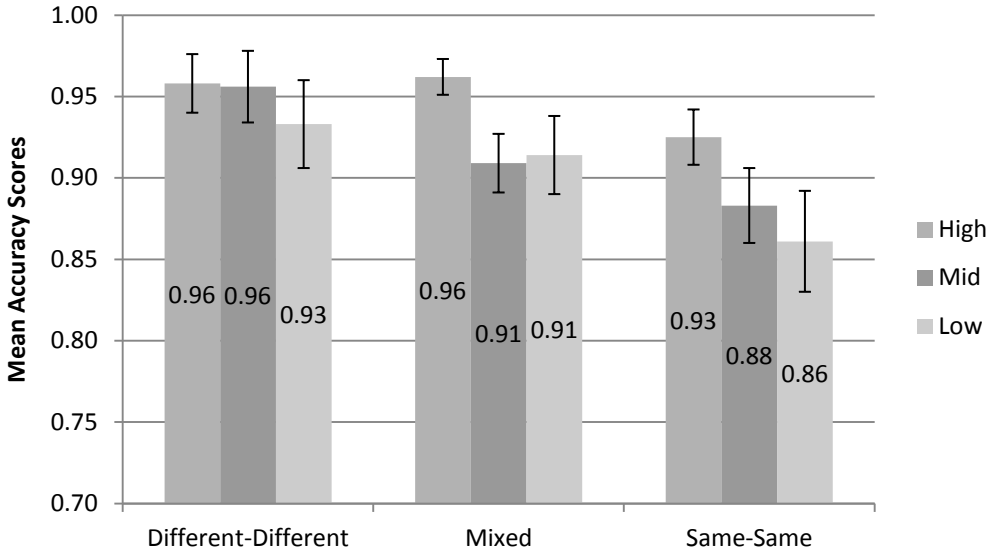


Figure 8.1. Mean Accuracy Score by Contrast and Quad Type.

A 3 x 3 repeated measures ANOVA with Trial Type (Different-Different, Mixed, Same-Same) and Contrast (High, Mid, Low) was done on the percent of correct responses. The assumption of sphericity was violated for each factor and interaction, and so Greenhouse-Geisser adjusted values will be cited. A main effect of Contrast was found ($F(1.67, 48.52) = 4.51$, $p = .021$), as well as Trial Type ($F(1.41, 40.95) = 4.46$, $p = .029$) but there was no significant interaction between Trial Type and Contrast ($F(2.18, 63.14) = .851$, $p = .44$). For the main effect of Trial Type (which collapses results across Contrast) there were no significant pairwise

comparisons, but there was a trend for the Same-Same pairs to have lower accuracy scores than the Different-Different pairs (mean difference = .059, p = .085). For the main effect of Contrast (which collapses results across Trial Type), pairwise comparisons are provided in Table 8.3.

Table 8.3
Experiment 4: Pairwise Comparisons by Contrast - Accuracy

Contrast (a)	Contrast (b)	Mean Difference	Sig.
High	Mid	0.032*	0.039
	Low	0.046	0.055
Mid	Low	0.013	1.00

Note. A "*" indicates a significant result.

It was found that the High Contrast stimuli yielded higher accuracy scores than Mid and Low pairs. Mid stimuli were not significantly different from Low stimuli.

8.6. Discussion: Accuracy

It was predicted that within a gradient view of contrast (Prediction 1), different levels of contrast would all be significantly different from one another, ranging from higher scores for High Contrast to lower scores for Low Contrast. In a binary view of contrast (Prediction 2), High and Mid contrast stimuli would yield similar results since the vowels to represent those levels of contrast are all contrastive. Results indicated that High Contrast stimuli yielded higher accuracy scores than Mid and Low Contrast stimuli, indicating that predictions based on a binary view of contrast were not borne out. While results do not perfectly match Prediction 1, results are trending in that direction. Furthermore, results do not indicate that participants were performing the task in a purely acoustic mode (Prediction 3) because no significant difference was found between the Mid and Low Contrast conditions.

8.7. Experiment 4: Reaction Times

Figure 8.2 shows the mean RTs by Trial Type and Contrast.

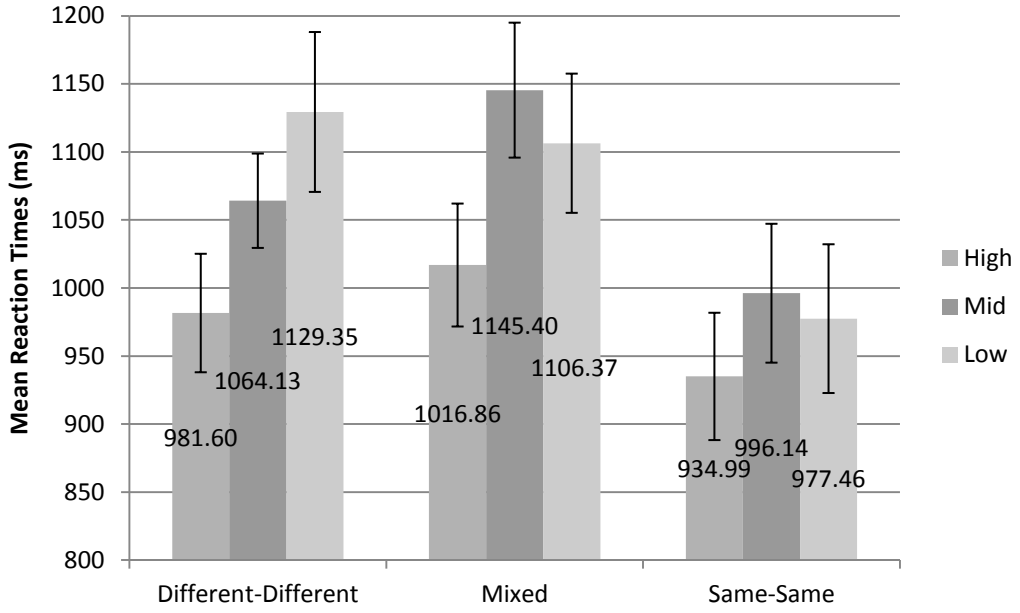


Figure 8.2. Experiment 4: Mean Reaction Times by Trial Type and Contrast.

A 3 x 3 repeated-measures ANOVA was done on Trial Type (Different-Different, Mixed, Same-Same) and Contrast (High, Mid, Low). The assumption of sphericity was violated for Trial Type, as well as for the interaction between Contrast and Trial Type, and so Greenhouse-Geisser adjusted values were used for these analyses. A main effect of Contrast was found ($F(2, 60) = 12.22$, $p < .001$), as well as for Trial Type ($F(1.58, 47.52) = 11.43$, $p < .001$). The interaction between the two factors was not significant ($F(2{,}68, 80.42) = 1.79$, $p = .161$). Looking at the effect of Trial Type, Same-Same quads yielded significantly faster RTs than both Different-Different quads ($p = .02$) and Mixed quads ($p<.001$). Looking at the effect of Contrast, Table 8.4 provides pairwise comparisons and significance values for High vs. Mid vs. Low pairs.

Table 8.4
Experiment 4: Pairwise Comparisons by Contrast Collapsed Over Trial Type - RTs

Contrast (a)	Contrast (b)	Mean Difference	Sig.
High	Mid	-93.264*	< .001
	Low	-105.904*	0.002
Mid	Low	-12.64	1.00

It was found that High Contrast quads yielded significantly faster RTs than Mid and Low quads, while Mid and Low quads were not significantly different from one another.

8.8. Reaction Times: Discussion

It was predicted according to the gradient view of contrast (Prediction 1) that High Contrast stimuli would yield faster RTs than Mid Contrast stimuli, which in turn would yield faster RTs than Low Contrast stimuli, while according to a binary view of contrast, High and Mid Contrast pairs would yield similar results, and Low Contrast (allophonic) stimuli would yield slower results (Prediction 2). Results indicate that High Contrast stimuli elicited faster RTs than both Mid and Low Contrast stimuli, while the Mid and Low conditions were not statistically different from one another. While this does not entirely support Prediction 1, it does suggest that Prediction 2 does not hold. A non-speech approach (Prediction 3) to the experiments was also not apparent in the results, since Mid and Low conditions were not different from one another.

8.9. Discussion: Overall 4IAX Results

A summary of the results is provided in Table 8.5.

Table 8.5
Experiment 4: Summary of Results

Results	**Prediction 1: Gradient Contrast H>M>L**	**Prediction 2: Binary Contrast H=M>L**	**Prediction 3: Non-speech Acoustics L>H>M**
Accuracy	(✓)	x	X
RTs	(✓)	x	X

Since no effect of Trial Type was found, only the results by Contrast will be discussed. While results for this task did not perfectly match Prediction 1 of higher accuracy for High Contrast stimuli, lower accuracy for Mid Contrast stimuli, and lowest accuracy scores for Low Contrast stimuli, findings do seem to refute Prediction 2. If relationships followed a strict contrastive/allophonic dichotomy, High and Mid stimuli should have yielded similar results; however, this was not found. With regards to RTs, results mirrored the accuracy results, with High Contrast stimuli yielding faster RTs than Mid and Low stimuli, showing a greater degree of certainty by participants and facilitation in processing. No evidence was found for Prediction 3, whereby if participants perceived stimuli in a non-speech manner, their results would match the weighted acoustic differences outlined in Chapter 6.

Chapter 9. Experiment 5: Similarity Rating Task

This chapter presents the experimental procedure, predictions and results for the similarity-rating task.

9.1. Predictions

There are several possible outcomes for this experiment, discussed in Chapter 6 and summarized in Table 9.1.

Table 9.1
Predictions for Experiment 5: Similarity Rating Task

Prediction	Ratings
1. Gradient Contrast	H>M>L
2. Binary Contrast	H=M>L
3. Acoustics: Non-Speech	M>H>L

The first possibility (Prediction 1) is that participants will judge the three vowel pairs based on a gradient conception of contrast as defined in Chapter 6, i.e. according to the functional load as measured in minimal pairs vowels and by the type frequency of the individual vowels in the sub-set of syllables used in this study, reproduced here in Table 9.2:

Table 9.2
Vowels by Minimal Pair Counts and Frequency

Contrast Level	**Individual Counts**		**Shared Counts**	**Frequency**
High	a	616	86	305
	ɔ	441		202
Mid	o	378	22	154
	ʊ	333		110
Low	y	207	0	63
	ʏ	245		76

In this scenario, pairs containing [a-ɔ] would be perceived as most different from one another; pairs containing [y-ʏ] will be perceived as most similar; and pairs containing [o-ʊ] will be perceived somewhere on the scale of similarity between the other two pairs. In this scenario: [a-ɔ] > [o-ʊ] > [y-ʏ], or H > M > L.

Another possibility is that vowels will be perceived according to a binary conception of contrast (Prediction 2) in which both [o-ʊ] and [a-ɔ] pairs will be perceived as more dissimilar while [y-ʏ] will be perceived as more similar since [y] and [ʏ] are allophonic and do not participate in any minimal pairs with each other, while the other vowel pairs do. In other words, if simply participating in a minimal pair is sufficient to qualify two vowels as contrastive, then the quantity of minimal pairs shared should be irrelevant: those that do ([o-ʊ] and [a-ɔ] pairs) will be easy to distinguish and be classified as more dissimilar, while those that do not ([y-ʏ] pairs) will be difficult to distinguish and be classified as more similar. In this scenario: [a-ɔ] = [o-ʊ] > [y-ʏ], or H = M > L.

A final possibility is that participants will perform this task as a non-speech task (Prediction 3) relying solely on acoustic differences to judge similarity between stimuli. If this is the case, it is expected that participants' judgments will roughly map to the weighted acoustic distances presented in Chapter 6, where [o-ʊ] was the most similar vowel pair, [a-ɔ] was the second-most similar, and [y-ʏ] was the third-most similar of the possible vowel combinations. These results can be schematized as: [y-ʏ] > [a-ɔ] > [o-ʊ], where ">" means "is more different than," or L > H > M.

9.2. Participants

Participants were the same as described in Chapter 7 (N=31).

9.3. Procedure

For this task, participants were told that they would hear two different syllables and their task was to decide how similar or how different the two syllables were on a scale of 1 to 6 with "1" being "Not very similar" ("Peu similaire") and "6" being "Very similar" ("Très similaire"). A six-point scale was used so as to avoid the use of a middle number as a placeholder when uncertain, as sometimes happens with odd-numbered scales (Matell & Jacoby, 1971). They were told that no two syllables were the same so that using a "6" did not mean that stimuli were identical.[24] Stickers with numbers were affixed to the lower letters of the keyboard (keys "x" to "m" were labelled "1" to "6") along with a reminder of what the extreme numbers meant. Everyone had the same scale, so that "x" was always "1 - Not very similar" and "m" was always "6 - Very similar." Participants were told that there was no time limit, that there was no correct answer, and to trust their own spontaneous judgment. It was not possible to replay any of the trials. The task lasted approximately five minutes and there were no breaks. As with the AX task, there was a tone at the beginning of each trial to draw the participants' attention.

9.4. Stimuli

Stimuli consisted of only different pairs of the same syllables used in the AX task. Participants heard each CVC-CVC stimulus pair once (consonants: [b], [f], [l],

[24] Pilot experiments included Same stimuli, but it became clear that participants relativized similarity when these were included, so that all Same stimuli received a "6" and all Different stimuli received values on the lower end of the scale, either 1, 2 or 3. It was therefore decided to only test Different stimuli.

[ʃ]; vowels: [a], [ɔ], [o], [ʊ], [y], [ʏ]), totalling 24 trials and 8 instances each of High, Mid, and Low Contrast stimuli. The two syllables were separated by 1500 ms.

9.5. Z Score Transformation

As in Boomershine et al. (2008) and Hall (2008), raw similarity-rating scores were transformed into z scores. This was done to normalize response patterns across individuals. For example, if one participant only used responses 1-4 (on the scale from 1-6) while another participant used predominantly 3-6, differences between conditions across participants could potentially be obscured, whereas transforming raw scores into z scores emphasizes differences between conditions. The z score transforms the mean of a participant's scores to 0 and counts how many standard deviations above or below the mean a given response lies. To illustrate, fictitious data is provided in Table 9.3 where ratings are given for three participants over three stimuli:

Table 9.3
Fictitious Raw Data

	Stimulus Pair X	**Stimulus Pair Y**	**Stimulus Pair Z**	**Mean**	**Standard Deviation**
Participant 1	1	2	3	2	1
Participant 2	4	5	6	5	1
Participant 3	1	3	4	2.67	1.53

To obtain the z score of these scores, each mean rating of the participant is subtracted from an individual rating and divided by the standard deviation:

$$\text{Z Score} = \frac{\text{Sample Value} - \text{Sample Mean}}{\text{Standard Deviation}}$$

For example, for Participant 1:

$$\frac{1\ [\text{stimulus pair X}] - 2\ [\text{mean for Participant 1}]}{1\ [\text{standard deviation}]} = -1$$

A negative value indicates the score was below the mean; a positive value indicates it was above the mean. Transforming the above scores into z scores, we obtain the following:

Table 9.4
Fictitious Z Score Data

Z-Scores	Stimulus Pair X	Stimulus Pair Y	Stimulus Pair Z
Participant 1	-1.00	0.00	1.00
Participant 2	-1.00	0.00	1.00
Participant 3	-1.09	0.22	0.87

From the transformation, we can see that despite the fact that Participant 1 and Participant 2 gave very different raw scores for the three stimuli pairs, the z-scores are the same: stimulus pair X was rated less similar than stimulus pair Y, which was rated less similar than stimulus pair Z, and, more to the point, by the same amount relative amount. Multiply this example over many stimuli, for instance, in the case where a participant may avoid one end of the similarity-rating scale, and such generalizations may be missed if only raw rating scores were used. Also note that the variance is minimized and is more representative of the differences between participants, which are in fact very minor. Based on this reasoning, it was therefore decided that z-score transformations would better represent the results than the raw scores, and so those will be reported below.

9.6. Results

As with Experiment 3 (AX task), Low stimuli (i.e. with [y] or [ʏ]) in [l_l] frames were removed. Figure 9.1 shows the normalized similarity ratings averaged across participants by Contrast (High, Mid, Low). Error bars show the standard error.

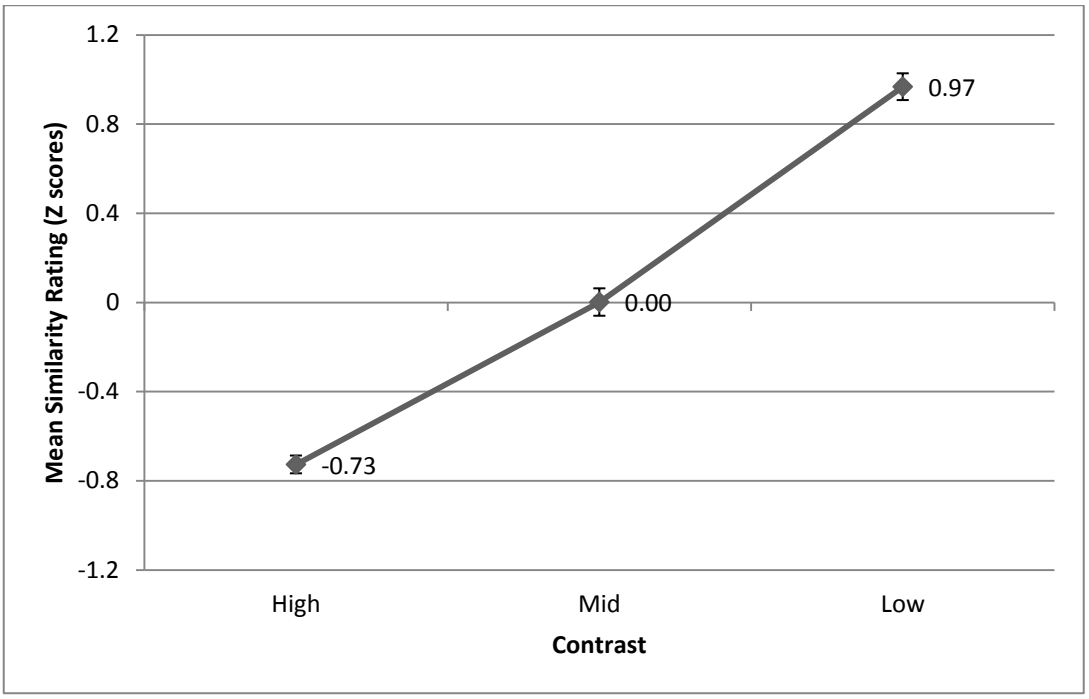


Figure 9.1. Average Similarity Rating (Z Scores).

A 1-way repeated measures ANOVA with 3 levels was done to determine if the differences between similarity ratings for High, Mid and Low Contrast pairs were significant. The assumption of sphericity was violated and so Greenhouse-Geisser adjusted values were used. Results showed that there was a main effect of Contrast ($F(1.62, 48.63) = 162.46$, $p < 0.001$). Pairwise comparisons provided in Table 9.5 indicate that all conditions were statistically significantly different from one another.

Table 9.5
Similarity Ratings: Pairwise Comparisons by Contrast

Contrast (a)	Contrast (b)	Mean Difference	Sig.
High	Mid	-.728*	<.001
	Low	-1.69*	<.001
Mid	Low	-.966*	<.001

Note: A "*" indicates a significant result.

Normalized rating scores show that High Contrast stimuli pairs ([a-ɔ]) were perceived as most different, that Low Contrast ([y-ʏ]) pairs were perceived as most similar, and Mid Contrast ([o-ʊ]) fell between the two.

9.7. Discussion

In summary, High Contrast pairs were judged to be the most different; Low Contrast pairs were judged to be the most similar; and Mid Contrast pairs were judged to fall between High and Low in terms of similarity (H>M>L). These findings are consistent with Prediction 1 in which it was expected that participants would perceive vowel pairs based on gradient levels of contrast. No support for the binary view of contrast was found (Prediction 2) where High and Mid vowels should have yielded identical results. Finally, no evidence was found that participants treated the task as non-speech in which the raw acoustics would determine the direction of results. A summary of these results is provided in Table 9.6.

Table 9.6
Experiment 5: Summary of Results

Predictions	Ratings
1. H>M>L	H>M>L
2. H=M>L	
3. L>H>M	

Chapter 10 General Discussion and Conclusion

This dissertation examined the notion that phonological relationships do not always perfectly match the criteria for being either contrastive or allophonic. Different possible levels of contrast were therefore tested to see whether phonological relationships are perceived as being entirely of one type or another (i.e. only contrastive vs. allophonic), or whether degrees of contrast can be perceived (i.e. on a scale from contrastive to allophonic). This chapter will summarize the findings of this dissertation, first with regards to allophony and contrast as binary categories of phonological relationship, and then with regards to the converging evidence to support a gradient view of contrast.

10.1. Allophony and Contrast: Two Extremes on a Continuum

In this dissertation, Experiments 1 and 2 tested the extremes of the range of possible phonological relationships from allophony to contrast. Native speakers of Laurentian French (LF) were tested on their perception of [t] and [s], other contrastive segments such as [p] and [f], and allophones [t] and [ts] (as well as their voiced counterparts). The results confirmed previous findings such as in Boomershine et al. (2008) that sounds in an allophonic relationship yield slower reaction times (RTs) compared to contrastive sounds. Accuracy scores yielded no statistically significant differences. Table 10.1 provides a summary of results by prediction for Experiments 1 and 2. A "✓" symbol indicates that the results supported the prediction of that column; an "x" indicates the results refuted the prediction; ">" indicates "should yield better results than"; "≠" indicates "yields results different from", "=" indicates "yields results similar to."

Table 10.1 Summary of Results for Experiments 1 and 2: Allophony vs. Contrast

Experiment	Results	Prediction 1: Relationship C=Ctrl>A	Prediction 2: Acoustics C≠Ctrl>A
1. AX	Accuracy	x	X
	RTs	✓	X
2. 4IAX	Accuracy	x	X
	RTs (Same-Diff.)	✓	X

Note: "C" = "Contrastive", "Ctrl" = "Control", "A" = "Allophonic."

If the results had been due only to acoustic differences between stimuli, one would also expect to see differences in RTs between the Control and Contrastive conditions (as in Prediction 2) since these stimuli were comprised of different contrastive consonant onsets with different acoustic properties; however, both Control and Contrastive conditions yielded virtually identical results. Instead, there was clear evidence that stimuli with consonants in an allophonic relationship were more difficult to perceive than stimuli with consonants in a contrastive relationship, with the results for allophones being significantly slower than for contrastive segments, even though no significant differences were found in accuracy scores. Results from the experiments testing these two extremes of the continuum of contrastive relationships confirm previous findings that phonological relationships are important to processing speech.

10.2. Testing Contrast on a Continuum

The experimental testing of contrast should not stop at the two ends of the scale of allophony and contrast, and these two ends of the scale cannot be taken as

representative of all possible phonological relationships. The criteria used in theoretical analyses to determine the nature of a phonological relationship, such as lexical distinction and predictability of distribution, can be problematic when interpreted as all-or-nothing criteria; for example, either there are minimal pairs or there are not, or the distribution is either completely predictable or it is not. Such criteria, however, can instead be interpreted and represented by way of gradient, quantifiable measures, and in doing so can be used to evaluate and potentially classify intermediate relationships that do not fall neatly into the category of "allophonic" or "contrastive." By way of functional load, it is possible to determine how many minimal pairs are shared between two segments, as well as how likely it is that a segment will appear in a particular phonetic environment, as was done in this dissertation. The resulting measures can then be used to evaluate how the selected criteria contribute to the strength of the contrast between two speech sounds.

In Experiments 3, 4 and 5, the criteria of lexical distinction (as measured by functional load) and predictability of distribution were calculated in order to establish the strength of the contrast between vowel pairs in LF. It was predicted that High Contrast vowels [a]-[ɔ] should yield higher accuracy scores, faster RTs and be judged least similar; Low Contrast vowels [y]-[ʏ], which are in allophonic relationship, were predicted to result in lower accuracy scores, slower RTs and be judged most similar; Mid Contrast vowels [o]-[ʊ] were predicted to fall between the other two vowel pairs in terms of results. A summary of predictions and results for Experiments 3 through 5 are provided in Table 10.2. A "✓" symbol indicates the

results fully supported the prediction in that column; an “x” indicates the results did not support the prediction; a “(✓)” symbol in parentheses indicates the prediction was partially supported by the results obtained.

Table 10.2 Summary of Results for Experiments 3, 4 and 5: Gradient Contrast vs. Binary Contrast vs. Non-Speech

Experiment	Results	Prediction 1: Gradient Contrast H>M>L	Prediction 2: Binary Contrast H=M>L	Prediction 3: Non-speech Acoustics L>H>M
3. AX	Accuracy	(✓)	x	x
	RTs	(✓)	x	x
4. 4IAX	Accuracy	(✓)	x	x
	RTs	(✓)	x	x
5. Similarity Rating	Ratings	✓	x	x

For Experiments 3 and 4, the results did not perfectly match either prediction because there were only significant differences found between certain pairs. For Experiment 5, the results from the similarity rating task supported the predictions based on gradient contrast. The details of the support and counter-evidence found for the three different predictions in Table 10.2 are discussed in turn below.

10.2.1 Evaluating Prediction 1: Gradient

For the results to be in line with a gradient view of contrast, one would expect significant differences between each of the High, Mid and Low Contrast conditions. In Experiment 3, an AX task, results from both accuracy and RTs yielded significant difference between High and Mid conditions, and RTs showed an additional

significant difference between High and Low conditions. There was, however, no significant difference in accuracy scores between High and Low conditions (p=.083). The only significant difference between Mid and Low pairs was found among Same trials in RTs only. Overall, results patterned according to the predictions based on a gradient view of contrast: in absolute values, accuracy was highest for the High Contrast condition, followed by Mid, then Low; in terms of RTs, response times were fastest for the High condition, followed by Mid, followed by Low. In Experiment 4 (4IAX task), for both accuracy and RTs, High pairs yielded significantly more accurate and faster results than both Mid and Low pairs, while no differences were found between Mid and Low pairs, corroborating what was found in Experiment 3 (H>M=L).

There is no theoretical motivation for the (traditionally contrastive) Mid pairs with [o-ʊ] vowels to pattern with (traditionally allophonic) Low pairs with [y-ʏ] vowels, as was found in Experiments 3 and 4. It is possible that there was not enough of a difference between Mid and Low pairs in terms of the measures used to quantify contrast to set them far enough apart on the scale from contrastive to allophonic to expect different results. In other words, it may be that High and Low pairs are better representations of the extremes of the scale, and the Mid pairs selected for these experiments happened to fall closer to the Low Contrast end of the scale, rather than being an ideal representation of the mid-point of the scale. This suggests that there may be a minimum difference required in the measures of contrast for two sounds to be perceived as contrastive. Recall that in Chapter 6, for the monosyllables surveyed, the number of minimal pairs between the High [a]-[ɔ]

vowels was 86 while for the Mid [o]-[ʊ] vowels, it was 22 and for the Low [y]-[ʏ] vowels, it was 0, meaning that for the criterion of lexical distinction, Mid vowels were not equidistant between High and Low for this measure and were closer to the Low end of the scale. With regards to relative frequency of the individual vowels, High Contrast vowels [a] and [ɔ] occurred 305 and 202 times, respectively in the monosyllables surveyed, while Mid vowels [o] and [ʊ] occurred 154 and 110 times respectively, and Low vowels [y] and [ʏ] occurred 63 and 76 times respectively, meaning that the difference in frequency between High and Mid Contrast vowels is greater than between Mid and Low Contrast vowels (in absolute terms). This may explain why across Experiments 3 and 4, the High condition was significantly different from both Mid and Low conditions, while the Mid and Low conditions were never significantly different from one another. While the result of Mid and Low conditions patterning together was initially unexpected, the differences between High and Mid pairs as well as High and Low pairs provide partial support to a gradient view of contrast.

It is also possible that High vowels were set apart for other reasons. For example, the [a]-[ɔ] pair involves the only non-round stimulus of the set of vowels used. This could cause the pair to be easier to distinguish acoustically. In addition, if one were to simply count features, [a]-[ɔ] differ along the dimensions of height, rounding and tenseness; [o]-[ʊ] differ in height and tenseness; and [y]-[ʏ] differ in tenseness. Simply counting featural differences would lead to the expectation that stimuli pairs would fall in the same order as predictions based on functional load and predictability of distribution. While this would explain what may set the High pair

apart from Mid and Low pairs, it does not explain, however, why Mid and Low would pattern together in Experiments 3 and 4. It is also possible that the perception of gradience is not mono-factorial, so that, for example, where there is greater *un*predictability in distribution, functional load could become more important to processing.[25]

In Experiment 5, a similarity rating task, it was found that the High Contrast stimuli were judged as being the least similar; Low Contrast (allophonic) stimuli were judged as being the most similar; and ratings for Mid Contrast stimuli fell between the other two pairs. All differences were statistically significant. While Boomershine et al. (2008) used a five-point Likert scale and a 1000 ms interstimulus interval (ISI) and the present study used a six-point Likert scale and a 1500 ms ISI, both studies show that phones in an allophonic relationship were perceived to be more similar than those in a phonemic relationship. The Boomershine et al. study did not, however, test segments in an intermediate relationship and therefore only presents evidence from two extremes of the scale of possible contrasts. Experiment 5 in the present study included stimuli from three strengths of contrast, and the clearest evidence that these different strengths of contrast affect speech perception was found in this experiment.

Considering that the results from Experiment 5 showed such clear differences between the three levels of contrast, it is not immediately apparent why these same differences were not mirrored in Experiments 3 and 4, or, conversely, why – if

[25] Thank you to Kathleen Currie Hall for a discussion regarding possible factors influencing the results (Personal Communication, 2015).

minimal pair counts and relative frequency make Mid pairs closer to Low pairs on the continuum of contrast and allophony – Mid and Low pairs did pattern together in Experiment 5. The cause may lie in the nature of the task being asked of the participants. In a similarity rating task, the judgments made tend to be relative in nature, which is why some of the similarity-rating literature has discussed the issue of how large of a Likert scale to use, or whether there should be an even or odd number of ratings to choose from (Colman, Norris, & Preston, 1997; Dawes, 2008). In a pilot of this study, it was found that leaving Same pairs in the task caused participants to rate all Same pairs as 6 and typically use 1 or 2 for all other pairs, essentially causing participants to perform an AX task and classifying all stimuli as either Same or Different instead of placing them on a scale of similarity from 1 to 6. Upon removing Same stimuli from the experimental design, participants used more of the scale, although some tended to use the lower end while others tended to use the upper end, and yet others made full use of the scale. Indeed, normalizing results into Z scores is a direct acknowledgment of the relative nature of such a task, since it is considered more representative of the perception process to accentuate differences *between* conditions than the average rating conditions were allotted on the scale of Same to Different. Since all stimuli heard by participants were Different, participants would instinctively avoid classifying all trials as “1” (“Different”), and instead rate the similarity between stimuli in a relative manner: High vowels were perceived as more different than Mid vowels, and Low vowels were perceived as more similar than the other two conditions. In other words, the ratings represent not only how different one stimulus in a single pair is to its mate, but also how different stimuli pairs are from one another within a single stimuli set.

In addition, results for Experiment 5 may have been more improved than for Experiments 3 and 4 since Experiment 5 was always the final task to be performed by participants. Participants would have been more familiarized to the stimuli by the time they undertook this task, which may have contributed to what appears to be improved perception. Be that as it may, it is also telling that Mid Contrast vowels [ɒ] and [ʊ] were not perceived to be as different as High Contrast vowels [a] and [ɔ], since both these conditions contain vowels that are contrastive in LF. For the non-speech prediction, by which participants would have perceived and classified vowels solely according to their acoustic properties, [a-ɔ] should have been perceived as more similar than [y-ʏ], which was not the case. The fact that participants' results divided cleanly into three categories of pair type suggests that, at the very least, they did not perceive stimuli according to strict binary relationships (i.e. contrastive vowels versus allophonic vowels). The confluence of evidence from Experiments 3 through 5 therefore provides support for a gradient view of contrast.

10.2.2 Evaluating Prediction 2: Binary Contrast

Although results from Experiments 3 and 4 do not perfectly support the prediction based on a gradient view of contrast with no significant differences between Mid and Low conditions, they do go against a purely binary view of contrast where a relationship can be considered contrastive so long as one criterion for contrast is satisfied (such as lexical distinction). For the purely binary view to have been supported, there should have been no difference between High and Mid Contrast conditions. Recall that there was no distinction found between different types of contrastive segments in Experiments 1 and 2, with contrastive test stimuli

patterning exactly as contrastive control stimuli. However, this raises an interesting point about the impact of experimental design on results: when minimal pair counts and relative frequency are not incorporated into the design, such as in Experiments 1 and 2, results appear as though contrast is an all-or-nothing relationship; when minimal pair counts and relative frequency *are* incorporated into the experimental design, differences between sets of contrastive stimuli can be found. The evidence from Experiments 1 through 4 converges to suggest that a purely binary view of contrast does not hold.

In addition to the results from Experiments 3 and 4, the results from Experiment 5 provide further evidence against a strictly binary view of contrast. In terms of similarity ratings, if the binary view of contrast held, it was predicted that the High and Mid Contrast vowels would have been perceived as equally different or similar as compared to the allophonic Low vowels. However, results showed that the three vowel pairs were classified in distinct ranges of similarity, with High Contrast vowels being perceived as more different from one another than Mid Contrast vowels, despite the fact that both pairs are considered contrastive under a binary view. If contrast and allophony were the only possible phonological relationships, and if participants perceived pairs of sounds solely in those terms, then results would have yielded one group of ratings for High [a-ɔ] and Mid [o-ʊ] vowels, and another group of ratings for Low [y-ʏ] vowels. Since this scenario did not come to pass, the results appear to allow for gradience in contrast.

10.2.3 Evaluating Prediction 3: Non-Speech

A third prediction for results was based on the weighted acoustic distances between vowels as measured and reported in Chapter 6: if participants were only judging stimuli based on minute acoustic differences as they would a non-speech task, results should have shown that Mid vowels [o-ʊ] were the most similar, followed by High vowels [a-ɔ] and then Low vowels [y-ʏ] as per the weighted acoustic differences detailed in Chapter 6; however, this was not borne out in any of the experiments. Rather, results patterned more consistently according to the gradient view of contrast as determined by minimal pair counts and relative frequency.

10.3 Comparing Current Findings to Previous Findings in the Literature

The results of Experiments 1 and 2 corroborate what has been found in previous literature regarding purely allophonic and contrastive relationships: phones in an allophonic relationship are more difficult to perceive than those in a contrastive relationship (Boomershine et al., 2008; Dupoux et al., 1997; Ettlinger & Johnson, 2009; Johnson & Babel, 2010; Kazanina et al., 2006; Peperkamp et al., 2003; Pruitt et al., 2006).

The ends of the continuum tested in Experiments 1 and 2 map to contrastive relationships (High) and allophonic relationships (Low) in Experiments 3 through 5. Focusing solely on results from the ends of the scale, results were very similar in both AX tasks in Experiments 1 and 3, in terms of accuracy scores (approximately 90% and above) as well as RTs (about 800 ms for contrastive pairs (High) and between 910-970 ms for allophonic pairs (Low)). For the 4IAX tasks in Experiments

2 and 4, accuracy scores and RTs were again comparable for like conditions (i.e. contrastive segments as compared to other contrastive segments and allophones as compared to allophones), with lower accuracy and slower RTs for allophonic stimuli.[26] Lastly, in Experiment 5, there were higher similarity ratings for the contrastive pairs (High) compared to the allophonic pairs (Low). These results show that for the extremes of the scale of allophony and contrast, findings are remarkably consistent across studies. With regards to specific results, it is difficult to compare the previous literature with the present study due to multiple experimental paradigms: different task types, different ISI between stimuli, different stimuli, different experimental presentation software, and different languages being studied. On the other hand, it may be all the more remarkable that despite all these differences, the same general result is obtained over and over, with sounds in an allophonic relationship being difficult to perceive than those in a contrastive one.

The only other study to quantify and test intermediate phonological relationships by experimental means (a similarity rating task) is Hall (2009). Using predictability of distribution as the main criterion as represented by the entropy of the segments tested, Hall (2009) tested four pairs of German consonants exhibiting different levels of predictability of distribution. She hypothesized that pairs with greater predictability would be perceived as more similar. The results were inconclusive, and Hall (2009) lists the potential causes for this as being a combination of: (a) the power of the experiment being too low, (b) entropy values between pairs being too close to one another, (c) raw acoustic differences between stimuli, (d) the way in which entropy

[26] Note, however, that due to the design in Experiment 2, they are not directly comparable since Experiment 2 had three response options and Experiment 4 had four response options.

was calculated, (e) frequencies of specific sounds in the corpus surveyed, and (f) differences in phonotactic licitness of the contexts in which the consonants occurred. As such, it is difficult to compare present results with Hall's (2009) study in a meaningful way.

The present study is the only study to explore intermediate phonological relationships across three different experimental paradigms, using stimuli that are close in acoustic similarity, and using multiple measures to represent the criteria used to determine phonological relationships. It therefore confirms previous studies where differences were found between the perception of sounds in a contrastive relationship as compared to an allophonic relationship. Moreover, it presents new data supporting the theory-based hypothesis that there are phonological relationships between these two extremes.

10.4. Results Across Different Experimental Paradigms

While the above-mentioned studies remain consistent in their findings across multiple paradigms, there are many factors that differ between studies whether in regards to stimuli, task type, experimental design, and other factors, making a direct comparison across studies difficult. One strength of this dissertation is that no other study has explored the perception of contrast using three different experimental paradigms – an AX task, a 4IAX task, and a similarity rating task, with the same stimuli and same participants. This is important because using three experimental paradigms (Experiments 3 to 5) to explore a question can provide insight as to what different types of paradigms can offer by way of experimental capabilities, and where to expect to find differences and similarities across results.

The fact that three different paradigms were used and yielded consistent results suggests that results are not a by-product of experimental paradigm, and significant findings are attributable to linguistic factors. While the differences between levels of contrast in Experiments 3 and 4 were not all statistically significant (High was significantly different from Mid and Low, but Mid was not significantly different from Low), they were consistent with the findings in Experiment 5 (similarity rating task) in raw data: High Contrast pairs resulted in the highest accuracy scores and fastest RTs; Low Contrast pairs resulted in the lowest accuracy scores and slowest RTs; Mid Contrast pairs fell between the other two levels of contrast. It is possible that with a greater number of participants, results would have reached significance in the AX task, or as mentioned above, if vowels for the Mid Contrast better represented an equidistant midpoint on the scale of contrast to High and Low vowels. Nevertheless, the consistency across paradigms speaks to the fact that no one result was by chance.

Since no other studies directly compare results from both AX and 4IAX tasks (Experiments 1 and 3 vs. Experiments 2 and 4), it should be noted that due to differences between trial types in each paradigm, it is unclear whether results from each paradigm should be expected to match or be expected to differ. For example, since differences between Quad Type in Experiment 4 (Different-Different vs. Mixed vs. Same-Same) were not statistically significant, results were collapsed across this condition and so stating that differences between Contrast conditions were found in the AX task refers to only Different pairs, while differences found in the 4IAX task refer to Contrast across Quad Type. On closer examination of Quad Type, the

results for Different-Different quads appear to match the results of Different trials in the analogous AX task (Experiment 3), with higher accuracy on High Contrast stimuli, followed by Mid Contrast, followed by Low Contrast. However, a similar comparison with regards to Same trials show remarkable consistency across Contrast conditions in the AX task, with High, Mid and Low conditions yielding 98, 96 and 98 percent correct responses, respectively, while in the 4IAX task, Same-Same quads decreased from 93, to 88 to 86 percent correct, respectively. It is not clear why trials with Same stimuli would not yield approximately equivalent results regardless of paradigm, but since these differences were not large enough to cause a significant interaction between Contrast and Quad Type in Experiment 4, it is uncertain whether such discrepancies represent differences that are truly attributable to the different paradigms.

The above observations are only among accuracy scores. With regards to RTs, results were more consistent, patterning exactly the same for Different pairs (AX paradigm) and Different-Different quads (4IAX paradigm), as well as Same pairs and Same-Same quads. Since no previous study on contrast has analyzed RTs from a 4IAX task, there is nothing to compare these results with within the same paradigm, and so in this respect, it is encouraging that RTs between AX and 4IAX task appear to reflect one another, even with slightly less power in the 4IAX task.

Overall, the consistencies between AX and 4IAX tasks in experiments 1 and 2 confirmed that there were indeed differences found when testing only the extremes of the scale of contrast, with participants showing signs of difficulty in processing segments in an allophonic relationship. The AX and 4IAX tasks in Experiments 3

and 4 provided evidence against a purely binary view of contrast, and the similarity rating task in Experiment 5 provided the clearest evidence of gradience of contrast, with allophones being judged more similar than phones in an intermediate relationship, which in turn were judged more similar than those in a contrastive relationship. From these comparisons, we can be more confident moving forward that results from comparable AX and 4IAX studies should match: differences found in one should be echoed in the results of the other. This confirms that results are representative of linguistic factors affecting the speech perception of participants and are not simply experimental artefacts.

10.5. Theoretical Implications

How can these findings be incorporated into current theoretical frameworks used to define and describe phonological relationships? As discussed in the review of the literature in Chapter 2, classifying segments as contrastive or not can influence how a phonological analysis proceeds. When segments contrast in some contexts and do not in others, this can create disagreement about whether or not those segments should be included in an underlying phonemic inventory, such as was the case with Japanese affricates (described in Larson-Hall, 2004). Furthermore, determining the set of underlying phonemes in an inventory is often a first step to determining what features are active in a language's phonological processes, and so this can impact how feature sets and specifications are determined as well, which are critical elements in any analysis of speech patterns.

Cohn (2006) explores various aspects of gradient phonology and suggests that often the "grey areas" of determining what is phonological in a language is due to

difficulties in drawing a line between the traditional generativist modules of phonetics and phonology. For example, lengthening of vowels before voiced consonants in English is systematic, but it is unclear whether a length distinction between vowels has been phonologized or if this lengthening is more properly the domain of phonetics. Cohn argues that whether there needs to be a line drawn between phonetics and phonology should be an empirical question, determined by which approach provides the best fit for the range of more categorical to more gradient phenomena. Indeed, a modular view of phonology and phonetics, as well as a modular view to what constitutes contrast and allophony, has been found to be inadequate to describing phenomena which fall between one and the other (see Hall, 2013, for an extensive list). Cohn (2006) provides the example of the /a/-/ɔ/ contrast in American English is only found before coronals and in open syllables and raises the issue of the nature of the relationship between these sounds before non-coronals. Cohn also raises an issue with contrasts with a low functional load as in /θ/ and /ð/ in English which distinguish between relatively few lexemes (e.g. ‘thigh’ vs. ‘thy’) and asks whether contrast is “realized the same way in these cases as in the more robust cases? Or should contrast also be understood as a gradient property?” (p. 34).

One goal of this dissertation was to answer these questions by means of a quantitative method for determining phonological relationships that would have the potential to move phonological analysis towards a more nuanced approach to analyzing the world’s sound systems – an approach that supports a gradient model of contrast in phonology. Theories that do not subscribe to a strict interpretation of a

binary view of contrast, or even to the phoneme as a phonological unit, are supported by the results obtained in this study since the measures of lexical distinction and predictability of distribution used here were found to capture not only the extremes of purely allophonic and purely contrastive relationships, but also provide evidence that relationships can fall between these extremes. Exemplar theory allows for variability in phonological relationships and no other framework is able to incorporate usage effects and translate these into expectations on the behaviour of its various phonological units. One potential consequence of this methodology of evaluating contrast would be to have larger phonemic inventories such that surface contrasts – for example, [t^s]-[d^z] in LF in words such as “tu” [t^sy] ‘you.sg.’ vs. “du” [d^zy] ‘some/of’ – would cause those sounds to be included where before they were not. Ultimately, even if such contrasts were to be included, questioning into the phonological relationships between sounds cannot stop at a yes or no answer regarding contrast. Contrast should be evaluated on a scale, schematized in Figure 10.1, where the arrow represents the scale of strengths of contrast, the lowest possible strength of contrast is pure allophony (on the left) and the upper limit to the possible strength of contrast (on the right) depends on the measures discussed here, namely of functional load and predictability of distribution of a given sound pair.

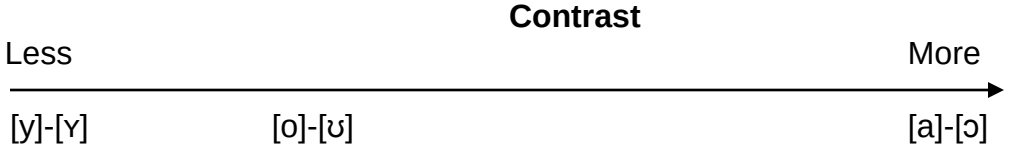


Figure 10.1 Scale of phonological relationships

This depiction also captures phonological relationships based on other factors (such as lexical distinction) used to determine phonological relationships, where purely allophonic relationships would be on one end and purely contrastive relationships on the other. Categories are partly created by virtue of the relationships between them, among other influences, and so robustness of category and the nature of the relationship between categories go hand in hand. The results in this dissertation show that between less robust categories (i.e. with lower minimal pair counts and lower frequency), relationships are weaker (i.e. towards the allophonic end of the scale) and these categories depend to a greater degree on surrounding environments to determine their surface form. Conversely, robust categories participate in stronger relationships with each other as well as other sounds, towards the contrastive end of the scale. Knowing what makes a category more or less robust and using a more objective metric to measure this robustness allows us to evaluate the phonological relationships between categories.

Based on the current findings from the present study, it should be possible to apply the same measures to segments that occur in any language and arrive at comparable results. One would predict, therefore, that speakers of another language with different distributions and functional load with regards to the segments tested in this dissertation would yield results that represent the relationships between these segments in their own language. For example, speakers of French from other dialects and for whom [y] and [ʏ] are not in allophonic relationship should yield

different results from LF speakers.[27] Applying this methodology in reverse, it may even have the potential to be used as a diagnostic for phonological relationships, where one could examine perceptual results and determine the relationship between segments based on the accuracy scores and RTs obtained. As in Hall (2008), such results could also be used to determine competency in another language, where measures such as perceived similarity reflected a shift from an allophonic L1 relationship to a contrastive L2 relationship between the same L1 segments.

10.6. Limitations and Future Research

While the success of the quantitative measures in this dissertation was not perfect, this may reflect differences in experimental paradigm rather than a lack of correlation between the measures used here and the results. The 4IAX task used here included four response options and an increased cognitive load as compared to the AX tasks, which could account for differences in the significance of the results. Minimally, a baseline has been established by which to compare future studies across experimental paradigms due to the similarities and differences found in results from AX to 4IAX task. Furthermore, as mentioned above, the fact that the similarity-rating task (Experiment 5) was always presented last in the sequence of experiments could explain the apparent improvement in participant's results; alternatively, a similarity-rating task may simply reflect relative judgments between sets of stimuli, in addition to judgments between syllables in a pair of stimuli.

[27] Note, however, that for speakers who do not have [Y] in their phonetic inventory, this would be akin to an L2 study where novel speech sounds are tested.

The next step for future research would be to test pairs of sounds that exhibit different types of intermediate relationship. For example, in the present study, the Mid [o-ʊ] pair is not allophonic according to all phonological frameworks (i.e. context never determines which vowel will appear between contexts where they are allowed to appear, and they serve to distinguish between morphemes). In future studies, pairs such as [o-ɔ] in LF should also be tested, where the pair of sounds sometimes distinguishes between lexemes as in “saute" [sot] ‘jump’ and “sotte” [sɔt] ‘idiot.fem.’, and are predictably distributed in contexts such as “sot” [so] ‘idiot.masc.’ and “sotte”, with [o] in an open syllable and [ɔ] in a closed syllable.[28] In other words, [o] is allowed in word-final position while [ɔ] is not. I am not proposing that the traditional criteria of lexical distinction and predictability of distribution (among other criteria) be done away with all together; simply, once it has been determined whether such criteria are applicable or not, it is then determined *to what degree* criteria apply in a given language to any given pair of segments.

In summary, this dissertation aimed to provide experimental evidence for what is being more frequently acknowledged in the theoretical literature, namely that there are phonological relationships that fall between purely allophonic or purely contrastive. An all-or-nothing view has proven problematic in analyses where some criteria for contrast are satisfied while others are not, or where one particular criterion is satisfied but only to a certain extent. The resulting ambiguities in phonological status may be resolved by using quantifiable measures for the criteria traditionally used to evaluate phonological relationships, and in doing so, better

[28] Note that the morphological relatedness of the words “sot” and “sotte” is part of what motivates calling this context-specific distribution of [o] and [ɔ] allophony.

represent the range of relationships between categories of speech sounds and further our understanding of sound patterns in human language.

References

Aliaga-García, C. & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production (*pp. 2-11). Newcastle upon Tyne: Cambridge Scholars Publishing.

Archangeli, D. (1988). Aspects of Underspecification Theory. *Phonology 5*, 183-207.

Avery, P., Dresher, E. & Rice K. *Contrast in Phonology.* Berlin: Mouton de Gruyter.

Best, C.T., McRoberts, G.W., & Sithole, N.N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental psychology: Human Perception and Performance*, 14, 345-360.

Bishop, D.V.M., Adams, C.V., Nation, K., & Rosen, S. (2005) Perception of transient non-speech stimuli is normal in specific language impairment: evidence from glide discrimination. *Applied Psycholinguistics*. 26, 175-194.

Boomershine, A., Currie Hall, K., Hume, E., & Johnson, K. (2008). The Impact of Allophony versus Contrast on Speech Perception. In P. Avery, E. Dresher, & K. Rice (Eds.), *Contrast in Phonology* (pp. 143-172). Berlin: Mouton de Gruyter.

Brannen, K. (2002). The Role of Perception in Differential Substitution. *Canadian Journal of Linguistics / Revue canadienne de linguistique, 47(*1/2), 1-46.

Brown, C. 1997: Acquisition of segmental structure: consequences for speech perception and second language acquisition. Unpublished PhD dissertation, McGill University, Montreal, Quebec.

Brown, A. (1988). Functional Load and the Teaching of Pronounciation. *Tesol Quarterly*, 22(4), 593-606.

Bybee, J. (2001). *Phonology and Language Use.* Cambridge: Cambridge University Press.

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language, 82*(4): 711-733.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403,* 428-430.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* New York, NY: Harper & Row.

Coh, A. (2006). Is there gradient phonology? In G. Fanselow, Fery, C., Vogel, R., & Schlesewsky, M. (Eds.). *Gradience in grammar: Generative perspectives* (pp. 25-44). Oxford: Oxford University Press.

Colman, A. M., Norris, C. E., & Preston, C. C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports, 80*, 355-362.

Côté, M.-H. (2012). Laurentian French (Quebec). In R. Gess, C. Lyche, & T. Meinsburg (Eds.), *Phonological Variation in French. Illustrations from Three Continents* (pp. 235-274). Philadelphia, PA: John Benjamins.

Curtin, S., Goad H. & Pater, J. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 389-405.

Dawes J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. International Journal of Market Research, 50(1), 61-77.

de Bree E, Zamuner TS, Wijnen F. (2014). Phonological neighbourhoods in the vocabularies of Dutch typically developing children and children with a familial risk of dyslexia. In R. Kager, J. Grijzenhout, & K. Sebregts (Eds.), Where the Principles Fail. A Festschrift for Wim Zonneveld on the Occasion of his 64th Birthday (pp. 17-28). Utrecht: Utrecht institute of Linguistics OTS.

Dresher, B. E. (2009). The contrastive hierarchy in phonology. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in phonology: Theory, perception, acquisition* (pp. 11-33). Berlin: Mouton de Gruyter.

Dresher, B. E., Piggott, G. & Rice, K. (1994). Contrast in phonology: Overview. In C. Dyck (Ed.), *Toronto Working Papers in Linguistics* (pp. iii-xvii). *13*.

Eckman, F. R., Iverson, G. K., Fox, R. A., Jacewicz, E., & Lee, S. (2009). Perception and production in the acquisition of L2 phonemic contrasts. In M. A. Watkins, A. S. Rauber & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 81-95). Cambridge: Cambridge Scholars Publishing.

Ettlinger, M. & Johnson, K. (2009). Vowel discrimination by English, French and Turkish speakers: Evidence for an exemplar-based approach to speech perception. *Phonetica, 66*, 222-242

Gerrits, E. (2001). *The categeorisation of speech sounds by adults and children* (Doctoral dissertation).

Gerrits, E. & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics, 66*(3), 363-376.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166-1183.

Goldsmith, J. (1995). Phonological Theory. In J. A. Goldsmight (Ed.), *The Handbook of Phonological Theory* (pp. 1-23). Cambridge, MA: Blackwell.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Graham, L., & House, A. (1970). Phonological oppositions in children: a perceptual study. Journal of the Acoustical Society of America 49, 559-566.

Gussenhoven, C., & Jacobs, H. (1998). *Understanding phonology*. London: Arnold.

Hall, D. C. (2007). *The role and representation of contrast in phonological theory* (Doctoral dissertation).

Hall, K. C. (2008). *Testing an exemplar-based model of contrast and allophony against evidence from second language acquisition.* Ohio State University, Columbus, OH. Retrieved from http://www.ling.ohio-state.edu/~kchall/Hall_Phonology_submission.pdf

Hall, K.C. (2009). *A probabilistic model of phonological relationships from contrast to allophony* (Doctoral dissertation).

Hall, K. C. (2012). Phonological relationships: A probabilistic model. In A. McKillen & J. Loughran (Eds.). *Proceedings of "Phonology in the 21st Century: In Honour of Glyne Piggott"*. McGill Working Papers in Linguistics 22. Montreal, QC. Retrieved from http://www.mcgill.ca/mcgwpl/files/mcgwpl/curriehall2012.pdf

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review, 30*(2), 215-275.

Hualde, J. I. (2005). Quasi-phonemic contrasts in Spanish. In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez & Benjamin Schmeiser (Eds.). *Proceedings of the 23rd West Coast Conference on Formal Linguistics* (pp. 374-398). Somerville, MA: Cascadilla Press.

Hume, E., Hall, K.C., Wedel, A., Ussishkin, A., Adda-Decker, M., & Gendrot, C. (2013). Anti-markedness patterns in French epenthesis: An information-theoretic approach. In C. Cathcart, I.-H. Chen, G. Finley, Kang, S., Sandy, C.S., & E. Stickles (Eds.). *Proceedings of the thirty-seventh annual meeting of the Berkeley Linguistics Society* (pp. 104-123). Berkeley: Berkeley Linguistics Society.

Hura, S., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping assimilation rules. *Language and Speech*, *35*(1, 2): 59-72.

Ito, J. & Mester, R.A. (1995). Japanese phonology. In John Goldsmith (Ed.). The handbook of phonological theory (pp 817-838). Cambridge: Blackwell: 817–38.

Jesney, K. (2005). Stridency and Differential Substitution in Two Dialects of French. Handout from presentation at the *Language Research Centre Graduate Student Forum*, Calgary, AB.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). San Diego, CA: Academic Press.

Johnson, K. & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics, 38*, 127-136.

Kuhl, P. & Iverson, P. (1995). Linguistic Experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.121-154). Baltimore, MD: York Press.

Ladefoged, P. (1975). A course in phonetics (1st ed.). New York, NY: Harcourt Brace Jovanovich.

Lahiri, A, & Marslen-Wilson, W.D. (1991). Themental representation of lexical form: Aphonological approach to the recognition lexicon. Cognition, 38, 245–294.

Larson-Hall, J. (2004). Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Language Research, 20*(1), 33-76.

Lee, S. and Iverson, G. (2012). Stop consonant productions of Korean-English bilingual children. *Bilingualism: Language and Cognition*, *15*(2), 275-287.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences,* Prague 1967, Academia Publishing House of the Czechoslovak Academy of Sciences.

Lombardi, L. (2003). Second language data and constraints on Manner: Explaining substitutions for the English interdentals. *Second Language Research,* 19(3), 225-250.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

McGuire, G. (2007). *Phonetic Category Learning* (Doctoral dissertation). Retreived from ProQuest Dissertations and Theses database. (UMI No. 3279823). Ohio State University, Columbus, OH.

McGuire, G. (2010). A Brief Primer on Experimental Designs for Speech Perception Research. Unpublished Draft, Santa Cruz: University of California, Santa Cruz, Summer 2010. Retrieved from http://people.ucsc.edu/~gmcguir1/experiment_designs.pdf

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The “soul” of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846-854.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 203-229.

New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004) Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 516-524.

Noordenbos M.W., Segers E., Serniclaes W., Mitterer H., & Verhoeven L. (2012a). Allophonic mode of speech perception in Dutch children at risk for dyslexia: a longitudinal study. *Research in Developmental Disabilities*, 33,1469–83.

Noordenbos M.W., Segers E., Serniclaes W., Mitterer H., & Verhoeven L. (2012b). Neural evidence of allophonic perception in children at risk for dyslexia. *Neuropsychologia*, 50(), 2010–7.

Peperkamp, S., Pettinato, M., and Dupoux, E. (2004). *Allophonic Variation and the Acquisition of Phoneme Categories.* Paper presented at the 27th Annual Boston University Conference on Language Development, Boston, MA.

Pierrehumbert, J. (2000) The phonetic grounding of phonology, *Bulletin de la Communication Parlée* 5, 7-23.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137-158). Amsterdam: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101–139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.). *Probabilistic Linguistics* (pp. 177-228). Cambridge, MA: MIT Press.

Pierrehumbert, J. (2006). The Next Toolkit. *Journal of Phonetics, 34,* 516-50.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception and Psychophysics, 13, 253-60.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition, 3*(1), 7-18.

Pisoni, D. B. & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15*(2), 285-290.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *Journal of the Acoustical Society of America,* 89(6), 2961-2977.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America, 119*(3), 1684-1696.

Saussure, F., Bally, C., Sechehaye, A., & Riedlinger, A. (1916). *Cours de linguistique générale*. Lausanne: Payot.

Scobbie, J. and Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In P. Avery, B. E. Dresher, & K. Rice (Eds.). *Contrast in phonology: Perception and acquisition* (pp. 87-114). Berlin: Mouton.

Steriade, D. (1995). Underspecification and Markedness. In John A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 14-74). Oxford: Blackwell.

Studdert-Kennedy, M., Shankweiler, D., & Pisoni, D. (1972). Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology, 3,* 3, 455-466.

Surendran, D. & Niyogi, P. (2003). Measuring the functional load of phonological contrasts. In *Tech. Rep. No. TR-2003-12.* Chicago. Retrieved from http://people.cs.uchicago.edu/~dinoj/research/FL22.pdf

Trubetzkoy, N. S. (1939). Grundzüge der Phonologie. *Travaux de Cercle Linguistique de Prague 7.* English Translation 1969: Principles of Phonology. Berkeley: University of California Press.

Twaddell, W. F. (1935). *On defining the phoneme*. Baltimore: Waverly Press.

Vance, T. (1987). An introduction to Japanese phonology. Albany, NY: State University of New York Press.

Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science*, 9(), 325-329.

Weinberger, S. (1990). Minimal segments in second language phonology. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the acquisition of second language speech* (pp. 137-173). New York: Mouton de Gruyter.

Werker J.F., Gilbert, J.H.V. & Tees, R.C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.

Werker J.F. & Tees, R.C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society of America, 75(6), 1866-1878.

Werker, J. F. & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*(1), 35-44.

Zamuner TS, Morin-Lessard E, Bouchat-Laird N. (2014 Forthcoming). Phonological patterns in the lexical development of Canadian French. In M. Yavas (Ed.), *Unusual Productions in Phonology: Universals and Language-Specific Considerations*.

Zamuner T.S. (2009). The structure and nature of phonological neighbourhoods in children's early lexicons. *Journal of Child Language*, 36, 3–21.

Appendix A: F1-F2 Measurements for Different Vowels

Note that measurements were taken from the steady-state portion of the vowel in Praat (Boersma & Weenik, 2014).

Stimulus	F1	F2	F3	F4
ʃʊʃ	430.50	1245.57	2768.30	3162.42
ʃʏʃ	484.95	1756.74	2540.34	3030.11
ʃaʃ	798.15	1675.32	2468.15	3527.52
ʃoʃ	558.06	966.77	2464.89	3271.49
ʃyʃ	321.60	1891.62	2535.44	3227.01
ʃɔʃ	579.72	1570.53	2461.16	3164.48
bʊb	398.36	915.94	2504.40	3279.92
bʏb	356.09	1694.55	2315.90	3136.46
bab	799.92	1640.71	2573.01	3375.95
bob	540.05	728.03	2743.03	3182.72
byb	304.34	1852.19	2586.67	3150.22
bɔb	577.18	1277.35	2546.67	3266.22
fʊf	462.18	888.45	2679.88	3391.71
fʏf	397.39	1778.16	2143.17	3197.55
faf	849.92	1550.68	2267.14	2999.71
fof	672.06	718.46	2497.94	3058.51
fyf	304.25	1980.27	2316.06	3313.62
fɔf	665.94	1215.33	2343.49	3172.06
lʊl	454.47	918.91	2421.90	3190.73
lʏl	386.68	1899.38	2199.56	3096.72
lal	843.46	1607.32	2454.92	3251.22
lol	542.35	898.33	2373.87	2994.44
lyl	347.86	1905.57	2481.78	3105.99
lɔl	543.72	1295.33	2402.16	3213.71

Appendix B: Description of Project for ISPR Participants

Research Project:	The Strength of Segmental Contrasts: A Study on Laurentian French
Principal Researcher:	Sophia Stevenson
REB Approval Code:	06-10-19B
Description:	Le but de cette étude est d'examiner la perception de la parole. L'étude portera spécifiquement sur l'effet des sons distinctifs d'une première langue sur la perception. Il s'agit de trois expériences de perception auditives où on vous demandera d'écouter et comparer plusieurs syllabes.
Duration:	60 minutes
Points:	2

Appendix C: Consent Form

Note that participants received copies of the consent form printed with University of Ottawa letterhead.

La puissance des contrastes segmentaux :
une étude du français laurentien

Formulaire de consentement

Chercheur Principal	Sophia Stevenson, candidate au doctorat Département de Linguistique, Université d'Ottawa
Coordonnées	70 avenue Laurier Est, pièce 401 Ottawa, ON K1N 6N5 Téléphone : Courriel :
Superviseur	Professeure Marie-Hélène Côté Département de Linguistique, Université d'Ottawa
Coordonnées	70 avenue Laurier Est, pièce 401 Ottawa, ON K1N 6N5 Téléphone : Courriel :

Invitation à participer: Je suis invité(e) à participer à une étude menée par Sophia Stevenson et supervisée par Dre Marie-Hélène Côté.

But de l'étude: Le but de l'étude est d'examiner la perception de la parole. L'étude portera spécifiquement sur l'effet des sons distinctifs d'une première langue sur la perception.

Participation: Ma participation à cette étude consistera à me présenter à une séance de une heure et demie, durant laquelle on me demandera de remplir un questionnaire, de participer à un court examen auditoire et trois expériences de perception, les résultats desquels seront enregistrés. Le questionnaire prendra environ quinze minutes et comportera des questions concernant mon expérience d'apprentissage de langues. Les expériences de perception seront complétées en trois temps, soit environ vingt minutes par section. Au cours de la première partie, on me demandera d'écouter des séquences de deux syllabes afin de signaler si j'entends les deux mêmes syllabes ou deux différentes. Au cours de la deuxième partie, on me demandera d'écouter quatre syllabes et de signaler si les deux premières ou les deux dernières sont pareilles. Finalement, on me demandera d'écouter deux syllabes et d'indiquer sur une échelle de 1 à 6 si elles sont similaires ou différentes.

Risques: Il n'y a pas de risques réels associés à cette étude. Cependant, j'aurai la possibilité d'interrompre l'étude et de faire une pause si je ressens le besoin.

Confidentialité et anonymat: L'information que je partagerai restera strictement confidentielle. L'anonymat sera garanti par l'utilisation de codes au lieu des noms. Mon nom et toute autre

information qui pourrait m'identifier ne seront divulgués à personne et dans aucune publication. Si des extraits des résultats de certains participants sont cités, l'anonymat sera garanti par l'utilisation de codes pour référer à ces individus.

Conservation des données: Les données recueillies, c'est-à-dire le questionnaire et les résultats des exercices de perception, seront conservées pendant une période de 10 ans dans un endroit sûr dans le laboratoire psycholinguistique mené par Professeure Laura Sabourin ou bien dans le bureau de Sophia Stevenson au Département de Linguistique à l'Université d'Ottawa. Seul le chercheur et son superviseur auront accès à ces données. En plus d'être utilisées pour la présente étude, il est possible que les données recueillies soient utilisées lors de futurs projets de recherche.

Participation volontaire: Ma participation est volontaire et j'ai le droit de ne pas participer ou de me retirer de l'étude en tout temps sans subir de conséquences négatives. Si je décide de me retirer de l'étude, toute information à mon sujet sera effacée et retirée complètement de l'étude.

☐ Comme on me l'a expliqué au début de la session, ma participation à cette étude me garantie ____ points dans mon cours, même si je décide de me retirer de l'étude.

Acceptation: Je, _____, accepte de participer à l'étude menée par Sophia Stevenson sous la supervision de Dre Marie-Hélène Côté.

Pour tout renseignement additionnel concernant cette étude, veuillez communiquer avec la chercheure ou sa superviseure. Pour tout renseignement sur les aspects éthiques de cette recherche, veuillez vous adresser au Responsable de l'éthique en recherche, Université d'Ottawa, Pavillon Tabaret, 550, rue Cumberland, pièce 159, (613) 562-5841 ou ethics@uottawa.ca.

Il y a deux copies. Veuillez en conserver une. Merci pour votre temps et considération.

_____ _____
Signature (participant) Date

_____ _____
Signature (chercheur) Date

Appendix D: Pre-Screening Questionnaire for ISPR participants only

Note: This questionnaire was used only to find potentially eligible participants and to screen out ineligible participants. The questionnaire was completed by participants on a different day before my study. Therefore it did not affect the cognitive burden of my data collection. The responses to the pre-screening questions are not reported in this dissertation.

The link to the complete questionnaire is found at url:
http://socialsciences.uottawa.ca/sites/default/files/public/psy/eng/documents/permanent_pre-screen_questions.pdf

Appendix E: Language Background Questionnaire (All Participants)

Note: Formatting has been changed to accommodate printing and official letterhead has been removed.

Brain and Language Laboratory / Laboratoire sur le cerveau et le langage

Questionnaire sur les antécédents linguistiques

1. Informations du participant: *(À être rempli par le chercheur)*

Code du projet: Date d'aujourd'hui: Numéro du participant:

2. Informations biographiques: *(À être rempli par le participant)*

3. Mois & année de naissance: Lieu de naissance:
[MOIS/ ANNÉE]

Genre: □ Masculin □ Féminin

Si vous avez vécu à d'autres endroits que votre domicile actuel, veuillez indiquer ces endroits de même que la période de temps (date de départ et d'arrivée) pendant laquelle vous y êtes demeuré. .

4. **De la naissance jusqu'à l'école: Informations biographiques sur vos parents/ gardien(nes):**

(CETTE SECTION CONCERNE CEUX QUI ONT SOIGNÉ DE VOUS PENDANT LA PÉRIODE ENTRE VOTRE NAISSANCE ET QUAND VOUS AVEZ COMMENCÉ L'ÉCOLE)

Premier parent/gardien(ne): Ville actuelle:

Langue(s) maternelle(s):
Durée/fréquence des contacts avec vous: Endroit de aissance:
Langue(s) de communication avec vous:

Si votre premier parent/gardien(ne) est né à un autre endroit ou a habité à un autre endroit que sa ville actuelle, veuillez indiquer ci-dessous où, quand et pour combien de temps votre parent/gardien(ne) a vécu à cet endroit.

Deuxième parent/gardien(ne): Ville actuelle:

Langue(s) maternelle(s):

Durée/fréquence des contacts avec vous: Endroit de naissance:
Langue(s) de communication avec vous:
Si votre deuxième parent/gardien(ne) est né à un autre endroit ou a habité à un autre endroit que sa ville actuelle, veuillez indiquer ci-dessous où, quand et pour combien de temps votre parent/gardien(ne) a vécu à cet endroit.

Autre parent/gardien(ne): Ville actuelle:

Langue(s) maternelle(s):

Durée/fréquence des contacts avec vous: Endroit de naissance:
Langue de communication avec vous:

Si votre autre parent/gardien(ne) est né à un autre endroit ou a habité à un autre endroit que sa ville actuelle, veuillez indiquer ci-dessous où, quand et pour combien de temps votre parent/gardien(ne) a vécu à cet endroit.

5. **Informations sur votre (vos) langue(s) maternelle(s):**

À quelle(s) langue(s) avez-vous été exposé(e)(s) dès votre naissance?

(1)
(2)
(3)

Compétences linguistiques actuelles:

(1) Langue:
Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

(2) Langue:
Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

(3) Langue:

Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

Commentaires généraux sur cette section:

6.**Informations sur les autres langues:**

Avez-vous appris ou étudié d'autres langues? □ Oui □ Non

Si non, passez à la Section 6, si oui, veuillez continuer cette section.

Pour chaque **autre** langue que vous avez apprise ou étudiée, veuillez répondre aux questions suivantes (s'il-vous- plaît demandez d'autres feuilles si nécessaire):

- Langue:
  Âge de la première exposition:
  Méthode d'acquisition (En classe et/ou naturellement):

- Si c'est en classe, veuillez indiquer quel type d'éducation vous avez eu (Complète/ totale, immersion, classes d'héritage, formation linguistique, etc.):

- Si c'est naturellement, veuillez expliquer:

Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

- Langue:
  Âge de la première exposition:
  Méthode d'acquisition (En classe et/ou naturellement):

- Si c'est en classe, veuillez indiquer quel type d'éducation vous avez eu (Complète/ totale, immersion, classes d'héritage, formation linguistique, etc.):

- Si c'est naturellement, veuillez expliquer:

Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle

Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

- Langue:
  Âge de la première exposition:
  Méthode d'acquisition (En classe et/ou naturellement):

- Si c'est en classe, veuillez indiquer quel type d'éducation vous avez eu (Complète/ totale, immersion, classes d'héritage, formation linguistique, etc.):

- Si c'est naturellement, veuillez expliquer:

Niveau de compréhension orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de production orale: □ Très bas □ Bas □ Intermédiaire □ Quasi-langue maternelle □ Langue maternelle
Niveau de compétence à l'écrit: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent
Niveau de compétence en lecture: □ Très bas □ Bas □ Intermédiaire □ Très bon □ Excellent

Commentaires sur les compétences:

Commentaires généraux sur cette section:

7.**Utilisation actuelle du langage:**

Veuillez indiquer quelle(s) langue(s) vous utilisez dans les contextes suivants. Si plus qu'une langue s'applique pour le même contexte, veuillez aussi indiquer le pourcentage d'utilisation pour chaque langue. (*À noter: Si vous utilisez plus d'une langue dans un contexte, votre réponse doit s'additionner à 100%)*

(4) **Langue(s) (%)**

Exemple: Avec des ami(e)s *Anglais (75%); Français (20%); Allemand (5%)*

- Avec votre mere _____
- Avec votre père _____
- Avec vos frère(s) et soeur(s)_____
- Avec votre famille élargie _____
- Avec votre partenaire _____
- Avec des ami(e)s _____
- À l'école _____
- Au travail _____

- ○ À l'épicerie _____
- ○ En écoutant la television _____
- ○ En lisant le journal _____
- ○ En lisant pour le plaisir _____
- ○ En voyageant (Comment souvent?) _____
- ○ Autre contexte (spécifiez): ○ Autre contexte (spécifiez): ○ Autre contexte (spécifiez):

Si vous avez d'autres commentaires à propos de votre utilisation du langage que vous pensez qui pourraient être pertinent, écrivez-les ci-dessous: *(Par exemple, si la langue utilisée avec vos parents ou vos frères et sœurs a changé au fil du temps)*

8.**Contexte d'acquisition du langage:**

Pour chaque langue que vous avez apprise, veuillez indiquer pour chacun des contextes suivants:

- les dates (ex. 1990-2000) ou les âges (ex. 5-8 ans) ou les années scolaires (ex. de la 8e à la 12e année) pendant lesquels vous l'avez apprise/utilisée;
- le nombre de mois ou semestres que vous avez passés à l'apprendre/utiliser, si pertinent;
- la fréquence d'utilisation/d'instruction (ex. 5 heures/semaine); et
- pour les contextes scolaires, si elle était la langue principale d'instruction ou si elle faisait partie des classes de français de base ou des classes d'immersion (complète, immersion à jeune âge, etc.)

S'il-vous-plaît, avertissez le chercheur si vous avez besoin de colonnes additionnelles.

Langue **Contexte**	**Français**			
Maison				
Garderie / École de petite enfance				
Maternelle				
École primaire				
École secondaire				
Collège				
Université				

École de langue				
Autre: (spécifiez)				

9. **<u>Autres informations linguistiques:</u>**

S'il-vous-plaît ajoutez toutes autres informations qui ont pu influencer votre utilisation d'une ou des langues que vous connaissez, ou qui permettraient de compléter votre historique linguistique.

_____

_____

_____

_____

Vous pouvez utiliser le verso de cette feuille si nécessaire. Merci!

Appendix F: Macbook Pro Details

Hardware Overview:

Model Name: MacBook Pro
Model Identifier: MacBookPro4,1
Processor Name: Intel Core 2 Duo
Processor Speed: 2.4 GHz
Number Of Processors: 1
Total Number Of Cores: 2
L2 Cache: 3 MB
Memory: 2 GB
Bus Speed: 800 MHz
Boot ROM Version: MBP41.00C1.B03
SMC Version (system): 1.27f3
Serial Number (system): W88141MDYJZ
Hardware UUID: C53A497B-2237-52A8-95D4-F1558F69F3F3
Sudden Motion Sensor:
State: Enabled

Apple Internal Keyboard / Trackpad:

Product ID: 0x0230
Vendor ID: 0x05ac (Apple Inc.)
Version: 0.70
Speed: Up to 12 Mb/sec
Manufacturer: Apple, Inc.
Location ID: 0x5d200000 / 3
Current Available (mA): 500
Current Required (mA): 40

MacPro KYBD Specifications

Hardware Overview:

Model Name: Mac Pro
Model Identifier: MacPro4,1
Processor Name: Quad-Core Intel Xeon
Processor Speed: 2.66 GHz
Number of Processors: 1
Total Number of Cores: 4
L2 Cache (per Core): 256 KB
L3 Cache: 8 MB
Memory: 6 GB
Processor Interconnect Speed: 4.8 GT/s
Boot ROM Version: MP41.0081.B07

SMC Version (system): 1.39f5
SMC Version (processor tray): 1.39f5
Serial Number (system): H09212754PD
Serial Number (processor tray): J591200L64MFC
Hardware UUID: A1053D18-9EEA-5A1C-A7DB-7CD8E83D5DA4

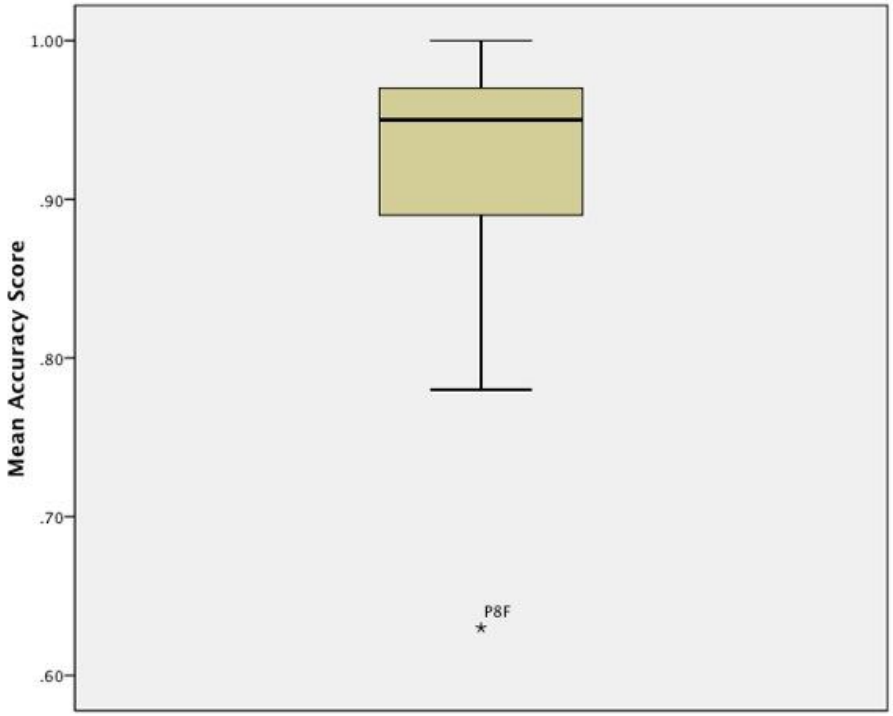
Hub in Apple Pro Keyboard:

Product ID: 0x1003
Vendor ID: 0x05ac (Apple Inc.)
Version: 4.20
Speed: Up to 12 Mb/sec
Manufacturer: Mitsumi Electric
Location ID: 0xfd330000 / 5
Current Available (mA): 500
Current Required (mA): 50

Apple Pro Keyboard:

Product ID: 0x020b
Vendor ID: 0x05ac (Apple Inc.)
Version: 4.20
Speed: Up to 12 Mb/sec
Manufacturer: Mitsumi Electric
Location ID: 0xfd333000 / 9
Current Available (mA): 250
Current Required (mA): 50

Appendix G: Boxplot for Mean Accuracy Scores for Experiment 3



This figure shows that, in terms of accuracy scores, participant P8F is an outlier compared to all other participants. For this reason, they were removed from the analyses for Experiment 3, and from 4 and 5 as well to make results directly comparable.