

RESEARCH

Gradient phonological relationships: Evidence from vowels in French

Sophia Stevenson and Tania Zamuner

University of Ottawa, 70 Laurier Ave. East, K1N6N5 Ottawa, CA

Corresponding author: Tania Zamuner (tania.zamuner@uottawa.ca)

The dichotomy of contrastive and allophonic phonological relationships has a long-standing tradition in phonology, but there is growing research that points to phonological relationships that fall between contrastive and allophonic. Measures of lexical distinction (minimal pair counts) and predictability of distribution were applied to Laurentian French vowels to quantify three degrees of contrast between pairs: high, mid, and low contrast. According to traditional definitions, both the high and mid contrast pairs are classified as phonologically contrastive, and low contrast pairs as allophonic. As such, a binary view of contrast (contrastive vs. non-contrastive) predicted that high and mid contrast pairs would pattern together on tasks of speech perception, and low contrast pairs would show a different pattern. The gradient view predicted all vowel pairs would fall along a continuum. Thirty-two speakers of Laurentian French participated in two experiments: an AX task and a similarity rating task. The results did not support a strict binary interpretation of contrast, since the high, mid, and low contrast vowel pairs pattern differently across the experiments. Instead, the results support a gradient view of phonological relationships.

Keywords: Contrast; gradient phonology; Laurentian French; speech perception; allophony

1 Introduction

The concept of contrast is at the heart of phonological analysis (Avery et al. 2008). In phonological theory, the relationships between speech sounds serve to differentiate words. The difference between the initial consonants of *fast* [fæst] and *vast* [væst], for example, signals a difference in meaning and distinguishes lexical items. Segments that distinguish between lexemes are considered to be in a contrastive relationship, and have traditionally been viewed as belonging to a stored inventory of underlying phonological representations. Segments that are not contrastive are in an allophonic relationship with each other. Recent research, however, suggests that the traditional dichotomy of contrastive versus allophonic phonological relationships is far more complex (Hall 2013). This is because the criteria that are commonly applied in phonological analyses to determine whether or not two sounds are contrastive do not account for intermediate phonological relationships, which fall between fully contrastive and fully allophonic. While the concept of gradient contrast is not new (Goldsmith 1995; Cohn 2006; 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 contrasts (see Hall 2013). Researchers have begun to re-examine the way phonological relationships are defined (Dresher 2008; Ernestus 2011; Hall 2009, 2013; Lu 2014; Hall 2015; Hall & Hall 2016), with the recent development of a model of contrast described along a continuum (Probabilistic Model of Phonological Relationships, PPRM;

Hall 2009, 2015). Despite this, there has been little research on how to precisely define phonological relationships. For what little experimental research exists testing phonological relationships, the results have varied in the literature. The goals of this research are to explore criteria for establishing phonological relationships, to apply these criteria to identify various degrees of contrast in Laurentian French (LF) vowels, and lastly, to test for evidence of gradient contrasts in two speech perception experiments. Laurentian French refers to dialects of French spoken in Canada excluding Acadian French (Côté 2012).

A typical phonological analysis begins by determining the relationships between speech sounds. In generative frameworks, contrast is often approached with an all-or-nothing view: two sounds either contrast or they do not, and gradience tends to fall under the domain of phonetics and not phonology (e.g. Chomsky & Halle 1968). Saying that two segments contrast indicates that they participate in a specific type of phonological relationship. It is taken to indicate that the two sounds are members of a phonological inventory and have distinct 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, in English, *leather* [lɛðɹ̩] and *letter* [lɛɹ̩] create a surface contrast between [ð] and [ɹ], where [ɹ] derives from /t/ intervocally when the first vowel is stressed (note that the realization of this pattern in English also depends on morphology and stress). However, because the realization of /t/ as [ɹ] can be explained, [ɹ] is not considered to be part of the phonemic inventory of English, even though it can be argued that two lexemes are distinguished by [ð] and [ɹ] (Boomershine et al. 2008). This is an example of a *surface contrast* where an apparent contrast exists but that one of the sounds involved is derived from an underlying phoneme and is not itself part of the phonemic inventory of the language. Because segments can appear to contrast in some word positions and not in others, creating surface contrasts such as the one above, the question has been raised as to whether contrast is binary and an all-or-nothing type of relationship, or gradient (Hall 2013).

The question of whether contrast is gradient or binary 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 a given analysis, as well as on the implications of what is assumed to be stored in an underlying representation. For example, exemplar theories, which allow for a gradient view of contrast due to the nature of phonetic categories, have different assumptions of what constitutes a category of speech sound and how the relationships between these categories are expressed and evaluated. Phonetic categories in an exemplar theory can be viewed as tokens of experience organized in a mental map of phonetic distributions, parameterized with acoustic, articulatory, and perceptual information (Pierrehumbert 2003). When similar remembered tokens reach a large enough number, this group is generalized and a category is formed (Pierrehumbert 2000, 2001; Bybee 2006). 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 (Pierrehumbert 2001; Bybee 2006). 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. This view of phonetic categories coincides with models where contrast is described along a continuum such as the Probabilistic Model of Phonological Relationships (Hall 2009, 2015) which focuses predominately on the continuum of predictability of distribution. The critical role that frequency plays in establishing phonetic categories and relationships between them will be captured in the measurable criteria used below to determine levels of contrast in our experimental stimuli, something which cannot be captured by a binary approach to contrast.

1.1 Criteria for contrastive relationships

There are multiple criteria that are typically used in a binary approach to contrast to determine whether two sounds are in an allophonic or contrastive relationship; however, the formulation and application of the criteria is not always clear and the default expectation is that there are only two types of phonological relationships. The most commonly used criteria to determine phonological relationships are outlined in Hall (2013: 223–225). Briefly, two sounds are typically considered contrastive if they define a lexical distinction, if they do not have a predictable distribution, or if they are written with different graphemes. Two sounds are allophonic if they participate in allophonic alternations conditioned by a specific phonemic environment, are judged to be the same sound by native speakers, and are written with the same grapheme. In addition to work by Hall, other authors have also begun to define contrast using a variety of phonetic and usage-based metrics, such as frequency and functional load, see work by Renwick (2014) and Renwick et al. (2016). The two most important and often-used criteria in a binary approach are *lexical distinction* (also called the distinctive function) and *predictability of distribution*. We refer to a “binary approach” because *how many* lexemes are differentiated or *how predictable* a distribution is typically ignored; as it pertains to phonological theory, all that matters is that at least one pair of lexemes is distinguished, or that a pair of sounds is unpredictably distributed to establish a contrastive relationship status. Thus, problems arise when the criteria conflict. For example, in Canadian Raising the diphthongs [aɪ] and [ʌɪ] are predictably distributed and often presented as allophones of a single phoneme. However, there are surface (i.e. not underlying) minimal pairs such as *title* [tʌɪl] and *tidal* [tʌɪr], where the diphthongs could be said to contrast before the flap [ɾ], which satisfies the criterion of lexical distinction (Hall 2012). On the one hand, one criterion for contrastive status is satisfied while on the other, a criterion for allophony is satisfied. Another example is found in Laurentian French. High tense vowels [i, y, u] become lax in closed syllables that do not end with [v, z, ʒ] and sometimes [ʁ] (Côté 2010), as in, for example, *petit* [pət̪^hi] ‘small.M’ and *petite* [pət̪^hit] ‘small.F’. In addition, loanwords from English create low-frequency minimal pairs such as *coule* [kɔl] ‘flow.3.SG’ and *cool* [kul] ‘cool’. It is not clear under a binary view of contrast whether only a handful of contrasts involving tense vowels is sufficient make these phones legitimately contrastive, and so should perhaps be a part of the phonological inventory, and begs the question of whether they are somehow less contrastive or exhibit a weaker contrastive relationship than high-frequency contrasts. These examples illustrate problems in establishing contrast because of conflicting or unclear criteria. This research aims to quantify two of the most used criteria to determine phonological relationships: lexical distinction and predictability of distribution. In doing so, a scale of contrast can be established and move beyond the all-or-nothing binary approach to classifying phonological relationships. As these criteria

form the basis for the current research, they will be discussed in greater detail, along with previous experimental results examining these criteria.

1.2 Determining contrast based on the distinctive function

If two sounds serve a distinctive function, i.e. they are used to distinguish two otherwise identical lexemes or morphemes, they are considered to be in a contrastive relationship, an approach used in both structural (Saussure et al. 1916; Twaddell 1935) and generative phonology (Chomsky & Halle 1968). However, different minimal comparisons can yield different conclusions about what to include in an underlying representation or phoneme inventory. Indeed, for such a common criterion, there are few attested formalizations of how minimal comparisons should be carried out, and there is a lack of commonality across the approaches. These formalizations have focused on the feature level (as opposed to the phonemic level), but nevertheless use lexical contrast to determine contrastive features. For example, Contrastive Specification posits that all and only contrastive features are specified underlyingly, and predictable feature values are eliminated (Steriade 1995), whereas Radical Underspecification claims that all and only unpredictable features are specified (Archangeli 1988). Some algorithms have been created as a way of determining underlying features, such as the Pairwise Algorithm which relies on minimal pair contrasts (Dresher et al. 1994) and the Successive Division Algorithm (Dresher 2008), which does not depend solely on minimal pairs for determining contrastive features.

Even with the aid of an algorithm, the minimal pair test by itself is insufficient to entirely determine what is contrastive. Take *differential substitution* where speakers of different first languages (L1s) will produce different phones for the same second language (L2) phone. Hungarian learners substitute [t] for English [θ], while European French learners substitute [s] for English [θ] (Weinberger 1990). This is explained by proposing different underlying representations being transferred or mapped onto the new target phone based on the L1 inventories. A problem arises with different dialects of the same L1, such as European French and Laurentian French. 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 is correct that L1 feature matrices are being transferred onto a novel L2 phone, 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 (see Jesney 2005 for an account of these cases by dialect-specific active phonological processes). The minimal pair test also leads to disagreements on the members of a phonemic inventory, depending on whether loanwords are considered part of the L1 lexicon. In Japanese [ϕ] and [t^s] only occur before [u]; however, in foreign words, [ϕ] and [t^s] may occur before other vowels (Vance 1987; Ito & Mester 1995; Brown 1997). Also see the example of Laurentian French vowel alternations described above (Côté 2010). These cases bring into question whether a single minimal pair is sufficient to classify the relationship between two phones as contrastive and whether loan words should be among the lexical items being compared.

One way to carry out minimal pair comparisons in a quantifiable way is to calculate the functional load of a language's contrasts. Functional load measures the frequencies of two contrastive sounds and the degree to which those two sounds contrast in all possible environments. This is to evaluate how much work the contrast does as compared to other contrasts (King 1967; Brown 1988; Wedel et al. 2013). Unlike 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. This allows for a more objective assessment of the contribution of a contrast to the overall phonological system of a language. Functional load was used in this study as a means to measure the degree to which two sounds are contrastive. The number of minimal pairs between two specific sounds was counted, as well as the number of minimal pairs in which a single sound participated. (This methodology is discussed in greater detail below.) Sounds that participated in a high number of contrasts were dubbed High Contrast, those with a small number were dubbed Low Contrast, and those in between were dubbed Mid Contrast (not to be confused with high, mid and low vowels in terms of tongue height). In addition, we recognize that the vowels used to exemplify High, Mid and Low Contrast pairs also differ in other ways, such as their acoustic properties. In order to rule out the possibility that perceived differences were due solely to these acoustic properties as opposed to phonological factors, we measured these differences using multiple methodologies and discuss below how there was no consistent effect of the acoustic properties on the results.

1.3 Determining contrast based on the predictability of distributions

The predictability of segmental distributions in a given language is also used to define contrast: “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” (Hall 2009: 2). Rather than being an all-or-nothing criterion, Hall 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 devises algorithms to calculate the uncertainty (i.e. the 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.

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. Determining phonological relationships becomes more complicated when the criterion of distinctiveness overlaps with the criterion of predictability of distribution. For example, in Laurentian French, the lexemes *saute* [sɔt] ‘jump.3P.SG.PRS’ and *sotte* [sɔt] ‘stupid.F’ 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.M’ and *sotte* [sɔt] ‘stupid.F’, where the open variant occurs in a closed syllable and the closed variant in an open syllable in morphologically related words, and never in open final syllables. Since their distribution is sometimes unpredictable (associated with contrast) and sometimes predictable (associated with allophony), it is not clear whether the relationship between these sounds should be classified as contrastive, allophonic or as something between the two. When the indicators for typically contrastive relationships contradict each other in this way, the criterion of predictability of distribution is often ignored as long as lexical distinctions exist. However, one might question whether sound-pairs that do not satisfy all criteria should be classified as having

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.

1.4 Experimental studies on gradient contrast

Various experimental methodologies have been used to explore the perception of phones in allophonic and contrastive relationships, such as the AX task, the 4-interval AX task, and the similarity-rating task. Generally speaking, phones of different categories are easier to discriminate while phones that are exemplars of a single category are difficult to discriminate (Kuhl & Iverson 1995). Evidence for this comes from Peperkamp et al. (2003), who 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 their task, participants heard pairs of two syllables and were asked to judge similarity across the pairs. They found that allophones are difficult to discriminate when embedded within their trigger phonological context.

Allophones have also been found to be perceived as more similar to one another than phonemes (Boomershine et al. 2008). Allophonic alternations entail a change in phonetic category, but not 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; e.g., 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. Boomershine et al. tested whether allophones are perceived as less distinct than contrastive sounds within a L1. They used similarity ratings as well as reaction times (RTs) from a speeded AX discrimination task, where longer RTs were associated with greater similarity and shorter RTs were associated with less similarity. Results indicated that English speakers perceived [d]/[r] (allophonic relationship) as more similar than [d]/[ð] (phonemic relationship), and Spanish speakers perceived [d]/[ð] (allophonic relationship) as more similar than [d]/[r] (phonemic relationship). The results for the pair [ð]/[r] patterned like the contrastive pairs for the respective language groups, which is likely due to them being allophones of different phonemes in each language. Another study to quantify and test intermediate phonological relationships, by experimental means of 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 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 provides a variety of potential causes for this, such as the entropy values between pairs being too close to one another, differences in phonotactic licitness of the contexts in which the consonants occurred, among others.

1.5 Current research

Multiple authors have found the need to appeal to terminology beyond the terms contrastive or allophonic. It appears that phonological relationships more likely fall on a scale from allophonic to contrastive, as opposed to the traditional view that contrast is an all-or-nothing phonological status. However, there has been very little experimental research supporting the view that contrast is gradient (see section above that describes the few studies that have been done). In addition to testing the extremes on the scale of contrastive to allophonic relationships, the current study will also test intermediate relationships

of contrast; specifically vowels pairs that exemplify High, Mid and Low degrees of contrast. These contrast types are tested in two studies: an AX task (Experiment 1) and a similarity rating task (Experiment 2) to facilitate comparisons across experimental paradigms and previous research, such as Boomershine et al. (2008). Different results are expected depending on whether phonological relationships are binary in nature (i.e. fully contrastive or fully allophonic) or gradient in nature (intermediate relationships between contrastive and allophonic). If a binary view of contrast is supported, it is predicted that High and Mid Contrasts will yield similar results of higher accuracy and faster RTs, while Low contrast will pattern different since they are in an allophonic relationship ($H = M > L$). If the gradient view of contrast is supported, it is predicted that for our experimental variable of Contrast (High, Mid, Low), the High Contrast vowel pair should be the easiest to discriminate, resulting in high accuracy and shorter RTs; the Mid Contrast vowel pair should result in lower accuracy scores and longer RTs; and the Low Contrast vowel pair should result in the lowest accuracy scores and longest RTs ($H > M > L$). We also looked at the acoustic differences between the specific vowel pairs to see whether acoustics could also account for the accuracy and speed of participants responses. We expand on this in our discussion of the stimuli below.

2 Experiment 1

2.1 Method

2.1.1 Participants

Participants were 32 native speakers of Laurentian French ($M = 28$ years, range 19–53, 6 males). Speakers of other dialects of French (Acadian, Haitian, Belgian, French, etc.) were not included as these dialects do not have the alternation between tense and lax vowels (described above) that Laurentian French does. All self-reported that they had normal hearing and no language disorders. Of the participants, 14 were born and raised in Ottawa, Ontario; 3 in Gatineau, Quebec; 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. An additional 2 participants were tested but not included in the analysis for equipment error ($N = 1$), falling asleep ($N = 1$).

2.1.2 Stimuli

There were two main criteria used to determine the phonological relationship between two sounds: lexical distinction and the predictability of their distributions.

2.1.2.1 Lexical distinction stimuli selection criteria

The OMNILEX database (Desrochers 2006) was used to establish a word list of French one-syllable words of CV, VC, CVC and CCV syllable structure. The database includes approximately 102,000 lexical entries originating from multiple French dictionaries and the Lexique corpus (New et al. 2004) and phonetic transcriptions are based on European French. Although the database provides minimal pair counts and neighbourhood density values, these could not be used for Laurentian French. The database was therefore re-transcribed to reflect a standard Laurentian pronunciation by a native speaker and expert in Laurentian French phonetics and phonology, paying particular attention to vowels [ɪ], [ɣ], [ʊ] and [ɜ] because these vowels do not occur in European French.

The resulting corpus was then processed using the software *Phonological Corpus Tools*, version 1.1.1 (Hall et al. 2015). The corpus was uploaded into the software and counts the number of minimal pairs in the corpus, in this case using the type frequency. The number of minimal pairs for each of 20 vowels was calculated by counting, for example, how many times a given vowel occurred after [b] in a monosyllable, then how many times

that vowel occurred after [d], and so on, for every consonant and consonant combination. From this, it was calculated (a) how many minimal pairs a single vowel participated in with all other vowels (referred to as *individual count*), 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). Appendix A summarizes these results. To represent a scale of contrast, the final selection of 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 (Table 1).

Note that the difference between High and Mid in terms of minimal pairs is not proportional 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 below). In addition, the Low contrast pair [y]-[ɣ] is allophonic in LF, sharing no minimal pairs, while the other two pairs contrast; for example, *gamme* [gam] ‘scale’ versus *gomme* [gɔm] ‘eraser’; *role* [ʁɔl] ‘role’ versus *roule* [ʁɔl] ‘roll.3.SG’. This will be important when testing the binary view of contrast.

2.1.2.2 Predictability of distribution stimuli selection criteria

Hall (2009: 40) depicts a “continuum of phonological relationships” based on her Probabilistic Model of Phonological Relationships. The PPRM focuses predominantly on the factor of predictability of distribution, i.e. how likely a segment is to occur in a given phonological environment, which is dependant on the phenomena of the language under study. Distributions that do not overlap at all are at one end, such as [y] and [ɣ] in LF, while distributions that overlap are at the other, and are contrastive, such as [b] and [g] in LF, with a range of possibilities in between. The *Phonological Corpus Tools* software (Hall et al. 2015) was used to calculate the predictability of distribution of the vowels tested in the current experiment. The vowels’ predictability was calculated according to functional load based on four local environments: before the end of a word, before another vowel (0 for this environment), before a consonant that is [v, z, ʒ] or [ʁ], and before a consonant that is *not* [v, z, ʒ] or [ʁ]. The predictability of distribution over the four environments is provided as entropy (uncertainty) as a number out of 1, where 1 indicates a perfectly overlapping distribution and therefore unpredictability, associated with contrast, and 0 indicates a perfectly complementary distribution. The specific algorithms are provided in the *Phonological Corpus Tools* help file (Hall et al. 2015). The High Contrast vowel pair exhibits the greatest level of uncertainty, which corroborates the level of contrast

Table 1: High, Mid, and Low vowel contrast pairs by individual minimal pair counts, shared minimal pair counts, predictability of distribution, and frequency.

Contrast level	Individual count	Shared count	Predictability of distribution	Frequency
High	a	809	0.95	410
	ɔ	595		112
Mid	o	496	0.73	191
	u	340		25
Low	y	266	0	73
	ɣ	262		0

assigned to that pair, followed by the Mid contrast vowel pair. The Low vowel pair have complementary distributions, which is associated with allophony (Table 1).

Predictability of distribution is tied in part to frequency, since it is based on the number of times a sound occurs in a particular phonotactic environment. The number of minimal pairs and the relative type frequency have been shown to be correlated to robustness of contrast and speed of processing (Vitevitch & Luce 1999; Wedel et al. 2012). Therefore, the relative frequency of a sound was a controlled factor. Frequency of vowels in Laurentian French were based on the OMNILEX database, and done by calculating the number of times each vowel occurred in monosyllabic words. Since there is no corpus with lexical frequencies for LF, only type frequency calculations were done. The stimuli's type frequencies based on the OMNILEX corpus are provided in Table 1. As with minimal pair counts, the high-contrast vowel pair [a-ɔ] consists of high type-frequency vowels; the low-contrast vowel pair [y-ɣ] consists of low type-frequency vowels; the mid-level contrast vowel pair [ou] falls between the two.

2.1.2.3 Stimuli creation

Stimuli were produced by a trained male phonetician who is a native speaker of Laurentian French from Quebec City. Four consonant C_C frames consisting of [l, b, f] and [ʃ] were combined with the six vowels [a], [ɔ], [o], [u], [y] and [ɣ] making for 24 unique syllables. 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. Stimuli were normalized in Praat (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 tokens in both Same and Different pairs were matched for intonation curve. In a few cases, the intonation curve was manipulated synthetically.

2.1.2.4 Acoustic measurements

Acoustic similarity was taken into consideration when choosing High, Mid and Low vowel stimuli pairs used in all experiments, so that the vowels were roughly acoustically matched by tongue position and lip rounding. This was done so as not to 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 Contrast pair of vowels be maximally acoustically different compared to the pair of Low Contrast vowels. For example, nasalization in English is allophonic but in French is contrastive. Research has shown that divergence along an acoustic cue is more distinct when the cue signals a contrast (Desmeules-Trudel 2015, 2016; Versteegh et al. 2014). This was done based on the phonetic properties of the chosen vowels, which was further verified with acoustic analyses. The stimuli's F1 and F2 measurements were taken in Praat from a steady-state portion of the vowel as close as possible to the mid-point. Figure 1 plots F1 and F2 for the stimuli, and Table 2 provides the F1 and F2 values. Values represent the mean of the two tokens that were selected as stimuli. The values are similar to Martin's (2002: 84) vowel space for male speakers of the Quebec dialect, indicating that the experimental stimuli are representative of LF vowels, and no vowels in the stimuli were anomalies of their phonetic category or unrecognizable as a member of their category.

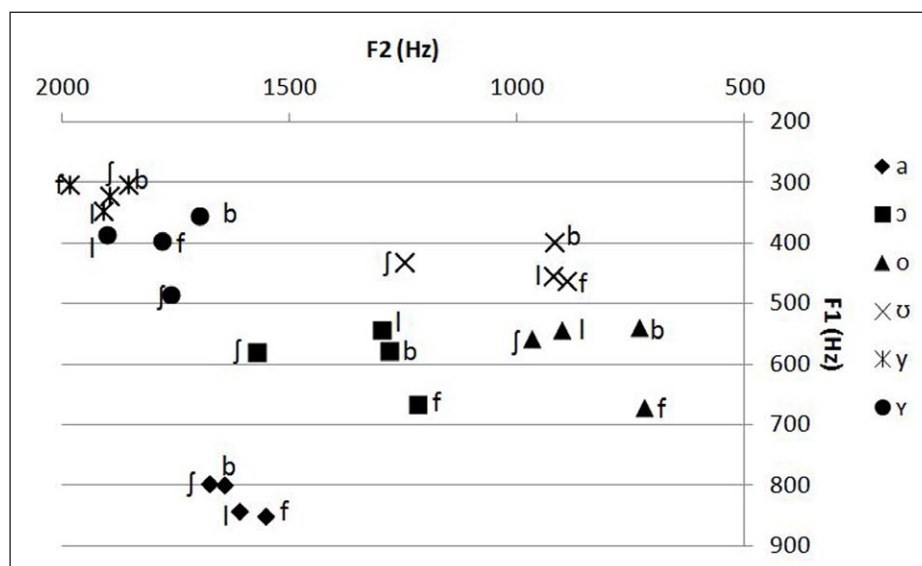


Figure 1: F1 and F2 values for vowels in CVC stimuli frames.

Table 2: F1 and F2 values in Hz for vowels from High, Mid, Low Contrast pairs, by stimuli frame.

Contrast	Stimulus	F1	F2	Stimulus	F1	F2	Difference F1	Difference F2
High	[aɪ]	798.15	1675.32	[ɔɪ]	579.72	1570.53	218.43	104.79
	[faɪ]	849.92	1550.68	[fɔɪ]	665.94	1215.33	183.98	335.35
	[laɪ]	843.46	1607.32	[ləɪ]	543.72	1295.33	299.74	311.99
	[baɪ]	799.92	1640.71	[bɔɪ]	577.18	1277.35	222.74	363.36
Average		822.86	1618.51		591.64	1339.64	231.22	278.87
Mid	[ɔʊ]	430.50	1245.57	[oʊ]	558.06	966.77	127.56	278.80
	[fɔʊ]	462.18	888.45	[foʊ]	672.06	718.46	209.88	169.99
	[ləʊ]	454.47	918.91	[ləʊ]	542.35	898.33	87.88	20.58
	[bɔʊ]	398.36	915.94	[boʊ]	540.05	728.03	141.69	187.91
Average		436.38	992.22		578.13	827.90	141.75	164.32
Low	[ɣʏ]	321.60	1891.62	[ɣʏ]	484.95	1756.74	163.35	134.88
	[fɣʏ]	304.25	1980.27	[fɣʏ]	397.39	1778.16	93.14	202.11
	[lɣʏ]	347.86	1905.57	[lɣʏ]	386.68	1899.38	38.82	6.19
	[bɣʏ]	304.34	1852.19	[bɣʏ]	356.09	1694.55	51.75	157.64
Average		319.51	1907.41		406.28	1782.21	86.77	125.21

Table 2 shows that on average, the largest difference between F1 and F2 is between the High Contrast pair [a-ɔ], followed by the Mid Contrast [o-u] pair, followed by the Low Contrast [ɣ-y] pair. Based on these average values, it is not possible to draw a clear line between judgments based on F1-F2 values and strength of contrast as calculated by minimal pair counts and relative frequency: result predictions appear go in the same direction whether based on average F1 and F2 differences or level of contrast. However, when comparing across the specific consonant frames, F1 and F2 differences do not yield

such a clear prediction. For example, there is roughly a 210 Hz difference in F1 and a 170 Hz difference in F2 between [fʊf] and [fof] while there is a roughly a 184 Hz difference in F1 and a 335 Hz difference in F2 between [faf] and [fɔf]. It is not clear in this case, based on absolute Hz differences, whether participants should find it easier to distinguish between one vowel pair over the other. In order to discourage an acoustic mode of perception, the interstimulus intervals (ISIs) were set to 1500 ms. Previous research has shown that a longer ISI encourages a phonological mode of processing, obscuring finer acoustic differences, and shorter ISIs encourages a more phonetic/auditory mode of processing (Werker & Logan 1985). Our use of an 1500 ms ISI should encourage participants to perform the task more in line with phonological relationships.

When one takes a closer look at the individual F1 and F2 differences for specific pairs, the amount of differences between F1 and F2 are not always in line with the contrast category. For example, for the [ʃ _ ʃ] frame, the greatest F1 difference between the pairs was for High, followed by Low, followed by Mid, and for the F2 the greatest difference was for Mid, followed by Low, followed by High. If the acoustic differences between the specific pairs is the most important factor for the accuracy and speed of participants' responses, we would predict that there would be a correlation between F1 and F2 difference scores and performance on the task. We ran statistical tests to explore this possible interaction and return to this issue in the results.

2.1.2.5 Machine-assisted calculations of acoustic similarity

The above is only one possible way of evaluating acoustic similarity. Another way of measuring acoustic similarity was developed by Mielke (2012), who uses a phonetically-based metric to assess the similarity of sounds. This metric combines 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 is also used to calculate phonetic distances between phones. For the present study, the acoustic distance was measured between the six vowels selected as stimuli ([a, ɔ, o, u, y, ʏ]) using the same methods as in Mielke (2012) developed for acoustic comparisons. 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 similarities between vowels. This provides a weighted acoustic distance measure 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). The distances were as follows: High Contrast pair [a-ɔ] = 111.4, Mid Contrast pair [o-u] = 94.17, Low Contrast pair [y-ʏ] = 130.8. Somewhat surprisingly, [ʏ] and [y] are not the most similar, which was expected based on the F1–F2 vowels (Table 2). As Praat calculates the weighted distances between vowels, all spectral information is used, regardless of how salient the frequencies 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 though, 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. It cannot be predicted from this analysis that, for

example, participants would perceive [o]-[u] 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 if participants were performing the experimental tasks as if with non-speech stimuli; if this is not obtained 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.

Given that higher frequencies are less relevant to human speech perception and these frequencies may have played a role in the resulting acoustic distances, the stimuli were downsampled to 11,000 Hz to eliminate periodicity above 5500 Hz and were then re-analyzed. Figure 2 shows the outcome of the re-analysis. Even with downsampling, the low pair [y]-[ʏ] still remained the most dissimilar. 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). Results would pattern according to these analyses in this section, with [o]-[u] perceived as most similar, [y]-[ʏ] being perceived as least similar, and results for [a]-[ɔ] falling between the other two pairs ($L > H > M$).

2.1.3 Design

Syllables were paired by consonant, e.g. [bob-bub]. There were two conditions: Trial Type (Different, Same) and Contrast (High, Mid, Low). Stimuli in the Same condition consisted of two acoustically different tokens, e.g., Same-Mid [bob₁-bob₂]. Stimuli in the Contrast condition consisted of two different vowels, e.g., Different-Mid [bob₁-bub₁]. There were 48 total trials, with 8 trials in each condition (Different: High, Mid, and Low Contrast; Same: High, Mid and Low Contrast; $6 \times 8 = 48$ trials). Stimuli were quasi-randomized to ensure 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, and

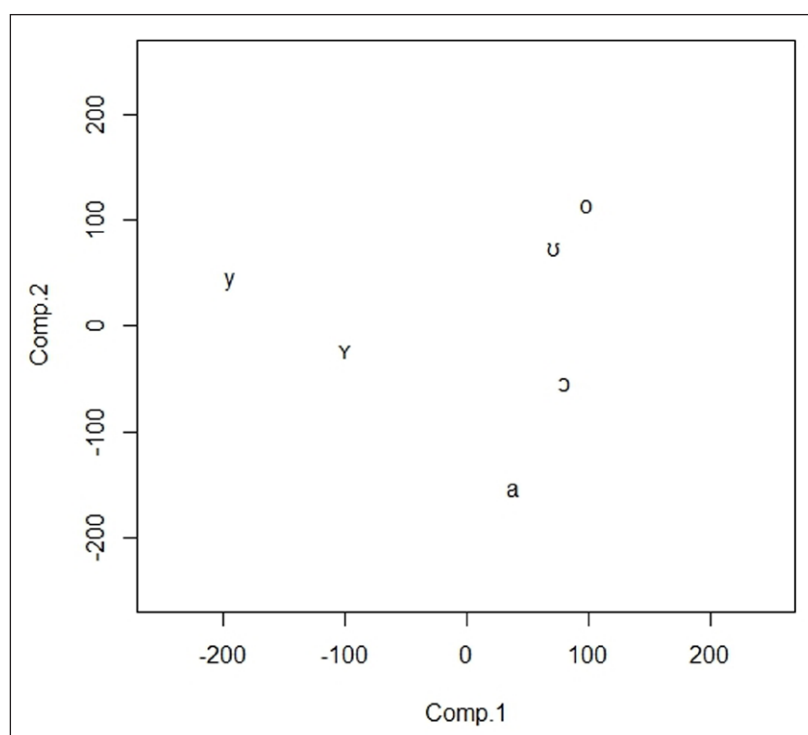


Figure 2: Principal component analysis with downsampled stimuli.

with 2 ordered lists. An ISI of 1500 ms was used to encourage a phonological mode of processing.

2.1.4 Procedure

Stimuli were presented using PsyScope software. 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, and counterbalanced across participants for whether Same or Different corresponded to the left or right side of the keyboard. The beginning of each trial was indicated by a tone. Participants wore headphones in a quiet room and were allowed to adjust the volume to a comfortable listening level.

2.2 Results and discussion

2.2.1 Accuracy

A repeated measures 2×3 ANOVA was done on the number of accurate responses with Trial Type (Same, Different) and Contrast (High, Mid, Low). For Contrast, the assumption of sphericity was violated and Greenhouse-Geisser adjusted values were used. The mean correct responses in the conditions were as follows: Same trials, High Contrast ($M = 7.81$, $SD = 0.47$), Mid Contrast ($M = 7.66$, $SD = 0.48$), Low Contrast ($M = 7.84$, $SD = 0.45$), Different trials, High Contrast ($M = 7.91$, $SD = 0.3$), Mid Contrast ($M = 7.63$, $SD = 0.49$), Low Contrast ($M = 6.5$, $SD = 1.37$). There were main effects of Trial Type ($F(1,31) = 13.3$, $p < .001$, $\eta_p^2 = .3$) and Contrast ($F(1.38,42.74) = 16.57$, $p < .001$, $\eta_p^2 = .35$), and a significant interaction of Trial Type \times Contrast ($F(1.51,46.76) = 25.12$, $p < .001$, $\eta_p^2 = .45$) (Figure 3). To examine this interaction, Contrast was analyzed separately for Different and Same trials with 1-way ANOVAs. For Different trials, the assumption of sphericity was violated and Greenhouse-Geisser-corrected values were used. There was a significant effect of Contrast ($F(1.26,39.1) = 24.65$, $p < .001$, $\eta_p^2 = .44$).

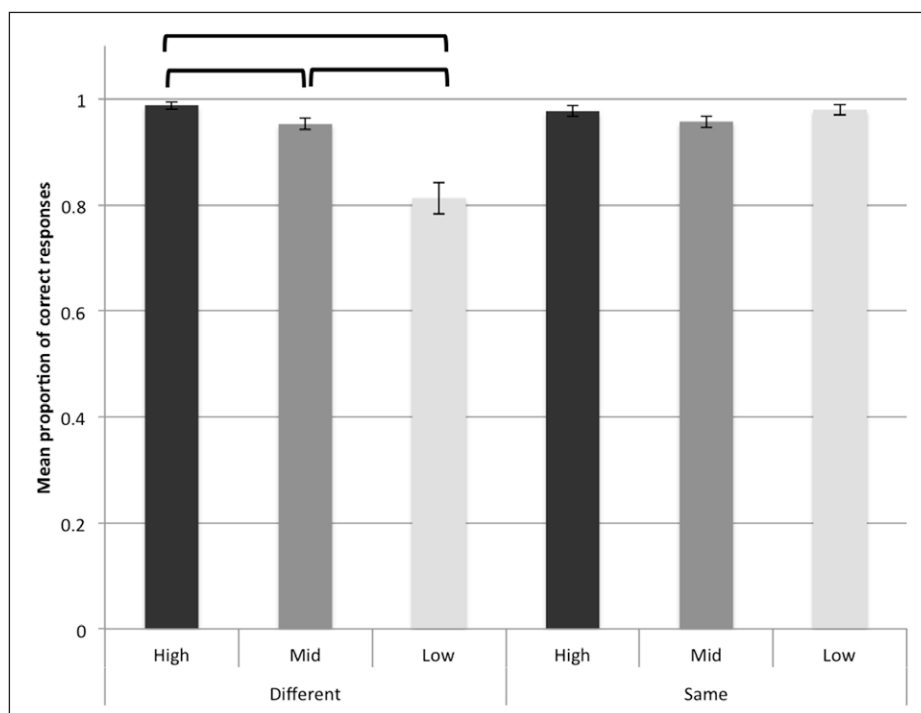


Figure 3: Mean accuracy by contrast and trial type for Experiment 1. Error bars show the standard error.

Pairwise comparisons were done using Bonferroni corrections for multiple comparisons. Participants were significantly more accurate on the High-Mid pairs ($p < .05$), High-Low pairs ($p < .001$), and Mid-Low pairs ($p < .001$). For example, participants were more accurate on High pairs such as [bab-bɔb] than Mid pairs such as [bob-bub]. For the Same trials, there was no effect of Contrast ($F(1,31) = 2.37, p = .13$). For the Different pairs, Pearson correlations were computed to assess the correlation between accuracy and F1 difference scores and between accuracy and F2 difference scores (see Table 2). There was a near significant positive correlation between accuracy and F1 difference scores ($r = 0.57, n = 12, p = 0.051$), indicating that as the difference between the pairs' F1 values increased, participants accuracy scores also increased. Similarly, there was a near significant positive correlation between accuracy and F2 difference scores ($r = 0.55, n = 12, p = 0.065$).

2.2.2 Reaction times (RTs)

RTs were based on correct responses ($N = 85$ responses removed). Responses below 200 ms and above 4 s were removed ($N = 4$), as well as responses within 3 SDs above or below each condition mean ($N = 64$). The mean RTs in the conditions were as follows: Same trials, High Contrast ($M = 794.97, SD = 185.45$), Mid Contrast ($M = 822.11, SD = 162.93$), Low Contrast ($M = 736.14, SD = 163.52$), Different trials, High Contrast ($M = 795.63, SD = 142.92$), Mid Contrast ($M = 874.54, SD = 169.80$), Low Contrast ($M = 917.87, SD = 183.56$). A 2×3 repeated measures ANOVA showed main effects of Trial Type ($F(1,31) = 15.71, p < .001$), Contrast ($F(2,62) = 9.45, p < .001$), as well as a significant interaction between Trial Type and Contrast ($F(2,62) = 30.13, p < .001$) (Figure 4). To examine the interaction, 1-way ANOVAs were done on Contrast with separate analyses for Same and Different trials. A significant effect of Contrast was found among Different pairs ($F(2,62) = 21.96, p < .001$), as well as among Same pairs ($F(2,62) = 16.08, p < .001$). Among Different pairs, RTs were significantly shorter for High-Mid pairs ($p < .001$), and High-Low pairs ($p < .001$), but Mid-Low pairs were not

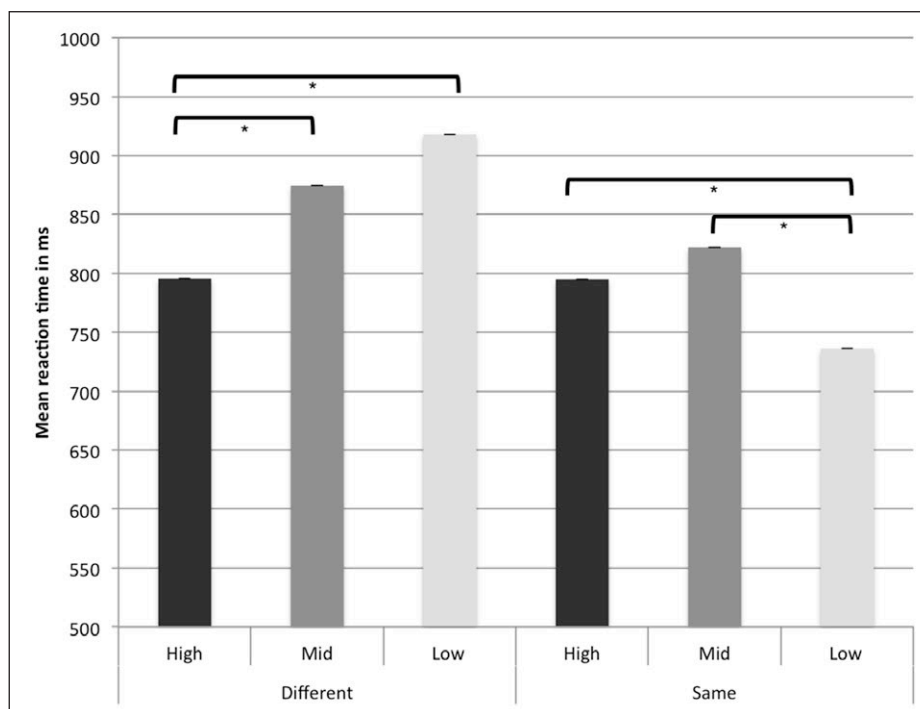


Figure 4: Mean reaction times by contrast and trial type for Experiment 1. Error bars show the standard error.

significantly different ($p = .10$). In the Same trials types, RTs were significantly shorter for Low-Mid pairs ($p < .001$) and Low-High pairs ($p < .001$), but not for the High-Mid pairs ($p = .4$). Additional analyses using Pearson correlations were computed to assess the correlation between RTs and the stimuli's F1 difference scores and between RTs and the stimuli's F2 difference scores. There was a weak negative correlation between RT and F1 difference scores ($r = -0.15$, $n = 12$, $p = 0.63$). While this was not significant, the direction of the correlation was opposite than what one would predict if acoustic differences between the stimuli pairs were driving the speed of participants' RTs. Similarly, there was a weak negative correlation between accuracy and F2 difference scores ($r = -0.10$, $n = 12$, $p = 0.75$).

For both accuracy and RT measures, there was a significant interaction between Trial type and Contrast type. Among the Different trials, participants were significantly more accurate when comparing High-Mid, High-Low, and Mid-Low contrast pairs ($H > M > L$), as according to the predictions based on gradient contrast. For RT measures though, only the High-Mid and High-Low conditions were statistically different, and Mid-Low did not reach significance. Importantly, as the results statistically set the High Contrast condition apart from the Mid Contrast condition, the results do not support a binary view of contrast. Among the Same pairs for accuracy, there were no significant differences. This is unsurprising since it is less difficult to confirm two things are the same than to identify them as different, resulting in the known response bias for AX tasks where participants tend to choose Same when unsure (Gerrits & Schouten 2004).

Based on a gradient view of contrast, it was predicted that the level of Contrast (High, Mid, Low) would be reflected in accuracy scores and RTs, with highest accuracy and shortest RTs for the High condition, followed by the Mid, then Low condition. This prediction was partially borne out. Significant difference for accuracy scores were found between High-Mid, High-Low, and Mid-Low pairs. Results from RTs showed High-Mid and High-Low Contrast pairs were significantly different. Overall, the results from the different trials demonstrate a facility for High Contrast stimuli. If contrast was strictly binary in nature, it was expected that High and Mid pairs would yield similar results, but this was not the case.

3 Experiment 2

Experiment 2 used a different methodology of similarity ratings to test High, Mid and Low degrees of contrast. If the contrast between vowels is binary, it is predicted that High and Mid Contrast would both yield similar results of faster RTs and higher accuracy, while Low would not since they are in an allophonic relationship ($H = M > L$). If contrast is more gradient, it is predicted that participants would rate the similarity between contrasts on a scale ($H > M > L$), where $>$ means "is more different than".

3.1 Method

3.1.1 Participants

Participants were the same as in Experiment 1. All participants first completed the AX task (Experiment 1), followed by a 4IAX task, and then the Similarity-Rating Task (Experiment 2). The results from the 4IAX task corroborated the results from the AX task and are not presented here for brevity (see Stevenson 2014).

3.1.2 Stimuli

Stimuli consisted of only the different pairs used in the AX task (e.g. Mid Contrast pair [bob-b**o**b]). Pilot experiments were run which included Same stimuli, but participants relativized similarity when these were included so that all Same stimuli received the

highest possible similarity rating, 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.

3.1.3 Design

Participants heard each CVC-CVC stimulus pair once (consonants: [b, f, l, ʃ]; vowels: [a, ɔ, o, u, y, ɪ]), totalling 24 trials and 8 instances each of High, Mid, and Low Contrast stimuli, e.g. participants heard the High Contrast pair [bab-bɔb] and rated the similarity of the syllables on a scale of 1 to 6. The two syllables were separated by 1500 ms.

3.1.4 Procedure

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. 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, a tone was used to draw participants’ attention to the beginning of each trial.

3.2 Results and discussion

Following Boomersshine et al. (2008), raw similarity-rating scores were transformed into z scores to normalize responses across participants. Figure 5 shows the normalized similarity ratings averaged across participants by Contrast (High, Mid, Low).

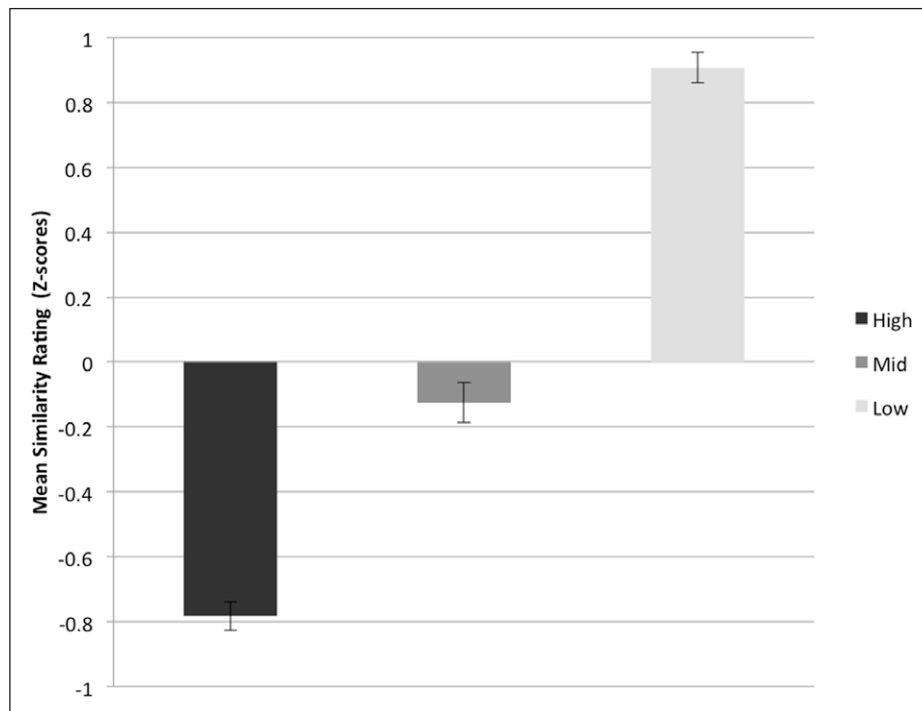


Figure 5: Average similarity rating for Experiment 2 (z-scores). Error bars show the standard error.

A 1-way repeated measures ANOVA was done to determine if the differences between similarity ratings for High, Mid and Low Contrast pairs were significant. Results showed that there was a main effect of Contrast ($F(2,62) = 184.85, p < 0.001$). Pairwise comparisons indicated that all conditions were statistically significantly different from one another (High–Mid: $p < .001$, High–Low: $p < .001$, Mid–Low: $p < .001$). As an example, participants rated High pairs such as [bab-bɔb] as more similar than Mid pairs such as [bob-bɔb].

High Contrast pairs were judged to be the most different; Low Contrast pairs were judged to be the most similar; and Mid Contrast pairs fell between High and Low in terms of similarity ($H > M > L$). These findings are consistent the view of gradient levels of contrast. No support for the binary view of contrast was found, otherwise High and Mid vowels should have yielded similar results.

4 General discussion

This research examined the notion that phonological relationships do not always perfectly match the criteria for being wholly contrastive or allophonic. High, Mid, and Low levels of contrast were tested to see whether phonological relationships are perceived as binary (i.e. only contrastive vs. allophonic), or whether degrees of contrast can be perceived (i.e. on a scale from contrastive to allophonic). In Experiment 1, results on the different trials yielded differentiation between High, Mid, and Low conditions. On the Same trials, RT differences were found between High-Low and Mid-Low pairs. The likely reason why the results were not mirrored on Different and Same trials lies in the nature of the task being asked of the participants. Different trials tested vowel contrasts, while Same trials tested participants' ability to judge acoustic similarities between two same vowels. In Experiment 2, 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. While Boomershine et al. (2008) used a five-point scale and a 1000 ms ISI and the present study used a six-point 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. The present study included stimuli from three strengths of contrast as quantified by predictability of distribution and functional load.

Although the results do not perfectly support the prediction based on a gradient view of contrast, they clearly do not support a purely binary view of contrast where a relationship can be considered contrastive as 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, regardless of acoustic differences between the vowels. For example, in terms of similarity ratings in Experiment 2, if the binary view of contrast held, the High and Mid Contrast vowels should 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.

The results corroborate previous literature regarding purely allophonic and contrastive relationships: phones in an allophonic relationship are more difficult to perceive than those in a contrastive relationship (e.g. Dupoux et al. 1997; Boomershine et al. 2008; Johnson & Babel 2010). Moreover, the current study provides new data supporting the hypothesis

that there are phonological relationships between these two extremes. How can these findings be incorporated into current theoretical frameworks used to define and describe phonological relationships? Classifying segments as contrastive or not can influence how a phonological analysis proceeds. When segments contrast in some contexts and not in others, this can create disagreement about whether those segments should be included in an underlying phonemic inventory (described in Larson-Hall 2004). 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 are 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 of contrast and allophony, is inadequate in describing phenomena which fall between one and the other (see Hall 2013 for an extensive list of authors that use terminology such as "quasi-phonemic" and "mushy contrast"). Hall's PPRM focuses on the factor of predictability of distribution to quantify the continuum of phonological relationships, measured as entropy (also see Hall 2015). While Hall's (2009) study did not yield definitive results, the idea of quantifying phonological relationships was extended in this paper to the measure of functional load, in addition to that of predictability of distribution, and evidence of phonological relationships between contrastive and allophonic was found. However, since the two measures did not offer different predictions from one another, our results cannot serve to distinguish between these two measures as one being a greater predictor of results over the other, or as a stronger measure of contrast. Further research is needed to determine whether these two factors are too highly correlated to be distinguishable from one another, or whether they can be isolated as independent factors. It may also be that if sound pairs are too close to one another in their measures – which is to say, too close in strength of contrast – no significant differences will be found.

As this is one of the first experimental studies to test gradient levels of contrast based on specific measures, it provides a reference from which different languages and experimental paradigms can be compared. The 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. Based on the present study, it should be possible to apply the same measures to segments that occur in other languages and arrive at comparable results. One would predict that speakers of another language would yield results that represent the lexical distinction and predictability of distributions between 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 Laurentian French speakers. Applying this methodology in reverse, it may have the potential to be used as a diagnostic for phonological relationships. One limitation of the current study is that it only examines processing of vowels. It has been argued that consonants and vowels may be processed differently, and that consonants play a greater role than vowels with regards to lexical processing (Nespor

et al. 2003; Havy et al. 2014). Thus, one might find less evidence for gradient contrast when examining the processing of consonants, given their preferential status in lexical processing. In addition, although no evidence was found of a direct correlation between acoustic differences and results, it would be ideal for future studies to tease these apart, using stimuli that are equally acoustically different and of different phonological relationships. Unfortunately, many previous studies do not include measurements of acoustic differences between stimuli.

In summary, this work provides 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 other criteria are not, or where one criterion is partially satisfied. The resulting ambiguities in phonological status may be resolved by using quantifiable measures for the criteria traditionally used to evaluate phonological relationships. In doing so, we may better represent the range of relationships between categories of speech sounds and further our understanding of sound patterns in human language.

Abbreviations

F = feminine, ISI = interstimulus interval, LF = Laurentian French, M = masculine, N = number, PPRM = Probabilistic Model of Phonological Relationships, RT = reaction time, SD = standard deviation, SG = singular

Additional File

The additional file for this article can be found as follows:

- Appendix A. Minimal pair counts by vowel in Laurentian French for CV, VC, CVC and CCV syllables. DOI: <https://doi.org/10.5334/gjgl.162.s1>

Competing Interests

The authors have no competing interests to declare.

References

- Archangeli, Diana. 1988. Aspects of Underspecification Theory. *Phonology* 5. 183–207. DOI: <https://doi.org/10.1017/S0952675700002268>
- Avery, Paul, Elan Dresher & Keren Rice (eds.). 2008. *Contrast in phonology*. Berlin: Mouton.
- Boersma, Paul & David Weenik. 2014. Praat: doing phonetics by computer. Version 5.3.79, retrieved from <http://praat.org/>.
- Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume & Keith Johnson. 2008. The impact of allophony versus contrast on speech perception. In Peter Avery, Elan Dresher & Keren Rice (eds.), *Contrast in phonology*, 143–172. Berlin: Mouton de Gruyter.
- Brown, Adam. 1988. Functional load and the teaching of pronunciation. *Tesol Quarterly* 22. 593–606. DOI: <https://doi.org/10.2307/3587258>
- Brown, Cynthia. 1997. Acquisition of segmental structure: Consequences for speech perception and second language acquisition. Montreal: McGill University dissertation.
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Cohn, Abigail. 2006. Is there gradient phonology? In Gisbert Fanselow, Caroline Féry, Matthias Schleswsky & Ralph Vogel (eds.), *Gradience in grammar: Generative perspectives*, 25–44. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199274796.003.0002>

- Côté, Marie-Hélène. 2010. La longueur vocalique devant consonne allongeante en contexte final et dérivé en français laurentien. In Carmen LeBlanc, France Martineau & Yves Frenette (eds.) *Vues sur les français d'ici, sous la direction de*, 49–75. Québec: Presses de l'Université Laval.
- Côté, Marie-Hélène. 2012. Laurentian French (Quebec). In Randall Gess, Chantal Lyche & Trudel Meinsburg (eds.), *Phonological variation in French. Illustrations from three continents*, 235–274. Philadelphia, PA: John Benjamins.
- Desmeules-Trudel, Felix. 2015. Perception of Québec French nasal vowels by Brazilian learners: A pilot study. *Langues et linguistique* 35. 40–54.
- Desmeules-Trudel, Felix. 2016. Propriétés aérodynamiques des voyelles nasales et potentiellement nasalisées en français québécois. In Santa Vinerte (ed.), *Proceedings of the 2015 Annual Conference of the Canadian Linguistic Association*. Ottawa, ON.
- Desrochers, Alain. 2006. OMNILEX: Une base de données sur le lexique du français contemporain. *Cahiers Linguistiques d'Ottawa* 34. 25–34.
- Dresher, B. Elan. 2008. The contrastive hierarchy in phonology. In Peter Avery, Elan Dresher & Keren Rice (eds.), *Contrast in phonology*, 11–33. Berlin: Mouton.
- Dresher, B. Elan, Glynn Piggott & Keren Rice. 1994. Contrast in phonology: Overview. In Carrie Dyck (ed.), *Toronto Working Papers in Linguistics* 13, 3–17.
- Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian & Jacques Mehler. 1997. A destressing “deafness” in French? *Journal of Memory and Language* 36. 406–421. DOI: <https://doi.org/10.1006/jmla.1996.2500>
- Ernestus, Mirjam. 2011. Gradience and categoricity in phonological theory. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2115–2136. Cambridge, MA: Blackwell.
- Gerrits, Ellen & M. E. H. Schouten. 2004. Categorical perception depends on the discrimination task. *Perception and Psychophysics* 66. 363–376. DOI: <https://doi.org/10.3758/BF03194885>
- Goldsmith, John. 1995. Phonological theory. In John Goldsmith (ed.), *The handbook of phonological theory*, 1–23. Cambridge, MA: Blackwell.
- Hall, Daniel Currie & Kathleen Currie Hall. 2016. Marginal contrasts and the Contrastivist Hypothesis. *Glossa: a journal of general linguistics* 1. 1–23. DOI: <http://doi.org/10.5334/gjgl.245>.
- Hall, Kathleen Currie. 2009. *A probabilistic model of phonological relationships from contrast to allophony*. Columbus, OH: The Ohio State University unpublished dissertation.
- Hall, Kathleen Currie. 2012. Phonological relationships: A probabilistic model. *McGill Working Papers in Linguistics* 22.
- Hall, Kathleen Currie. 2013. A typology of intermediate phonological relationships. *The Linguistic Review* 30. 215–275. DOI: <https://doi.org/10.1515/tlr-2013-0008>
- Hall, Kathleen Currie. 2015. Categorical segments, probabilistic models. In Eric Raimy & Charles Cairns (eds.), *The segment in phonetics and phonology*, 129–145. Cambridge, MA: Blackwell. DOI: <https://doi.org/10.1002/9781118555491.ch6>
- Hall, Kathleen Currie, Blake Allen, Michael Fry, Scott Mackie & Michael McAuliffe. 2015. Phonological CorpusTools, Version 1.0. [Computer program]. Available from: <https://sourceforge.net/projects/phonologicalcorpustools/>.
- Havy, Mélanie, Josette Serres & Thierry Nazzi. 2014. A consonant/vowel asymmetry in word-form processing: Eye-tracking evidence in childhood and in adulthood. *Language and Speech* 57. 254–281. DOI: <https://doi.org/10.1177/0023830913507693>
- Ito, Junko & Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Cambridge: Blackwell.

- Jesney, Karen. 2005. Stridency and differential substitution in two dialects of French. *Handout from presentation at the Language Research Centre Graduate Student Forum*, Calgary, AB.
- Johnson, Keith & Molly Babel. 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. DOI: <https://doi.org/10.1016/j.wocn.2009.11.001>
- King, Robert D. 1967. Functional load and sound change. *Language* 43. 831–852. DOI: <https://doi.org/10.2307/411969>
- Kuhl, Patricia & Paul Iverson. 1995. Linguistic experience and the “Perceptual magnet effect”. In Winifred Strange (ed.), *Speech perception and linguistic experience*, 121–154. Baltimore, MD: York Press.
- Ladd, Robert. 2006. “Distinctive phones” in surface representation. In Louis Goldstein, Douglas Harry Whalen & Catherine T. Best (eds.), *Laboratory Phonology* 8, 3–26. Berlin: Mouton.
- Larson-Hall, Jenifer. 2004. Predicting perceptual success with segments: A test of Japanese speakers of Russian. *Second Language Research* 20. 33–76. DOI: <https://doi.org/10.1191/0267658304sr230oa>
- Lombardi, Linda. 2003. Second language data and constraints on Manner: Explaining substitutions for the English interdental. *Second Language Research* 19. 225–250. DOI: <https://doi.org/10.1177/026765830301900304>
- Lu, Yu-An. 2014. Mandarin fricatives redux: The psychological reality of phonological representations. *Journal of East Asian Linguistics* 23. 43–69. DOI: <https://doi.org/10.1007/s10831-013-9111-5>
- Martin, Pierre. 2002. Le système vocalique du français du Québec. De l’acoustique à la phonologie. *La Linguistique* 38. 71–88. DOI: <https://doi.org/10.3917/ling.382.0071>
- Matell, Michael & Jacob Jacoby. 1971. Is there an optimal number of alternatives for likert scale items? *Educational and psychological measurement* 31. 657–674. DOI: <https://doi.org/10.1177/001316447103100307>
- Mielke, Jeff. 2012. A phonetically based metric of sound similarity. *Lingua* 122. 145–163. DOI: <https://doi.org/10.1016/j.lingua.2011.04.006>
- Nespor, Marina, Marcela Peña & Jacques Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2. 203–229. DOI: <https://doi.org/10.1418/10879>
- New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36. 516–524. DOI: <https://doi.org/10.3758/BF03195598>
- Peperkamp, Sharon, Michèle Pettinato & Emmanuel Dupoux. 2003. Allophonic variation and the acquisition of phoneme categories. In Barbara Beachley, Amanda Brown & Frances Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* Vol. 2, 650–661. Sommerville, MA: Cascadilla Press.
- Pierrehumbert, Janet. 2003. Probabilistic phonology: Discrimination and robustness. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probability theory in linguistics*, 177–228. Cambridge, MA: MIT Press.
- Renwick, Margaret E.L. 2014. *The phonetics and phonology of contrast: The case of the Romanian vowel system*. Berlin: Mouton. DOI: <https://doi.org/10.1515/9783110362770>
- Renwick, Margaret EL, Ioana Vasilescu, Camille Dutrey, Lori Lamel & Bianca Vieru. 2016. Marginal contrast among Romanian vowels: Evidence from ASR and functional load. In *Proceedings of Interspeech 2016*, 2433–2437. San Francisco, CA.
- Saussure, Ferdinand, Charles Bally, Albert Sechehaye & Albert Riedlinger. 1916. *Cours de linguistique générale*. Lausanne: Payot.

- Scobbie, James & Jane Stuart-Smith. 2008. Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In *Contrast: Perception and acquisition: Selected papers from the Second International Conference on Contrast in Phonology*, 87–113. Berlin: Mouton.
- Steriade, Donca. 1995. Underspecification and markedness. In John Goldsmith (ed.), *The handbook of phonological theory*, 114–174. Cambridge, MA: Blackwell.
- Twaddell, W. Freeman. 1935. *On defining the phoneme*. Baltimore, MD: Waverly Press.
- Vance, Timothy. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Versteegh, Maarten, Amanda Seidl & Alejandrina Cristia. 2014. Acoustic correlates of phonological status. In *Fifteenth Annual Conference of the International Speech Communication Association*, 91–95. Singapore.
- Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40. 374–408. DOI: <https://doi.org/10.1006/jmla.1998.2618>
- Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128. 179–186. DOI: <https://doi.org/10.1016/j.cognition.2013.03.002>
- Weinberger, Steven. 1990. Minimal segments in second language phonology. In Allan James & Jonathan Leathe (eds.), *New sounds 90: Proceedings of the 1990 Amsterdam Symposium on the acquisition of second language speech*, 137–173. New York: Mouton.
- Yip, Moira. Perceptual influences in Cantonese loanword phonology. *Journal of the Phonetic Society of Japan* 6. 4–21.

How to cite this article: Stevenson, Sophia and Tania Zamuner. 2017. Gradient phonological relationships: Evidence from vowels in French. *Glossa: a journal of general linguistics* 2(1): 58.1–22, DOI: <https://doi.org/10.5334/gjgl.162>

Published: 16 June 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 