

Phonetic & Phonological Salience Effects in Different Speech Processing Tasks

Maya L. Barzilai
Georgetown University

1 Introduction

This study examines the relative effects of phonetic and phonological salience in speech sound processing. Three experiments are reported, each examining the relative processing of two speech sounds by speakers of two languages. In each experiment, one sound is more phonetically salient, and the other is assumed to be more phonologically prominent in one of the languages, based on its morphophonological patterning in that language. In addition to examining different pairs of speech sounds, the three experiments also employ different methods, ranging from a short-term recall task to a longer-term artificial language learning task. The results presented here suggest that phonetics and phonology can exert separate effects on processing, and that the two effect types interact differently in different types of speech processing tasks.

The following section provides an overview on phonetics and phonology in processing, and §3 briefly introduces the present study. §4 presents the results from Experiment 1, examining the relative processing of aspirated and unaspirated stops; §5 presents the results from Experiment 2, examining the relative processing of high and low tones; and §6 presents the results from Experiment 3, examining the relative processing of consonants and vowels. §7 provides a discussion of the results and their implications, and §8 concludes.

2 Phonetics vs. phonological salience

Phonological systems of the world's languages are often influenced by the relative perceptibility of the segments involved, such that licit structures are those in which the segments that comprise them are maximally perceptible (e.g., Steriade, 1999; Mielke, 2002). This generalization is operationalized in the Perceptibility-map (P-map; Steriade, 2001), which generates phonological constraints based on the relative perceptibility of segments in different phonological contexts.

Though perceptibility has been shown to play an active role in the grammar, the relationship between phonology and perception is bidirectional (Hume & Johnson, 2001). Foundational work on the interaction between phonology and speech perception shows that speakers exhibit categorical perception, distinguishing only between sounds that are contrastive in their native language (Abramson & Lisker, 1970). These effects are famously seen very early in language acquisition, with infants that have achieved a certain stage of phonological acquisition failing to discriminate among sounds that are not contrastive in their native language (Werker et al., 1981; Werker & Tees, 1984a). The effects of a listener's native language on perception are not limited to its phonological inventory. Mielke (2002) shows that a listener's likelihood of correctly perceiving /h/ in a word is based on the phonological patterning of /h/ in that listener's native phonology; speakers of languages that allow for non-prevocalic /h/ were the most likely to perceive /h/ in this context, whereas speakers of languages in which /h/ cannot surface pre-vocalically did not perceive /h/ in this context. Similarly, speakers perceptually repair stimuli that are phonotactically illicit in their native language, perceiving illusory consonants in between consonants in a stimulus if their native language does not permit consonant clusters, but perceiving the cluster correctly if their native language does permit consonant clusters (Dupoux et al., 1999). Non-segmental phonological processes also impact perception; work on the perception of tone by native Mandarin speakers shows that the most confusable tones are those that are allophonic in Mandarin, despite the fact that these tones are not the most phonetically similar (Hume & Johnson, 2003; Huang & Johnson, 2010).

* I am grateful to Elizabeth Zsiga, Hannah Sande, Adam Ussishkin, and the audience at AMP 2020 for feedback on this project.

In addition to the many effects of phonology on perception, there are also experimental findings that show that the effect of phonology is not always present in speech perception. For example, in early work documenting categorical perception, adults were able to discriminate among sounds that were not contrastive in their language when they were told that they would be asked to perceive non-speech sounds or sounds in a different language (Werker & Tees, 1984b). Relatedly, Babel & Johnson (2010) report that though speakers of different languages processed fricatives differently from each other, these differences narrowed when the participants were instructed to make their discrimination judgements in under 500ms. These findings suggest that though a listener's native phonology leads to categorization and prevents fine-grained discrimination, there is no loss of the physical ability to process auditory percepts; rather, certain experimental methodologies can encourage the listener to avoid phonological perception and instead access purely phonetic processing.

This phonetic processing is inherently separate from phonological processing, and therefore carries its own set of predictions. Many have argued that in the absence of a phonological effect, acoustic salience predicts how sounds are perceived, such that more acoustically salient sounds are more easily processed or more accurately perceived than less acoustically salient sounds (e.g., Crowder, 1971; Cutler et al., 2000; Silverman, 2003). Though acoustic salience does not correspond to one objective acoustic measure, and therefore can be a nebulous concept, acoustic properties such as intensity and duration have been said to contribute to a sound's overall acoustic salience.

Importantly, since phonetics and phonology exert separate effects on perception, it follows that these effects can be in direct opposition to one another. Given two speech sounds, it can be the case that phonology predicts one outcome regarding the relative processing of these sounds, and that phonetics predict the opposite perceptual outcome. The literature discussed above also points to an impact of experimental design on perceptual results. This study investigates these generalizations, investigating which effect—that of phonetics or phonology—is stronger in speech perception tasks, and also asking whether the nature of the task influences these relative effects.

3 The present study

The remainder of this paper provides the results from three experiments, each examining the relative processing of two speech sounds by speakers of two different languages. In each of the experiments, one of the speech sounds is considered more acoustically salient than the other, based on acoustico-perceptual data, as well as on findings from the phonological typology and language acquisition literatures. The other sound in each experiment is more phonologically salient in the native language of one of the speaker groups; phonological salience is determined by the phonological or morphophonological patterning of the given sound in the grammar of said language. If the effect of phonetic salience is stronger than that of phonological salience, it can be predicted that the more phonetically salient sound in the pair will be processed more accurately than the less phonetically salient sound by both speaker groups. If, on the other hand, the effect of phonological salience can outweigh the phonetic effect, it can be predicted that the speakers of the language in which the less phonetically salient sound is more phonologically salient will process that sound more accurately. In other words, an effect of phonological salience would reveal a difference in relative processing of the two sounds between the two speaker groups.

Each experiment also uses a different task to investigate the relative processing of the two sounds. The tasks range from short-term to long-term, and from requiring less to more phonological processing. It is hypothesized that the differences in tasks will lead to a difference in the relative effects of phonetics and phonology, such that each effect may emerge in some but not all tasks.

For more details on the experiments reported here, as well as on related sets of experiments, see Barzilai (2020).

4 Experiment 1

The first experiment investigates the relative processing of aspirated and unaspirated stops. Unaspirated stops are stops that have a voice onset time (VOT) of around 0 ms, and aspirated stops have a longer VOT. The noisy airflow that occurs between the oral release in an aspirated stop and the onset of voicing for the following vowel is assumed to be perceptually salient (Silverman, 2003). This salient period of aspiration also contributes to the overall duration of the stop, leading to overall "greater phonetic richness" (Kim et al., 2012:p.444) relative to unaspirated stops. In other words, aspirated stops are more phonetically salient than

their unaspirated counterparts.

Despite the high acoustic salience of aspirated stops, there are languages with phoneme inventories that do not contain aspirated stops at all. Among these languages in Spanish, which contrasts only voiced stop, which have negative VOT, and voiceless stops, which in Spanish have a VOT of around 0 ms (Lisker & Abramson, 1964).¹ Given that unaspirated stops are the only voiceless stops in the Spanish phoneme inventory, it can be assumed that these are more phonologically prominent to a Spanish listener than the acoustically salient aspirated stops.

This experiment investigates the relative effects of the acoustic salience of aspirated stops and the phonological salience, for Spanish listeners, of unaspirated stops. The control group in this experiment is comprised of Thai speakers, as aspirated and unaspirated stops are both phonemic in Thai (Lisker & Abramson, 1964; Tingsabath & Abramson, 1993; Tsukada & Roengpitya, 2008); any difference in the processing of aspirated and unaspirated stops among this speaker must be due to phonetic effects, as the Thai phonology does not privilege one stop type over the other.

4.1 Participants The participants in this experiment were 20 native speakers of Spanish and 19 native speakers of Thai, all over the age of 18. Participants were recruited in the Washington, DC area, and were all at least moderately proficient in American English.

4.2 Materials The stimuli in this experiment were sequence of 6 CV syllables, comprised of the inventory /p t k p^h t^h k^h m s l i u a/. Each stimulus sequence contained the same vowel in all syllables, and either all aspirated stops (e.g., /ma k^ha p^ha t^ha la k^ha/) or all unaspirated stops (e.g., /ma ka pa ta la ka/). Filler sequences had the same consonant in all syllables, with the vowels differing (e.g., /sa si su sa su si/). All stimulus syllables were recorded once by a native speaker of Korean, a language with both stop types in its phonemic inventory.²

4.3 Methods The experiment was run on a laptop computer using PsychoPy (Peirce, 2007), in a sound-attenuated booth in Georgetown University's Linguistics Lab. In an effort to maximally prime Spanish phonology, Spanish speakers were given all experimental instructions by an advanced Spanish speaker.³ All written instructions were provided in the participants' native language for both speaker groups.

Stimulus sequences were presented auditorily on a laptop computer. The laptop screen was gray while each sequence played. The laptop screen turned blue 1500 ms after the end of each sequence; participants were instructed to repeat the sequence aloud once the screen was blue. After 8 seconds of response time, the screen turned gray again the following sequence was played. All stimuli were randomized for each participant.

Responses were recorded and transcribed by a native English speaker; transcriptions were checked by a different native English speaker. Transcribed syllables were coded for accuracy. Each syllable received one point if it was correctly reproduced and zero points otherwise. Coding did not measure the VOT produced by the speakers, but rather only accounted for whether the place of articulation in each response syllable was the same as that in the corresponding stimulus syllable. In other words, coding measured whether the presence of aspiration on a stop impacted whether the place of articulation of that stop was correctly recalled. This methodology was used to avoid penalizing Spanish speakers for not reproducing differences in VOT that are

¹ Some varieties of Spanish have been shown to have surface aspirated stops; these only surface, though, as a result of underlying /sC/ clusters, in which the /s/ is debuccalized and the stop becomes aspirated (Torreira, 2007). Therefore, even in the Spanish varieties that have surface aspirated stops, these stops are not phonemic but rather surface as the result of a series of phonological processes.

² The Korean speaker produced the aspirated and lenis stops, and not the Korean tense stop, as Korean is undergoing tonogenesis such that the main cue to the lenis/tense contrast is F₀ and not VOT (Silva, 2006). Korean aspirated and lenis stops correspond most closely to a canonical aspirated/unaspirated contrast.

³ Given that the crucial linguistic property of Spanish in this experiment is the lack of aspirated stops, and that the speakers were also proficient in English, which contains aspirated stops, a study was conducted to verify that the Spanish speakers produced unaspirated stops when speaking Spanish (Barzilai, Submitted). The findings showed that the Spanish speakers did in fact have an average VOT of about 23 ms when producing phonologically voiceless stops in Spanish, a value that matches previously reported Spanish VOT data (Lisker & Abramson, 1964). Though this result does not necessarily reflect any the perceptual patterns of these Spanish speakers, it at least reveals that these speakers' proficiency in English does not render their Spanish production different from that of monolingual Spanish speakers.

	<i>Aspirated</i>	<i>Unaspirated</i>
<i>Spanish</i>	0.571 (0.012)	0.553 (0.012)
<i>Thai</i>	0.626 (0.012)	0.572 (0.012)

Table 1: Mean syllable score (SE) by L1 and stop type

not contrastive in their native language.

Transcribed responses that were not exactly six syllables in length were re-aligned so that the final syllable produced in the response was scored as corresponding to the final syllable in the stimulus sequence (see Barzilai, 2020 for more details on this scoring scheme). This methodology has been used for a previous experiment with a near-identical design (Barzilai, 2019), and is intended to avoid obscuring any recency effects in the recall of these sequences (Crowder, 1971; Frankish, 1996).

4.4 Results Table 1 shows the mean scores in this experiment by participant L1 and aspiration type.

Figure 1 shows the mean recall scores. Both groups had a mean accuracy of about 0.56 when recalling unaspirated syllables, and remembered aspirated syllables with a slightly higher accuracy. Thai speakers had higher mean accuracies when remembering aspirated syllables compared to Spanish speakers.

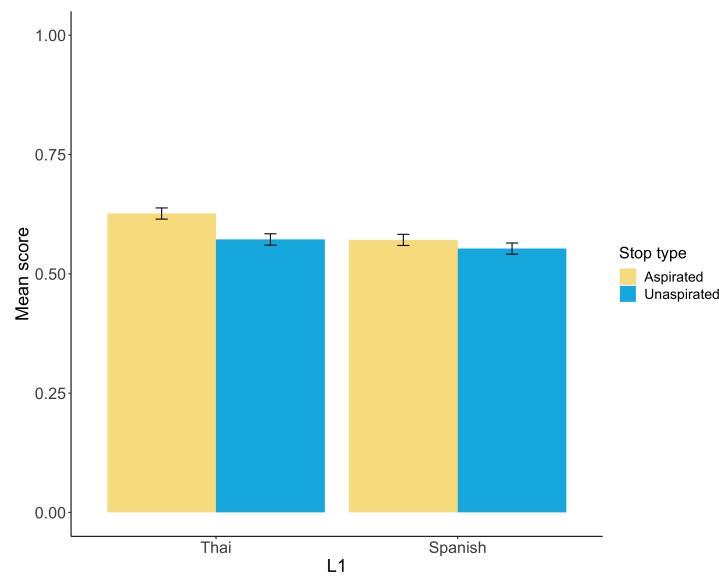


Figure 1: Recall scores by L1 and aspiration type

A mixed-effects logistic regression model was fit using the `glmer` function in the `lme4` R package (Bates et al., 2015) to predict mean syllable accuracy on this task (Table 2). The model found no significant main effect of L1 ($p = 0.5173$). Though the pairwise comparisons show no significant difference between aspirated and unaspirated recall for Spanish speakers ($p = 0.8332$), and only a marginally significant difference for Thai speakers ($p = 0.0687$), the regression model reveals a significant main effect of aspiration type ($p = 0.0166$) for the data as a whole, such that aspirated stops were significantly easier to recall than unaspirated stops. The interaction between aspiration and L1 in this task was only very marginally significant ($p = 0.0934$), with Spanish speakers remembering aspirated stops slightly less accurately than Thai speakers.

Syllable position in Table 2 was modeled as initial, medial, or final. This grouping into three levels is in keeping with the finding that recall tasks such as this one show ‘bowl-shaped’ results, such that not only are final elements easier to recall than medial ones, as discussed above, but initial elements are also easier to recall than medial ones (Crowder, 1971; Frankish, 1996). This latter effect type is known as a primacy effect.

<i>Fixed Effects</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>p-value</i>	
(Intercept)	0.1837	0.1203	1.526	0.1269	
<i>Aspiration</i>					
Aspirated	0.2515	0.1050	2.395	0.0166	*
<i>L1</i>					
Spanish	-0.0963	0.1487	-0.647	0.5173	
<i>Aspiration * L1</i>					
Aspirated:Spanish	-0.1699	0.1013	-1.678	0.0934	.
<i>Position</i>					
Initial	1.2228	0.08	15.283	<0.001	***
Final	-0.2911	0.0672	-4.334	<0.001	***

Table 2: Mixed-effects logistic regression model: recall accuracy. Thai as reference level for L1; unaspirated as reference level for aspiration type; medial as reference level for syllable position. Speaker and syllable sequence as random effects.

As shown in Table 2, recall of initial syllables was significantly higher than that of medial syllables ($p < 0.001$), revealing a primacy effect in this experiment. However, though there is a significant difference in the mean accuracies of medial and final syllables ($p < 0.001$), it is the medial syllables that are more likely to be correctly recalled. In other words, there is no recency effect in this experiment; the positioning of a syllable at the end of a sequence does not facilitate its recall.

4.5 Discussion The results of this experiment suggest a phonetic effect in the processing of aspirated and unaspirated nouns. The significant main effect of aspiration ($p = 0.0166$) reveals that aspirated stops were significantly easier to remember than unaspirated stops across all participants. The lack of a significant interaction between aspiration and L1 ($p = 0.0934$) suggests that the relative recall of aspirated and unaspirated stops did not differ based on participants' L1. In other words, there is no evidence here for an effect of Spanish phonology on the relative recall of the two stop types.

The results also show an interesting effect of syllable position, in which initial syllables were significantly easier to remember than medial syllables, which were in turn easier to remember than final syllables ($p < 0.001$). In other words, there was a primacy effect (Crowder, 1971; Frankish, 1996) in these results, but no recency effect. Though the latter result is surprising, it mirrors those from other recall experiments with effectively identical methodologies (Barzilai, 2019, 2020).

5 Experiment 2

The second experiment in this study investigates the relative processing of high (H) and low (L) tones. There is a cross-linguistic tendency for H tones to co-occur with metrical prominence (De Lacy, 1999, 2007), which has led researchers to posit that a tonal prominence scale in which H tones are more prominent than lower tones. The notion that H tones are more perceptually prominent also emerges from the first and second language acquisition literatures. For instance, infants acquiring Yoruba, a tone language, can discriminate H tones from other tones, but are not as successful at discriminating among non-H tones (Harrison, 1998). It has also been shown that learners acquiring a tone language may attend more to higher F_0 targets, therefore acquiring H tones faster than tones with lower targets (Riesterberg, 2017). Though these claims do not come directly from the acoustic properties of the different tones, they combine to suggest that H tones are more acoustically salient than L tones.

Despite this generalization, however, there are languages in which L tones are the only active tone in the phonology, and H tones surface by default on underlyingly toneless syllables. This type of language is referred to in the literature as L-marked (e.g., Hyman, 2001, 2007). T̥ɬchɔ̃ (ISO 639-3 dgr)⁴, an endangered

⁴ T̥ɬchɔ̃ is often referred to in the literature as Dogrib; T̥ɬchɔ̃ is used here as this is the name preferred by the speaker community.

and under-documented Northern Athabaskan Dene language spoken in the Northwest Territories, Canada, is an example of an L-marked language; L tones in Tłı̨cẖ are active in phonological processes (see discussion in Barzilai, Forthcoming), and H tones surface only on syllables that are unspecified for tone (Hyman, 2001; Krauss, 2005; Jaker, 2012). In other words, separate from the relative acoustic salience of H and L tones, it can be assumed that it is the L tones that are the phonologically salient tones to speakers of Tłı̨cẖ.

This experiment investigates the relative effects of the acoustic salience of H tones and the phonological salience, for Tłı̨cẖ speakers, of L tones. The control group in this experiment is comprised of French speakers, as French is a non-tone language, and it has been shown that French speakers do not rely on F_0 as a cue to phonological prominence (Dupoux et al., 1997; Frost, 2011). In other words, it can be assumed that for French speakers, any difference in the processing of H and L tones must be due a phonetic effect, as the French phonology does not create a bias towards either tone height.

5.1 Participants The participants in this experiment were 17 native speakers of French and 14 native speakers of Tłı̨cẖ, all over the age of 18. French speakers for this test case were recruited in the Washington, DC area, and Tłı̨cẖ speakers were recruited and participated in Canada's Northwest Territories. All participants in this study were also proficient in North American English.

5.2 Materials The stimuli in this experiment were sequences of 6 CV syllables, comprised of the inventory /p t s i u a/. A native Thai speaker produced the nine syllables generated from this inventory once with each of the five lexical Thai tones; the L and H⁵ level tones were extracted from the resulting recording and used to generate the sequences tested here.

Stimulus sequences contained only H or L tones, with at least two H syllables and at least two L syllables, in varying orders, in each. There were no more than two consecutive syllables hosting the same tone in any stimulus sequence. Each stimulus sequence was followed by a test syllable. This test syllable either matched one of the syllables in the stimulus sequence or did not match any of the stimulus syllables. Matching test syllables were segmentally and tonally identical to one of the syllables in the sequence. Importantly, the test syllables were resynthesized to have the tone contour from another utterance of a syllable with the same tone; as a result, the test syllables were acoustically distinct from the syllable they matched. Non-matching test syllables were segmentally distinct from all of the syllables in the sequence.

5.3 Methods The experiment was run on a laptop computer using PsychoPy (Peirce, 2007). French speakers participated in the experiment in a sound-attenuated booth in Georgetown University's Linguistics Lab; Tłı̨cẖ speakers participated in the experiment in a quiet office in the Tłı̨cẖ government offices in Behchoḵ, Northwest Territories, Canada.

Stimulus sequences were presented auditorily on a laptop computer; test syllables played approximately 1500 ms after the end of the stimulus sequence. The participant was told that their task was to determine whether the test syllable they heard was the same as one of the syllables they heard in the sequence or not. The right and left arrows on the computer keyboard were used as the response keys; the key corresponding to a matching syllable was counterbalanced across participants. All sequences were randomized across the testing phase. There were three practice sequences before the beginning of the actual testing portion of the experiment, neither of which was repeated during the testing phase.

Keyboard responses were recorded and coded for accuracy. Responses received one point if the participant correctly indicated whether the test syllable matched a syllable in the previous stimulus sequence; incorrect responses received zero points.

5.4 Results An initial examination of the data revealed that one of the native Tłı̨cẖ speakers produced the same response for all trials in this experiment, suggesting that they did not understand the task; this person was removed from the analysis. Similarly, one participant failed to give a response for over 15 of the trails in this experiment and therefore was also removed. The results below are from the remaining 12 speakers.

Table 3 shows the mean scores in this experiment by participant L1 and target syllable tone.

Figure 2 shows the mean recall scores. Both groups had a mean accuracy of approximately 0.60 when

⁵ It was in fact the phonologically falling tone in Thai that was produced by the speaker with the most level high pitch and therefore used for the stimuli of this experiment. They are referred to here as H for clarity, as their phonological identity in Thai is irrelevant to this experiment.

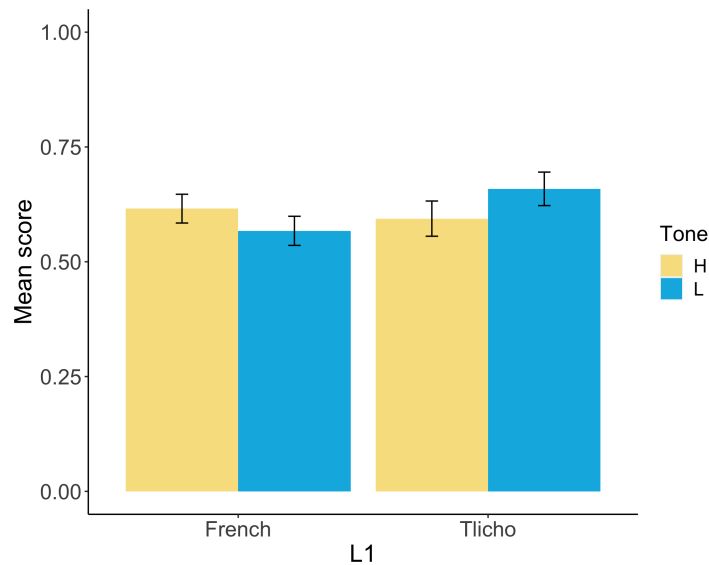
	<i>H</i>	<i>L</i>
<i>French</i>	0.616 (0.031)	0.567 (0.032)
<i>Tɨchɔ</i>	0.594 (0.038)	0.659 (0.036)

Table 3: Mean score (SE) by L1 and target syllable tone

<i>Fixed Effects</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>p-value</i>
(Intercept)	0.5431	0.321	1.690	0.0910
<i>Tone</i>				
<i>L</i>	-0.2593	0.2802	-0.925	0.3548
<i>L1</i>				
<i>Tɨchɔ</i>	-0.0597	0.4562	-0.131	0.8959
<i>Tone * L1</i>				
<i>L:Tɨchɔ</i>	0.6612	0.3290	2.010	0.0444 *

Table 4: Mixed-effects logistic regression model: recall accuracy. French as reference level for L1; H as reference level for target syllable tone. Speaker and syllable sequence as random effects.

recalling H syllables. Within the groups, French speakers had a lower mean score when recalling L syllables, whereas Tɨchɔ speakers had higher mean scores when recalling L syllables.

**Figure 2:** Recall scores by L1 and target syllable tone

A mixed-effects logistic regression model was fit using the `glmer` function in the `lme4` R package (Bates et al., 2015) to predict mean score on this task (Table 4). No significant main effect of target syllable tone or L1 was found. However, the interaction between target syllable tone and L1 was significant ($p = 0.0368$); the relative means of H and L accuracy was significantly different for Tɨchɔ speakers than for French speakers. Though the pairwise comparison revealed no significant difference between recall rates for H versus L tones for the French speakers ($p = 0.768$) or for Tɨchɔ speakers ($p = 0.563$), this significant interaction implies that the relationship between H tone recall and L tone recall was significantly different across the L1 groups.

5.5 Discussion The statistically significant interaction between L1 and test syllable tone ($p = 0.0444$) reveals that syllables with H and L tones were correctly remembered at significantly different rates for each of the speaker groups. This suggests an effect of phonology in tone processing, such that speakers of L-marked languages remember L tones more often than H tones, and that this pattern is opposite than that of speakers with no such phonological bias.

These results may also provide evidence for an effect of acoustic salience in tone processing. Pairwise comparisons show that French and Tḥçḥ speakers remembered H tones at equal rates ($p = 0.999$). In other words, the acoustic salience of H tones appears to be such that all speakers, so long as their L1 does not further facilitate H recall, remember them equally well. The difference between the groups then comes from the fact that Tḥçḥ speakers are impacted by an additional effect of the phonological prominence of L tones, which boosts their recall of L-toned syllables.

6 Experiment 3

The third and final experiment in this study investigates the relative processing of consonants and vowels. Vowels are said to be more acoustically salient than consonants, due to their relatively long duration and acoustic steady state (Crowder, 1971; Cutler et al., 2000). The relatively high acoustic salience of vowels as compared to consonants has been confirmed by recall experiments, similar to the ones detailed above, in which vowels were more accurately reproduced than consonants when presented either visually (Drewnowski, 1980) or auditorily (Crowder, 1971; Kissling, 2012; Barzilai, 2019).

Despite the high acoustic salience of vowels, there are languages in which the consonants are solely responsible for conveying lexical information, and vowels carry only morphosyntactic information (Ryding, 2005; Nespor et al., 2003; Toro et al., 2008). Hebrew is one of these languages, with a root-and-pattern morphological system in which the lexical root is comprised entirely of consonants. This morphophonological property can be assumed to increase the phonological salience of consonants for Hebrew speakers.

This experiment investigates the relative effects of the acoustic salience of vowels and the phonological salience, for Hebrew speakers, of consonants. The control group in this experiment is comprised of German speakers, as German has a non-root-and-pattern morphological system, in which both consonants and vowels provide lexical root information. In other words, it can be assumed that for German speakers, any difference in the processing of vowels and consonants must be due to a phonetic effect, as the German morphophonology does not create a bias towards one segment type or the other.

6.1 Participants The participants in this experiment were 20 native speakers of German and 28 native speakers of Hebrew. German speakers were recruited in the Washington, DC area and Hebrew speakers were recruited in Tel Aviv, Israel. All speakers had at least some proficiency in English.

6.2 Materials The stimuli in this experiment comprised two separate artificial languages, a vowel-variable language and a consonant-variable language. Words in each language were produced by a speaker of Minnesota English.

Each participant learned only one of the two artificial languages. In the vowel-variable language, the stimuli were all of the shape /tVkV/, with the consonants held constant throughout and one of the vowels from the inventory /a e i o u/ appearing in each of the two vowel slots (e.g., /tika/, /tuko/, /teki/, etc.). No stimulus had the same vowel in both vowel positions (i.e., /taka/ was not a stimulus in the vowel-variable language). Fillers were words with consonants other than /t/ and /k/, but also containing two different vowels (e.g., /sima/, /mulo/, etc.).

In the consonant-variable language, the stimuli were all of the shape /CaCi/, with the vowels held constant throughout and one of the consonants /t k z m b/ appearing in each of the consonant slots (e.g., /tami/, /mabi/, /zaki/, etc.). No stimulus had the same consonant in both consonant positions (i.e., /tati/ was not a stimulus in the consonant-variable language). Fillers in this language were words with vowels other than /a/ and /i/, but also containing two different consonants (e.g., /tomu/, /kezo/, etc.). All stimuli were recorded by a native speaker of American English.

6.3 Methods The experiment was run on a laptop computer using PsychoPy (Peirce, 2007). German and Amharic speakers participated in the experiment in a sound-attenuated booth in Georgetown University's

	<i>Vowel-Variable</i>	<i>Consonant-Variable</i>
<i>German</i>	0.712 (0.032)	0.833 (0.027)
<i>Hebrew</i>	0.750 (0.026)	0.850 (0.021)

Table 5: Mean score (SE) by L1 and experimental language

Linguistics Lab; Hebrew speakers participated in the experiment in a quiet office on the Tel Aviv University campus.

Participants were told that their task was to learn the names of objects in a new language. Each stimulus was associated with an image of an object in each language; the same object images were used for both languages. In the training phase, the participants heard a stimulus while the accompanying image was displayed on the screen. Each image remained on the screen for three seconds, with an interstimulus interval of one second. The participant was exposed to each stimulus and corresponding image two times, in a randomized order. In the testing period, participants saw one of the objects from the training period and heard two words. One of the words was the name for the object and the other word was a distractor word. Distractor words were other non-filler words in the experimental language. Participants were instructed to press the 1 key on the keyboard if the first word heard was the correct name for the object and the 2 key if the second word was correct. The order of the correct and incorrect words was counterbalanced across trials.

Keyboard responses were recorded and coded for accuracy. Responses received one point if the participant indicated the correct word for the image; incorrect responses received zero points.

6.4 Results Table 5 shows the mean scores in this experiment by participant L1 and experimental language.

Figure 3 shows the mean ALL scores. The German and Hebrew speakers performed almost identically on this task, with a mean score for the vowel-variable language of around 0.725 and a mean score for the consonant-variable language of around 0.85.

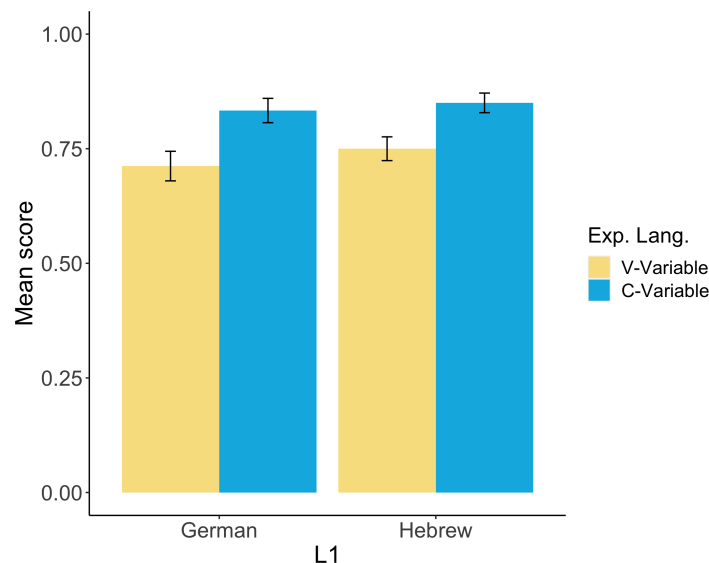


Figure 3: ALL scores by L1 and experimental language

A mixed-effects logistic regression model was fit using the `glmer` function in the `lme4` R package (Bates et al., 2015) to predict mean word accuracy on this task (Table 6). Though the pairwise comparisons did not reveal statistically significant differences within each L1 group (German: $p = 0.2104$; Hebrew: p

<i>Fixed Effects</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>p-value</i>	
(Intercept)	1.0296	0.3283	3.137	0.002	**
<i>Exp. Lang.</i>					
C-Variable	0.9006	0.4809	1.873	0.0611	.
L1					
Hebrew	0.2512	0.4001	0.628	0.5300	
<i>Exp. Lang. * L1</i>					
C-Variable:Hebrew	-0.1108	0.5885	-0.188	0.8507	

Table 6: Mixed-effects logistic regression model: ALL accuracy. German as reference level for L1; vowel-variable as reference level for experimental language. Speaker and word as random effect.

= 0.1855), there was a marginally significant main effect of experimental language ($p = 0.0611$), with the consonant-variable being significantly easier to learn than the vowel-variable language for all groups. The model found no main effect of L1, showing that the language groups performed this task with equivalent accuracy.

Crucially, the interaction between experimental language and L1 was not significant; the relationship between vowel-variable language accuracy and consonant-variable language accuracy was equivalent across L1 groups.

6.5 Discussion The lack of a significant interaction between L1 and experimental language type here ($p = 0.8507$) indicates that the difference in scores between each experimental language is not significantly different depending on the language. Rather, the marginally significant main effect of experimental language ($p = 0.0611$) shows that both groups learned the C-variable language with more accuracy than the V-variable language. This result is perhaps not surprising for the Hebrew speakers, as it reveals a phonological effect, such that the more phonologically prominent segment type was more easily learned.⁶

This result is surprising, however, for the German speakers. Given that German does not employ a root-and-pattern morphological system, and rather has lexical roots comprised of both consonants and vowels, it was hypothesized that any effect among these speakers would be phonetic; the V-variable language was expected to be more easily learned by the German speakers because vowels are more phonetically salient than consonants. This surprising result is best explained by the CV hypothesis, a cross-linguistic generalization that consonants tend to convey lexical information and vowels tend to encode morphosyntactic information (Nespor et al., 2003). Though Hebrew is an extreme example of this generalization, with only consonants carrying lexical information, the generalization still holds, albeit more weakly, for German. In other words, it is not the case that the German grammar does not have any property that would facilitate consonant learning, it is merely that the Hebrew grammar was predicted to even more strongly facilitate consonant learning because of its root-and-pattern morphology.

7 Discussion & Implications

7.1 Task-dependent phonetic and phonological effects Results from experiments 1 (§4) and 2 (§5) show effects of phonetic salience on speech processing. In experiment 1, all participants performed better on the more phonetically salient sound, regardless of their native language. In experiment 2, though there was a statistically significant difference in performance between the two language groups, the effectively equal recall of H tones may suggest that an effect of phonetic salience was at play. Taken together, these two results suggest that though acoustic salience may not have a quantifiable spectral correlate, it can be detected and measured experimentally.

In addition to evidence for phonetic salience, results from experiments 2 (§5) and 3 (§6) both show

⁶ This experiment was also designed to investigate whether Hebrew's orthography biases the listener towards consonants; no evidence of an orthographic effect was found (see details on complete experiment in Barzilai (2020); see Barzilai (2019) for discussion of a potential orthographic effect for Arabic speakers.)

evidence of a phonological effect on processing. In experiment 2, the statistically significant interaction between language and syllable tone show that Tḥçḳ's L-markedness makes L-toned syllables more accurately remembered by its speakers than they are for speakers of non-tone languages. In experiment 3, the phonological effect was found among both speaker groups, arising from the cross-linguistic tendency for consonants to carry lexical information, and therefore for the C-variable artificial language to be more natural and easier to learn than the V-variable one. These results show that phonetic effects on processing, described above, can be overridden by effects of phonological salience.

The experiments reported here increase both in time required to complete the task and in type of processing required: experiment 1 requires the repetition of syllables 1500 ms after they are heard, experiment 2 requires remembering syllables and making a choice about their correspondence, and experiment 3 requires learning words in an artificial language over the course of approximately 20 minutes. Having established this continuum, the results from all three experiments together suggest that phonetic effects are more likely to emerge in shorter-term tasks, and phonological effects are active in tasks that require more time, or tasks that require the formation of a form/meaning mapping. This finding is not only important to the understanding of speech processing, but also raises an important methodological point; the absence of phonetic or phonological effects in experimental research may be an artifact of task choice, and may not prove that these effects do not in fact exist.

7.2 Limitations Due to the practical constraints of conducting a set of experiments with a diverse set of participants, only a subset of the speakers participated in this study in a sound-attenuated booth. Rather, the Tḥçḳ speakers participated in the study in a conference room and the Hebrew speakers in an office with several windows. Though the data from these speakers was as easily coded and scored as that from other speakers, it is possible that this difference presents an experimental confound.

Several Tḥçḳ speakers reported low levels of exposure to and familiarity with computers. Additionally, though some of them had heard of other community members having worked with linguists, this previous work had always been focused on description and documentation; anecdotally, some speakers reported surprise at the fact that their work with a linguist did not involve speaking their own language but rather responding to nonce sounds on a computer. This is in stark contrast to the rest of the participants in this study, who were recruited through a university and were overall more likely to have participated in or even conducted experimental research. Though this may have led to overall lower results by the Tḥçḳ speakers compared to the other speakers, the results are nonetheless interpretable, and it is unlikely that the major findings are skewed by this confound.

8 Conclusion

This study presents results from three psycholinguistic experiments, each examining the relative effects of phonetics and phonology on speech sound processing. Results show evidence for both a phonetic and a phonological effect in processing. Crucially, the results suggest that the rates at which these effects influence processing results are mediated by the task, such that shorter-term tasks may be more susceptible to phonetic effects while tasks requiring longer-term, more phonological processing may be more susceptible to phonological effects.

References

- Abramson, Arthur S & Leigh Lisker (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the sixth international congress of phonetic sciences*, Academia Prague, vol. 196, 569–573.
- Babel, Molly & Keith Johnson (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory phonology* 1:1, 179–205.
- Barzilai, Maya L. (2019). Effects of templatic morphology on segmental recall. Paper presented at the Linguistics Society of America annual meeting.
- Barzilai, Maya L. (2020). *The Relative Effects of Phonetic and Phonological Salience in Speech Sound Processing*. Ph.D. thesis, Georgetown University.
- Barzilai, Maya L. (Forthcoming). Phonetic and phonological salience in tone processing. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique*.
- Barzilai, Maya L. (Submitted). Phonetic salience in processing of stop aspiration. Manuscript submitted for publication.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1, 1–48.

- Crowder, Robert G (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior* 10:6, 587–596.
- Cutler, Anne, Nuria Sebastián-Gallés, Olga Soler-Vilageliu & Brit Van Ooijen (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & cognition* 28:5, 746–755.
- De Lacy, Paul (1999). Tone and prominence. *Rutgers Optimality Archive* 333.
- De Lacy, Paul (2007). The interaction of tone, sonority, and prosodic structure. *The Cambridge handbook of phonology* 281–307.
- Drewnowski, Adam (1980). Memory functions for vowels and consonants: A reinterpretation of acoustic similarity effects. *Journal of Verbal Learning and Verbal Behavior* 19:2, 176–193.
- Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian & Jacques Mehler (1997). A destressing “deafness” in French? *Journal of Memory and Language* 36:3, 406–421.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of experimental psychology: human perception and performance* 25:6, p. 1568.
- Frankish, CR (1996). Auditory short-term memory and the perception of speech. *Models of short-term memory* 179–207.
- Frost, Dan (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association* 41:1, 67–84.
- Harrison, PA (1998). Yoruba babies and unchained melody. *UCL WPL* 10, 33–50.
- Huang, Tsan & Keith Johnson (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica* 67:4, 243–267.
- Hume, Elizabeth & Keith Johnson (2001). A model of the interplay of speech perception and phonology.
- Hume, Elizabeth & Keith Johnson (2003). The impact of partial phonological contrast on speech perception. *Proceedings of the fifteenth international congress of phonetic sciences*, vol. 2385.
- Hyman, Larry M (2001). Privative tone in Bantu. *Cross-linguistic studies of tonal phenomena* 237–257.
- Hyman, Larry M (2007). Universals of tone rules: 30 years later. *Tones and tunes: Studies in word and sentence prosody* 1–34.
- Jaker, Alessandro Michelangelo (2012). *Prosodic reversal in Dogrib (Weledeh dialect)*. Stanford University.
- Kim, Sahyang, Taehong Cho & James M McQueen (2012). Phonetic richness can outweigh prosodically-driven phonological knowledge when learning words in an artificial language. *Journal of Phonetics* 40:3, 443–452.
- Kissling, Elizabeth M (2012). Cross-linguistic differences in the immediate serial recall of consonants versus vowels. *Applied Psycholinguistics* 33:3, 605–621.
- Krauss, Michael (2005). Athabaskan tone. *Athabaskan prosody* 269, p. 55.
- Lisker, Leigh & Arthur S Abramson (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:3, 384–422.
- Mielke, Jeff (2002). Turkish/h/deletion: Evidence for the interplay of speech perception and phonology. *PROCEEDINGS-NELS*, CiteSeer, vol. 32, 383–402.
- Nespor, Marina, Marcela Peña & Jacques Mehler (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio* 2:2, 203–230.
- Peirce, Jonathan W (2007). Psychopy - psychophysics software in Python. *Journal of neuroscience methods* 162:1-2, 8–13.
- Riesterberg, Katherine (2017). *Acoustic Salience and Input Frequency in L2 Lexical Tone Learning: Evidence from a Zapotec Revitalization Program in San Pablo Macuiltianguis*. Ph.D. thesis, Georgetown University.
- Ryding, Karin C (2005). *A reference grammar of modern standard Arabic*. Cambridge University Press.
- Silva, David J (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 287–308.
- Silverman, Daniel (2003). On the rarity of pre-aspirated stops. *Journal of Linguistics* 39:3, 575–598.
- Steriade, Donca (1999). Alternatives to syllable-based accounts of consonantal phonotactics.
- Steriade, Donca (2001). The phonology of perceptibility effects: the p-map and its consequences for constraint organization. *Ms., UCLA*.
- Tingsabath, MR Kalaya & Arthur S Abramson (1993). Thai. *Journal of the International Phonetic Association* 23:1, 24–28.
- Toro, Juan M, Marina Nespor, Jacques Mehler & Luca L Bonatti (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science* 19:2, 137–144.
- Torreira, Francisco (2007). Pre- and postaspirated stops in Andalusian Spanish. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4* 282, p. 67.
- Tsukada, Kimiko & Rungpat Roengpitya (2008). Discrimination of English and Thai words ending with voiceless stops by native Thai listeners differing in English experience. *Journal of the International Phonetic Association* 38:3, 325–347.
- Werker, Janet F & Richard C Tees (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development* 7:1, 49–63.
- Werker, Janet F & Richard C Tees (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America* 75:6, 1866–1878.
- Werker, Janet F, John HV Gilbert, Keith Humphrey & Richard C Tees (1981). Developmental aspects of cross-language speech perception. *Child development* 349–355.