Perceptual Assimilation of Arabic Voiceless Fricatives by English Monolinguals

Michael D. Tyler & Sarah E. Fenwick
Marcs Institute and School of Social Sciences and Psychology
University of Western Sydney, Sydney, Australia
m.tyler@uws.edu.au, s.fenwick@uws.edu.au

Abstract

Native language experience strongly influences non-native speech discrimination. According to the Perceptual Assimilation Model (PAM) [1], discrimination is most accurate when two non-native sounds map onto different native phonemes (Two-Category assimilation), poorer when they differ in goodness-of-fit to the same native phoneme (Category-Goodness assimilation), and worst when perceived as equally good or poor versions of the same native phoneme (Single-Category assimilation). The Three Factor Model [2] suggests that discrimination accuracy is poorer at a 1500 ms inter-stimulus interval (ISI), when only phonemic information is available, than at 500 ms, when both phonemic and phonetic information is available. To test the models, monolingual English participants assigned Arabic fricatives to English categories, and discriminated contrasting fricative pairs in an AXB task with a 500 ms or 1500 ms ISI. PAM discrimination predictions were upheld, but there was no influence of ISI.

Index Terms: cross-language speech perception, perceptual assimilation, human speech discrimination

1. Introduction

Research into cross-language speech perception has shown that listeners have difficulty discriminating certain non-native phonological contrasts. Early theories suggested that discrimination was only possible if listeners had been exposed to the same phonetic categories in their native language. Contrasts shared by the native and non-native language would be discriminated well, whereas those non-native contrasts not in the native language inventory would be discriminated poorly. In a classic study, [3] showed that native English listeners, with no previous exposure to click languages, were nevertheless able to discriminate Zulu click contrasts because they were not perceived as speech. This study led to the development of the Perceptual Assimilation Model (PAM) [1]. The aims of the present study were to test the predictions of PAM on contrasts from a previously untested language, Arabic, and to test whether PAM can shed further insight on the Three Factor Model (3FM) of Werker and Logan [2].

According to PAM [1], a non-native phone can be 1) heard as speech and assimilated to a native category, with a goodness-of-fit ranging from good to poor, 2) heard as speech but not categorised as any one native category, or 3) not heard as speech. This is assessed using a categorisation task in which participants assign a native-language speech category label to a non-native phone, and then rate its goodness-of-fit to the chosen category. Contrast assimilation patterns are then derived from pairs of contrasting non-native phones. Here we consider three PAM contrast assimilation types: 1) Two-category (TC) assimilations, where the two non-native phones are assimilated to different native categories, 2) Category Goodness (CG), where both phones are assimilated as the same native category, but one is perceived as more phonetically similar to the native category than the other, and 3) Single Category (SC), where both are assimilated as the same native category as equally good or poor versions. A clear discrimination prediction of PAM is that TC should be discriminated more accurately than CG, which should be discriminated more accurately than SC (TC > CG > SC).

A number of studies have provided support for PAM assimilation types and predictions [see 4, 5]. For example, [6] investigated native English speakers’ discrimination of three Zulu contrasts (/ɬ/-/b/, /k/-/k'/, /b/-/b/). Results upheld PAM predictions as the SC contrast /b/-/b/ was discriminated more poorly than the CG contrast, /k/-/k'/, while the TC contrast, /ɬ/-/b/, was the easiest to discriminate. To test whether PAM is a general model of cross-language speech perception, data is required from a variety of language and listener combinations. Therefore, the first aim of the present study was to test PAM discrimination predictions using voiceless Arabic fricatives and Australian English monolingual listeners. Arabic fricatives were chosen as they contrast several constriction locations that are non-native to English. For example, Arabic has glottal /h/, pharyngeal /h/, and velar /x/ fricatives, whereas English only has /h/. Additionally, both Arabic and English contain an alveolar /s/ and post-alveolar /ʃ/ fricatives, while Arabic also contains a pharyngealised alveolar /ɾ/. Combinations of these Arabic categories are likely to result in CG and/or SC assimilations for monolingual English listeners. However, as Arabic has consonant-conditioned allophonic variation of vowels, listeners may successfully discriminate certain consonants on the basis of the surrounding vowel context, rather than on phonetic characteristics of the consonant. To overcome this, we took the novel approach of varying the vowel context within a discrimination trial (e.g., /hɪ/ /hʊ/ /xʊ/).

In addition to testing the basic predictions of PAM, we sought to extend the model by combining it with the 3FM [2]. Werker and Logan proposed that listeners have access to three levels of processing for discrimination of non-native phones: 1) Acoustic—Very small differences can be detected (e.g., differences between tokens of the same category), 2) Phonetic—Differences that signal a phonological difference in a non-native language, but are simply phonetic variants of a single phoneme in the native language, can be detected, 3) Phonemic—Only those differences that signal a phonological difference in the native language can be detected. Importantly, these levels differ in the decay rate of memory traces. The acoustic level is only available up to an interstimulus interval (ISI) of 250 ms, the phonetic level up to 500 ms, and the phonemic level is available for a longer term (although a 1500 ms ISI is generally used). Thus, according to the 3FM, it is variation in ISI that can facilitate or constrain sensitivity to non-native contrasts. In terms of PAM, TC assimilations should not be affected by ISI because differences can be detected on the basis of native-language phonological distinctions. However, both CG and SC differences should be detectable at the phonetic level, and discrimination should be poorer for both at 1500 ms ISI than 500 ms ISI.
2. Method

2.1. Participants

Forty monolingual Australian-English speaking students, from Introductory Psychology at the University of Western Sydney, participated for course credit. Participants were randomly assigned to one of two ISI conditions: 500 ms ISI: 2 males and 18 females, M_age = 22.2 years, range = 18–35 years; 1500 ms ISI: 7 males and 13 females, M_age = 23, range = 18–34 years. An additional 26 participants were tested, but their data were discarded due to prolonged foreign language exposure (n=13), possible hearing, learning or physical impairments (n=11), technical difficulties (n=1), or not following instructions (n=1).

2.2. Stimuli and Apparatus

All speech stimuli were produced by a phonetically trained, 26 year-old native Arabic speaking female from Amman, Jordan. The stimuli were recorded in an anechoic chamber at a 44.1 kHz sampling rate and 16-bit resolution using a Shure SM10A headset microphone, and a laptop computer with an EDIROL UA-25 USB audio capture device.

The six Arabic voiceless fricatives /h, h, s, s, s, / were chosen as certain contrast combinations were likely to produce TC, CG, and SC assimilation types. We selected /β-/s/, /β-/s/, /β-/c/ /h-/s/, /h-/c/, /h-/s/, /h-/h/. The first members of the pair are considered to be more articulatorily similar to an Australian English phonological category than the second member of each contrast pair.

The fricatives were recorded in consonant-vowel (CV) syllables in each of three vowel contexts /a, i, u/. The plosives /d/ and /b/ were also recorded as practice stimuli. Two tokens of each CV syllable were chosen, resulting in 36 CV syllables. The vowels were normalised to 50% of peak intensity, and truncated to 75 ms (following [7]).

Stimulus presentation and response collection was controlled by PsyScope X B53 on MacBook computers, and auditory stimuli were presented binaurally through Sennheiser HD650 headphones. A background questionnaire was used to ensure there was no reported hearing, vision, reading, speaking or learning difficulties, as well as to monitor exposure to languages other than English.

2.3. Procedure

Participants were tested on individual computers in groups of up to four. To familiarise participants with the varying vowel procedure, the six AXB practice trials were completed first, followed by seven AXB discrimination tasks, and a categorisation task. No feedback was provided. The 500 ms ISI condition lasted for 45 minutes and the 1500 ms ISI condition lasted for 60 minutes. At the end of the session the background questionnaire was completed to ensure that the participant had met the selection criteria.

2.3.1. AXB Discrimination

Three CV tokens were presented on each trial and participants were instructed to indicate, by pressing ‘1’ or ‘3’ on the computer keyboard, whether the consonant of the second token of the trial (X) matched the consonant of the first (A) or the third (B) token of that trial. As the vowel varied on each trial, the matching consonant was necessarily a different stimulus token. Once a response had been registered for a trial, the following trials would begin after 1 s. A trial would be randomly repeated if a response had not registered within 3.5 s.

Each of the 7 experimental blocks consisted of 48 trials, and there were four possible AXB trial types (AAB, BBA, BAA, ABB), presented 12 times each. The three vowel contexts (/a, i, u/) were presented on each trial, with the position of the vowel (A, X, or B) counterbalanced across the block. Two tokens of the six CV syllables were presented an equal number of times across each block, and each token occurred an equal number of times in each trial position (A, X, or B). To facilitate counterbalancing, two sets of 24 AXB trial orders were created, such that each contained one set of 6 CV syllables (i.e., 2 contrasting consonants × 3 vowel contexts). The 48 trials were presented in random order and the order of the seven discrimination tasks was counterbalanced across participants using a latin-square design.

Table 1. Mean percent categorisation and mean goodness ratings out of 7 (in parentheses) of each Arabic voiceless fricative, collapsed across stimulus token and vowel context, for participants across both ISI conditions. Boldfaced values indicate the most frequently chosen response label per target.

<table>
<thead>
<tr>
<th>Auditory Stimulus</th>
<th>English Response Label</th>
<th>h</th>
<th>sh</th>
<th>s</th>
<th>k</th>
<th>ch</th>
<th>f</th>
<th>th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal /h/</td>
<td></td>
<td>99.31</td>
<td>0.14</td>
<td>0.21</td>
<td>0.07</td>
<td>0.07</td>
<td>0.21</td>
<td>0.07</td>
</tr>
<tr>
<td>Pharyngeal /h/</td>
<td></td>
<td>99.58</td>
<td>0.07</td>
<td>0.14</td>
<td>0.07</td>
<td>0.21</td>
<td>0.21</td>
<td>(4.75)</td>
</tr>
<tr>
<td>Velar /s/</td>
<td></td>
<td>86.60</td>
<td>0.07</td>
<td>11.25</td>
<td>1.60</td>
<td>0.07</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td>Post-alveolar /β/</td>
<td></td>
<td>0.07</td>
<td><strong>86.04</strong></td>
<td>10.63</td>
<td>3.19</td>
<td>0.07</td>
<td>0.07</td>
<td>(4.00)</td>
</tr>
<tr>
<td>Alveolar /s/</td>
<td></td>
<td>0.07</td>
<td>1.11</td>
<td><strong>98.12</strong></td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>(3.00)</td>
</tr>
<tr>
<td>Pharyngealised</td>
<td></td>
<td>0.07</td>
<td>0.83</td>
<td><strong>98.47</strong></td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>(1.00)</td>
</tr>
</tbody>
</table>
2.3.2. Categorisation

On a given trial, the participant heard a CV syllable and selected using the mouse a native English consonant category from one of nine orthographic response labels (f, h, k, p, s, t, ch, sh, th) on a visually presented 3 × 3 grid. The same speech token was then repeated and participants rated how well the CV syllable matched the chosen English consonant category on a 7-point Likert scale, with ‘1’ indicating “a very strange sounding example of that category”, ‘4’ “an O.K. version”, and ‘7’ “a perfect example of that category”. The participants were instructed to focus on the consonant of the CV syllable, and were not provided feedback for their selected answers. Participants were familiarised with the procedure before commencing by being shown pictures of the stimulus response screens, and verbal examples using English stimuli. The categorisation task consisted of 216 trials (18 CV syllables × 2 stimulus tokens × 6 repetitions), divided into 12 randomised blocks to provide the participant with an optional break.

3. Results

As the PAM predictions for discrimination are contingent on the results of the categorisation task, we report the results of categorisation followed by discrimination.

3.1.1. Categorisation Results

As the assimilation patterns did not differ across ISI condition, for brevity the results are presented in Table 1 for all 40 participants as one group. However, statistics for determining assimilation type were run separately. Table 1 presents the percent English label choice for each Arabic voiceless fricative category (collapsed across vowel context and stimulus token), along with the mean goodness rating (out of 7). A stimulus category was deemed categorised if the percent categorisation was greater than 50%. Thus, the glottal, pharyngeal and velar fricatives were all categorised as English /h/, the post-alveolar as English /ʃ/, and the alveolar and pharyngealised alveolar as English /s/.

The contrast assimilation types for each of the seven discrimination contrasts are listed in the x-axis labels of Figure 1. If each member of the contrast was categorised as a different English phoneme it was deemed to be a TC contrast. Conversely, if both members of the contrast were categorised as the same English phoneme, a significant difference in the goodness ratings would indicate a CG contrast, otherwise it would be deemed an SC contrast (an individual’s mean rating was only included if the relevant label had been chosen more than 50% of the time). The goodness ratings as /h/ did not differ for the glottal versus pharyngeal (500 ms ISI: \( t(38) = 0.36 \); 1500 ms ISI: \( t(38) = 0.07 \)) and ratings as /ʃ/ did not differ for the alveolar versus pharyngealised alveolar, (500 ms ISI: \( t(38) = 1.44 \); 1500 ms ISI: \( t(38) = 1.74 \)). Ratings as /h/ did differ for the glottal versus velar (500 ms ISI: \( t(35) = 2.05 \), \( p < .05 \); 1500 ms ISI: \( t(35) = 2.85 \), \( p < .005 \)) and the pharyngeal versus velar (500 ms ISI: \( t(35) = 4.64 \), \( p < .0001 \); 1500 ms ISI: \( t(35) = 6.90 \), \( p < .0001 \)).

3.1.2. AXB Discrimination Results

Mean percent correct discrimination scores for participants in each ISI condition, collapsed across AXB trial type, are presented in Figure 1. It can be seen that, in general, scores appear to be highest for TC, followed by CG, then SC contrasts, but there appears to be little effect of ISI on accuracy. To test the predictions of PAM and the 3FM, we conducted a mixed ANOVA with planned contrasts. A single between-groups ISI group contrast compared results of the 500 ms and 1500 ms ISI conditions. The remainder were repeated measures contrasts. The phonological boundary contrast compared the overall accuracy of TC versus CG contrasts, and the phonetic sensitivity contrast compared CG and SC contrasts. As these statistical contrasts are not orthogonal, a Bonferroni-adjusted alpha rate of .025 was used for those two statistical contrasts and for any interaction contrast that included them. The matching position contrast compared trials in which X matched the first item (AAB, BBA) versus the third item (BAA, ABB), and the nativelikeness of X contrast compared trials in which X was more nativelike (AAB, BBA) versus less nativelike (ABB, BBA). The aggregated means relevant to these contrasts are presented in Table 2.

The phonological boundary contrast was significant, \( F(1, 38) = 7.31 \), \( M_{diff} = 4.01\% \), \( SE = 1.48\% \), 97.5% confidence interval (CI) = 0.55%-7.48%, such that TC contrasts were responded to more accurately than CG contrasts. The phonetic sensitivity contrast was also significant, \( F(1, 38) = 198.36 \), \( M_{diff} = 32.52\% \), \( SE = 2.31\% \), 97.5% CI = 27.13%-37.90%, such that CG contrasts were discriminated more accurately than SC contrasts. The ISI, matching position, and nativelikeness of X contrasts were not significant. There was a significant two-way interaction between phonological boundary and nativelikeness of X, \( F(1, 38) = 6.35 \), Contrast \( M = 1.66\% \), \( SE = 0.66\% \), 97.5% CI = 0.12%-3.21%, such that participants were slightly more accurate when X matched the more nativelike item than the less nativelike item in TC contrasts, whereas the opposite pattern was observed for CG contrasts. There was also a significant three-way interaction between ISI, matching position, and nativelikeness of X, \( F(1, 38) = 4.14 \), Contrast \( M = 2.37\% \), \( SE = 1.17\% \), 95% CI = 0.01%-4.73%. From Table 2 it can be seen that at 500 ms ISI responses are generally more accurate when X matches the first item than when it matches the third item, and more accurate when X is more nativelike rather than less nativelike. At 1500 ms, on the other hand, responses are generally more accurate when X matches the third item.
Results of the AXB discrimination task supported PAM predictions. With the new listener and language combination of English monolinguals and Arabic stimuli, TC assimilations were discriminated more accurately than CG assimilations, which were discriminated more accurately than SC assimilations. Furthermore, to our knowledge, this study is the first to show that participants are able to discriminate consonants in an AXB task under varying vowel conditions. Contrary to the predictions of the 3FM [2], however, we did not observe any interaction between ISI and assimilation type.

Several possible factors might account for the lack of ISI effect. First, task demands may be an issue. We used AXB discrimination for comparability with previous studies in support of PAM, but it is possible that the higher memory demand of AXB, relative to the AX task used by [2], may have mitigated against ISI effects in this study. It should be noted, however, that there was a small effect of ISI in this study, collapsed across all contrasts, such that memory demands appeared to affect performance only at 1500 ms ISI. Second, following a recent suggestion [8], the complexity of the AXB discrimination task may cause participants to rely more on phonological processing, and thus restrict access to any phonetic level of processing. This may have been further exacerbated by the varying vowels on AXB trials. Nevertheless, it should be noted that phonetic sensitivity is required to discriminate CG contrasts, and participants discriminated those well, even at an ISI of 1500 ms. It is possible that effects of ISI are only observed for SC assimilation types (such as the dental-retroflex Hindi contrasts used by [2]) and that the near-chance performance of our participants on those contrasts masked any influence of ISI. Future research could address these various factors by testing the same Arabic contrasts used here with an AX task design.

5. Acknowledgements

This study was supported by an Australian Research Council Discovery Grant to the first author (DP0880913). We thank Mark Antoniou and Angie Azar for their research assistance.

6. References