Title
Patterns of Misperception of Arabic Consonants

Journal
Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 41(41)

ISSN
2377-1666

Author
Sanker, Chelsea

Publication Date
2015

DOI
10.20354/B4414110017

Peer reviewed

PROCEEDINGS OF THE FORTY-FIRST ANNUAL MEETING OF THE

BERKELEY LINGUISTICS SOCIETY

February 7-8, 2015

General Session

Special Session
Fieldwork Methodology

Editors
Anna E. Jurgensen
Hannah Sande
Spencer Lamoureux
Kenny Baclawski
Alison Zerbe

Berkeley Linguistics Society
Berkeley, CA, USA

Berkeley Linguistics Society
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720-2650
USA

ISSN: 0363-2946

LCCN: 76-640143

Contents

Acknowledgments v
Foreword vii

The No Blur Principle Effects as an Emergent Property of Language Systems
Farrell Ackerman, Robert Malouf. 1

Intensification and sociolinguistic variation: a corpus study
Andrea Beltrama 15

Tagalog Sluicing Revisited
Lena Borise 31

Phonological Opacity in Pendau: a Local Constraint Conjunction Analysis
Yan Chen 49

Proximal Demonstratives in Predicate NPs
Ryan B. Doran, Gregory Ward 61

Syntax of generic null objects revisited
Vera Dvořák 71

Non-canonical Noun Incorporation in Bzhedug Adyghe
Ksenia Ershova 99

Perceptual distribution of merging phonemes
Valerie Freeman 121

Second Position and "Floating" Clitics in Wakhi
Zuzanna Fuchs. 133

Some causative alternations in K'iche', and a unified syntactic derivation
John Gluckman 155

The 'Whole' Story of Partitive Quantification
Kristen A. Greer 175

A Field Method to Describe Spontaneous Motion Events in Japanese
Miyuki Ishibashi 197

On the Derivation of Relative Clauses in Teotitlán del Valle Zapotec
Nick Kalivoda, Erik Zyman 219

Gradability and Mimetic Verbs in Japanese: A Frame-Semantic Account
Naoki Kiyama, Kimi Akita. 245

Exhaustivity, Predication and the Semantics of Movement
Peter Klecha, Martina Martinović. 267

Reevaluating the Diphthong Mergers in Japono-Ryukyuan
Tyler Lau 287

Pluractionality and the stative vs. eventive contrast in Ranmo
Jenny Lee 307

Labial Harmonic Shift in Kazakh: Mapping the Pathways and Motivations for Decay
Adam G. McCollum 329

Reference to situation content in Uyghur auxiliary 'bolmaq'
Andrew McKenzie, Gülnar Eziz, Travis Major 353

Case-Marking in Estonian Pseudopartitives
Mark Norris 371

Discourse Coherence and Relativization in Korean
Sang-Hee Park 397

Negotiating Lexical Uncertainty and Speaker Expertise with Disjunction
Christopher Potts, Roger Levy 417

Patterns of Misperception of Arabic Consonants
Chelsea Sanker. 447

The Imperative Split and the Origin of Switch-Reference Markers in Nungon
Hannah Sarvasy 473

Asymmetries in Long-Distance QR
Misako Tanaka. 493

The cross-linguistic distribution of sign language parameters
Rachael Tatman 503

Homophony and contrast neutralization in Southern Min tone sandhi circle
Tsz-Him Tsui 515

Cultural Transmission of Self-Concept from Parent to Child in Chinese American Families
Aya Williams, Stephen Chen, Qing Zhou 533

Fruits for Animals: Hunting Avoidance Speech Style in Murui
Katarzyna Izabela Wojtylak 545

A Quest for Linguistic Authenticity: Cantonese and Putonghua in Postcolonial Hong Kong
Andrew D. Wong. 563

Acknowledgments

As the Executive Committee of the 41st Annual Meeting of the Berkeley Linguistics Society, we would like to express our gratitude to the conference participants, volunteers, session chairs, faculty, and staff members for their participation. It was your contributions that made the conference a success. We are especially grateful for the patience, hard work, and support of Paula Floro and Belén Flores, without whom BLS 41 would not have been possible. We would also like to thank the following departments and organizations of the University of California, Berkeley for their generous financial support:

Department of Linguistics
Student Opportunity Fund
Graduate Assembly
Department of Philosophy
Department of Psychology
Department of Ethnic Studies
French Department
Department of Spanish and Portuguese
Berkeley Language Center
Office of Academic Affairs Vice President

Foreword

This monograph contains a number of the talks given at the 41st Annual Meeting of the Berkeley Linguistics Society, held in Berkeley, California, February 7-8, 2015. The conference included a General Session and the Special Session *Fieldwork Methodology*. The 41st Annual Meeting was planned and run by the second-year graduate students of the Department of Linguistics at the University of California, Berkeley: Kenny Baclawski, Anna Jurgensen, Spencer Lamoureux, Hannah Sande, and Alison Zerbe.

The original submissions of the papers in this volume were reviewed for style by Anna Jurgensen and Hannah Sande. Resubmitted papers were edited as necessary by Anna Jurgensen and Kenny Baclawski, and then compiled into the final monograph by Anna Jurgensen. The final monograph was reviewed by Spencer Lamoureux. The endeavor was supported by Alison Zerbe's management of the Berkeley Linguistic Society's funds for publications.

The BLS 41 Executive Committee
July 2015

Patterns of Misperception of Arabic Consonants

Chelsea Sanker
*Cornell University**

1 Introduction

There has been much investigation into perception of speech sounds, demonstrating a range of influences including listeners' native language (e.g. Cutler et al. 2004), the sounds' position in the syllable (e.g. Wang and Bilger 1973), and the presence of different types of masking noise (e.g. Phatak, Lovitt, and Allen 2008). However, there is no data on patterns of misperception of guttural consonants (uvulars, pharyngeals, and glottals); data on consonant discrimination in Arabic (e.g. Kishon-Rabin and Rosenhouse 2000) and the phonological characteristics of gutturals (cf. McCarthy 1994) suggest that they may pattern distinctly from consonants with other places of articulation. The study introduced here will investigate patterns in directionality of misperception of guttural consonants by presenting syllables containing guttural and non-guttural consonants in noise to be identified by native and non-native speakers of Arabic.

While there have been studies on how patterns of misperception of phonological structure differ depending on whether the structure is licit or not within the listener's native language (e.g. Dupoux et al. 1999), there have been few studies on the patterns of misperception that are present in identification of structures which are licit within a language. This study investigates such patterns and how they differ between guttural and non-guttural consonants.

1.1 Phonetic and Phonological Background

The guttural consonants, which are produced by constriction at the back of the mouth or in the throat, often exhibit shared characteristics, which are similar across languages with these sounds; e.g. causing neighboring vowels to be lowered or retracted. They also can share phonotactic restrictions that other consonants within a language do not have, such as not occurring within clusters or word-finally, or not co-occurring within a word. A number of features have been proposed to characterize the gutturals or certain sets of them, including advanced tongue root and retracted tongue root (Rose 1996; Shahin 2002), and +pharyngeal (McCarthy 1994; Esling 2005). However, the featural specifications used for gutturals do not present a clear picture of the guttural class, because the features shared under such systems often do not align with the physical gestures which these segments have in common.

During the past twenty years, there has been much progress towards clarifying the set of articulatory gestures involved in producing the guttural sounds in a number of languages, with data from laryngoscope (e.g. Esling, Fraser, and Harris 2005) and ultrasound (e.g. Zeroual, Esling, and Hoole 2011) expanding the information available from earlier X-ray

*I thank Michael Weiss, Sam Tilsen, Alan Nussbaum, and Abby Cohn for their insightful comments and advice on earlier versions of this paper and Munther Younes for his help in designing the experiment. I would also like to thank the audiences of the 41st annual meeting of the Berkeley Linguistics Society for their thought-provoking questions and suggestions.

work (e.g. Ghazeli 1977). Patterns of confusions can help elucidate the salient acoustic features shared by consonants (Chang, Plauché, and Ohala 2001), so will be of particular value for the guttural consonants.

1.2 Misperception Studies

There have been many perception experiments designed to test the stability and confusability of consonants in English (e.g. Miller and Nicely 1955; Bell et al. 1989), some in other Indo-European languages (e.g. Lecumberri and Cooke 2006), and very few in languages from other families (e.g. Singh and Black 1966; Bradlow et al. 2010). To my knowledge, there is no existing data on the misperception patterns of guttural consonants.

Misperception studies have looked at a range of influences. A number of studies have demonstrated that listeners distinguish between phonemes less accurately in phonological systems of languages which they do not speak natively (e.g. Mattys et al. 2010) and that non-native listeners are more strongly influenced by challenging listening conditions (e.g. Cutler et al. 2004). Other factors which can potentially influence results include the set of sounds given which listeners can choose from (e.g. Chang, Plauché, and Ohala 2001) and the phonological system which the sounds occur in (e.g. Shafiro et al. 2012).

The accuracy with which listeners identify segments is influenced by the type of masking noise which is present (e.g. van Engen and Bradlow 2007; Phatak, Lovitt, and Allen 2008) and the loudness of that noise relative to the speech stimuli (e.g. Miller and Nicely 1955). Misperceptions are based on the acoustic cues which remain apparent in the degraded signal (Soli and Arabie 1979), so the type of noise which most inhibits accurate identification varies by segment (e.g. Phatak, Lovitt, and Allen 2008; Dubno and Levitt 1981).

Consonant identification can also be influenced by the vowel context; however, the patterns of vowel effects are not consistent across studies, perhaps due to other differences in the phonetic environment, in the language being investigated, or the consonants being investigated. In English, Dubno and Levitt (1981) found that consonants are more accurately identified in the environment of /a/ than /i/ or /u/; Singh and Black (1966) found that consonants' place of articulation was identified less accurately before /i/ than before other vowels in English but not in other languages. Redford and Diehl (1999) found consonant identifications to be more accurate in cods after high vowels than after /a/, and Wang and Bilger (1973) found that coda consonants, though generally identified less accurately than onsets, were identified more accurately than onsets when the vowel was /i/.

Most perception studies use only single consonants, not clusters; this is probably in part because the number of possible responses gets exponentially larger when there are two consonants, which is prohibitive for the forced-choice design of many misperception experiments. Within the existing work on cluster perception, listeners are more accurate with identification of single consonants than clusters; most errors in clusters are due to deletions (Meyer, Dentel, and Meunier 2013). Much of the identifying information for the consonant not adjacent to the vowel is obscured in a cluster, based on the lack of formant transition cues from the vowel; these are the consonants which are most frequently neglected in complex syllables (Redford and Diehl 1999).

Much of the existing work on cluster misperception is based on perception of clusters which are illicit in the listeners' native language; such sequences are often poorly distinguished from sequences with a vowel dividing the consonants. For example, Berent et al. (2007) found that English speakers are less accurate in identifying number of syllables in monosyllabic words beginning with clusters that are with falling sonority than onset clusters with plateau sonority, which are themselves identified less accurately than onset clusters than rising sonority. Russian speakers, for whom these clusters are all acceptable, can identify onset clusters of all of these types with high accuracy. However, Russian speakers were slower to identify the more marked clusters. In a discrimination task, Dupoux et al. (1999) found that Japanese listeners had trouble distinguishing between VCCV and VCuCV sequences in a discrimination task; they also frequently reported vowels within VCCV sequences in an identification task, whereas French listeners distinguished between the presence or absence of a vowel in the same stimuli with high accuracy.

The previous literature does not provide data on patterns of misperception of guttural consonants or the interaction of place of articulation and noise type with accuracy of identification of syllable structure. The aim of the present study is to determine which cues listeners use for identifying guttural consonants by examining which other sounds these consonants are confused with and how noise of different frequencies influences identification, comparing these patterns in different phonetic environments and among native and non-native speakers of Arabic.

If the primary cues used for identification of guttural consonants are the transitions from neighboring vowels or the pseudoformants present within the consonants themselves, it is expected that the gutturals will be less accurately identified in low frequency noise than high frequency noise; if the primary cues for identification of gutturals are in the spectrum of fricative noise, it is expected that they will be less accurately identified in high frequency noise than in low frequency noise. Comparison of accuracy of consonant identification when adjacent to a vowel or to a consonant may indicate whether that consonant is identified by features within itself or more by transitions to adjacent vowels; however, some differences will likely be due to patterns of differences in production of these consonants in different environments.

The extent to which there is a difference in accuracy of identifications in each noise type furthermore indicates whether there are multiple cues at different frequencies which listeners can use; similar accuracy in both noise types would suggest that listeners are able to use multiple cue types and will focus on whichever cues are clearest in the signal. Non-native listeners are likely to have more difficulty identifying guttural consonants than other consonants, with a potentially larger effect of noise type, because most of these consonants are not present within the phonological system of English and thus these listeners may not have developed the same strategies that native listeners have for identifying these consonants.

2 Experiment Design

2.1 Participants

Participants were 9 native listeners of Arabic and 16 non-native listeners of Arabic, all reporting normal hearing. The ages of participants were 18–43 years (mean 24.2 years). Participants received monetary compensation for their participation.

Of the native Arabic listeners, six spoke a Levantine dialect, one a Maghrebi dialect, one a Gulf dialect, and one an Egyptian dialect; all of them had spent the majority of their childhood in a country where Arabic was the primary spoken language.

The non-native listeners of Arabic were all native speakers of American English who had been studying Arabic for 0.5–10 years. None of them had lived in any Arabic-speaking region for more than one year nor had parents who spoke Arabic natively. Fluency in Arabic varied across the participants, but all were comfortable reading Arabic orthography.

2.2 Materials

In the first task, stimuli were nonce words[1] with syllables shapes of CV[2] and ʔVC. In the second task, syllable shapes were ʔVC, ʔVRC, ʔVRV, and ʔVRəC. Consonants were produced in the context of each of the contrastive Arabic long vowels: /iː/, /aː/, /uː/. The stimuli were elicited using a list of words written in the Arabic orthography and were produced in isolation. Isolated productions were used because there is generally resyllabification in running speech in Arabic (Ryding 2005).

There were no vowel-initial stimuli used in this experiment, as Arabic words never begin with a vowel. Vowel-initial words resulting from borrowing or morphological processes end up beginning with a glottal stop; this non-lexical glottal stop has little influence on surrounding sounds (Ghowail 1987). The choice of [ə] in ʔVRəC sequences was because this vowel has a weaker effect on guttural formants than other vowels do and also because this is the lexical vowel that is closest to neutral realizations of Arabic inter-consonantal epenthetic vowels.

The consonants included were the seven guttural consonants of Arabic: /q/, /χ/, /ʁ/, /ħ/, /ʕ/, /h/, and /ʔ/; 3 "emphatic" consonants (oral consonants with secondary uvularization): /tˤ/, /dˤ/, and /sˤ/; and 10 oral consonants for comparison: /w/, /t/, /d/, /s/, /z/, /r/, /l/, /dʒ/, /j/, and /k/.

Stimuli were produced in Modern Standard Arabic (MSA) by two male native speakers of Arabic, recorded in a sound attenuated booth. Their dialects were Palestinian and Iraqi Arabic, which have very similar standard phonetic realizations of consonants, but different syllabification patterns. Their pronunciations in the stimuli were generally leveled towards features of Modern Standard Arabic, which is the form of Arabic that is typically used in reading aloud. The one salient idiosyncratic feature present was the Iraqi speaker's pronunciation of /dˤ/, which was noticeably fronted. Speakers of the same gender were chosen to avoid potential interactions of speaker gender and noise type, as noise weighted at different frequencies is likely to affect voices of different fundamental frequencies unequally.

Each stimulus was mixed with one of two types of noise, which began 100 ms before the stimulus and continued until 100 ms after the stimulus. The types of noise used were pink noise, in which the intensity is inversely proportional to the frequency, and blue noise, in which the intensity is proportional to the frequency. Because the perception of loudness depends on frequency, the amplitude of the noise was calculated to have perceived loudness approximately equal to the average perceived loudness of the stimuli.

Perceived loudness was calculated using the average of the values produced by the ANSI S3.4 and DIN 45631 methods of calculating loudness, performed in Matlab. Although there are differences between the results produced by each method, they have both been found to correlate with listeners' subjective evaluations of loudness (e.g. Schlittenlacher et al.

[1]Some combinations of segments happened to be real words. See Section 3.5 for a discussion.
[2]Here and in following discussions of structure, C = any consonant, V = any vowel, R = a liquid ([l] or [r]).

2011). Having normalized loudness in this way, the effects of each type of noise can be more meaningfully compared, as results should reflect differences in noise type and not differences in perceived loudness.

2.3 Procedure

Each listener completed two experimental tasks, followed by a demographic questionnaire; each task was divided into blocks. Instructions were given in English in order to not bias responses based on exposure to a particular dialect during the experiment, as many Arabic speakers show accommodation to features of dialects used by their interlocutors (Abu-Melhim 1995). The experiment was run in Matlab on a laptop computer with 13 inch screen and a wireless mouse, with stimuli played over headphones in a quiet room.

In both tasks, the listeners saw an array of buttons of nonce words written in the Arabic script, listened to stimuli, and clicked the button corresponding to the word which they thought they heard. The order of stimuli was randomized within blocks.

The presentation of stimuli was self-paced. The next stimulus would not play until after the listener had made a response. Button selection and response time, measured from the end of the stimulus, were recorded in Matlab. The experiment lasted 40-60 minutes, depending on the participant's speed of responding.

In the first task, there were 20 buttons, each with a different consonant. The order of consonants within the array remained constant, following the alphabetical order of the consonants. There were 40 stimuli in each block, with each consonant presented twice, once from each speaker. In this task, there was alternation between blocks of CV stimuli and blocks of ʔVC stimuli; the vowel cycled between blocks among each of the three contrastive Arabic vowel qualities: /iː/, /aː/, /uː/, producing six combinations of consonant position + vowel. This was done to minimize possible carry-over effects between blocks based on sharing shape or vowel with the previous block. Participants had one of six orders of blocks; in each of these orders, the pattern was the same except for where in the order the experiment began, in order to separate effects based on phonetic environment and effects based on when in the task a stimulus was presented.

In the second set of blocks (mixed-structure blocks), there were 42 buttons. These blocks included a mix of syllable structures, to test errors in identifying structure; options were monosyllabic with a single coda consonant (ʔVC), monosyllabic with a coda cluster beginning with a liquid (ʔVRC), disyllabic with an intervocalic liquid and no final consonant (ʔVRV), and disyllabic with an intervocalic liquid and a final coda consonant (ʔVRəC). Due to constraints of array size and time, a smaller set of consonants was used in these blocks: /w/, /t/, /d/, /s/, /r/, /l/, /j/, /k/, /q/, /χ/, /ʁ/, /ħ/, /ʕ/, /h/, and /ʔ/. The liquids (/l/ and /r/) and glides (/w/ and /j/) did not appear in biconsonantal forms.

To aid participants in finding buttons, each type of syllable shape was given a different border color. There were 80 stimuli in each block in this task. As in the monosyllable blocks, selection options were written in the Arabic script. All short vowels and absence of vowels was marked in the script, to avoid ambiguity between written forms. In this task, there was alternation between blocks in which biconsonantal forms had [l] and blocks in which biconsonantal forms had [r]; as in the first task, the vowel cycled between blocks among each of the three contrastive Arabic vowel qualities: /iː/, /aː/, /uː/.

3 Results

There were differences between the two groups in accuracy of identifications, in which sounds were most frequently confused, and in the directions of confusion, as well as effects of phonetic environment, the consonant's place of articulation, and the type of masking noise. The lower accuracy among non-native listeners was most extreme for the guttural consonants. In both groups, identifications were more accurate in the first task (consistent blocks of CV or ʔVC stimuli) than in the second task (blocks with a mix of syllable structures) and were less accurate in low frequency noise than high frequency noise. The difference in accuracy between noise types was greatest for guttural consonants.

There was a general trend of increase in accuracy over time, though the pattern of change in accuracy differed between the listener groups listener groups in the two tasks; this increase in accuracy perhaps indicates listener acclimatization to the dialectal and idiosyncratic features of the stimuli, as well as the presence of the masking noise.

3.1 Structure Confusions

The accuracy of identifications of structure was higher in the native listener group (84.6%) than in the non-native listener group (74%). Both listener groups made more errors that involved not perceiving segments that were present than errors that involved perceiving segments which were not present.

Table 1: Native Listeners' Structure Identifications (as percents)

stimulus ↓ / item selected →	ʔVC	ʔVRC	ʔVRV	ʔVRəC
ʔVC	**95**	3	1	1
ʔVRC	8	**82**	1	8
ʔVRV	5	9	**67**	19
ʔVRəC	0	14	8	**78**

Table 2: Non-native Listeners' Structure Identifications (as percents)

	ʔVC	ʔVRC	ʔVRV	ʔVRəC
ʔVC	**91**	6	1	3
ʔVRC	22	**61**	2	15
ʔVRV	3	9	**70**	18
ʔVRəC	2	20	12	**67**

Accuracy of structural identifications increased across blocks in both listener groups. Among native listeners, mean structural identification accuracy increased from 81.5% to 86.6% from the first to the last block of the mixed-structure task ($p = 0.0084$); among non-native listeners, structural accuracy increased from 71.8% to 75.2% ($p = 0.057$).

Response times were significantly faster for accurate responses than inaccurate responses. For non-native speakers, the mean response time in the second task for structurally accurate identifications was 4.3 seconds and for structurally inaccurate identifications was 5 seconds ($p < 0.0001$); for native speakers, the mean response time in the second task for structurally accurate identifications was 4.5 seconds and for structurally inaccurate identifications was 5.7 seconds ($p = 0.00064$).

Response times differed by type of error; inaccurate structural identifications that involved perceiving segments which were not present had a higher mean response time than identifications that involved not perceiving segments that were present. For native listeners, the mean response time was 7.2 s for errors of segment addition and 4.5 s for errors of segment loss ($p < 0.0001$); for non-native listeners, the mean response times were 6 s and 4.3 s, respectively ($p < 0.0001$).

In errors in which the listener failed to identify one of the consonants in the stimuli as being present, the consonant neglected was most frequently the second consonant. Among native listeners, 92% of the cluster simplification errors involved deleting the second consonant. The identity of the first consonant, [r] or [l], was always preserved. Among non-native listeners, 82% of cluster simplification errors involved deleting the second consonant. The identity of the first consonant, [r] or [l], was preserved in almost all cases. In 6% of non-native listeners' cluster simplification errors, the second consonant was preserved and the first consonant deleted. However, in 13% of cluster simplification errors made by non-native listeners, the consonant selected did not match either consonant present in the stimulus; in nearly half of these errors, the consonant selected was [ʁ]. It is unclear whether these cases are best characterized as neglect of one consonant and misperception of the other or an interaction between features of the liquid and the following consonant.

Response times varied by the type of structure as well as the set of options available within a block; responses were faster in blocks of consistent structure than in blocks of mixed structure. The slower response times in mixed-structure blocks could be in part due to the larger set of options to select from and in part due to the greater complexity of the stimuli in the number of segments present and the need to attend to structure in addition to consonant quality, whereas listeners did not need to consider structure within the consistent-structure blocks. Response times also differed among identifications for each type of structure present among the stimuli of the mixed-structure blocks, as presented in Table 3, which demonstrates that the response latency was influenced both by the structure of the stimulus and by the options present in the block.

Table 3: Response Times by Structure Type (in seconds)

Mixed-Structure Blocks

	Native	Non-native
ʔVC	4.5	4.2
ʔVRC	4.6	4.4
ʔVRV	6.9	5.4
ʔVRəC	4.6	4.6

Consistent-Structure Blocks

	Native	Non-native
CV	2.8	2.4
ʔVC	2.6	2.6

3.1.1 Effects of Phonetic Environment on Structural Accuracy

Some features of the phonetic environment influenced accuracy in structure identifications – in particular, which vowel was present in the word. There was no effect of liquid environment on accuracy of structure identifications in either listener group, but there was a significant effect of vowel environment, which was stronger among native listeners than non-native listeners.

The vowel whose quality is being considered here is the first vowel of the word; in sequences with a subsequent vowel between consonants, this vowel was always [ə]. Among native listeners, accuracy was much lower following [a:] (77.8%) than following [i:] (86.9%); $p < 0.0001$. Accuracy was also lower following [a:] than following [u:] (89.1%); $p < 0.0001$. The difference in accuracy following [i:] vs. [u:] was not significant ($p = 0.2$). Among non-native listeners, there was a significant difference between accuracy following [a:] (72.1%) and accuracy following [u:] (76.2%), $p = 0.022$. The accuracy following [i:] (73.6%) was not significantly different from either [a:] ($p = 0.42$) or [u:] ($p = 0.13$).

The difference in structure identification accuracy following high vowels and the low vowel may in part be due to coarticulatory effects of a later [ə] being attributed to the prior vowel, as the largest difference between these environments was greater frequency of errors neglecting inter-consonantal [ə]. There were also many more errors in stimuli starting with [a:] of neglecting word-final vowels, particularly [u]; most such errors were for stimuli with trilled [r], perhaps because the [r] itself is very short but strongly influences the following vowel; the mix of trilled and bunched *r* stimuli may have allowed this *r*-colored [u] to be interpreted as a bunched [r]. This effect was weaker with [i], perhaps due to its more distinct F2. Word-final [a] was never neglected, but was more prone to interpretation as a consonant.

3.1.2 Effect of Consonant Place on Structural Accuracy

There are different patterns of structure identification errors based on the region of the consonant's place of articulation; the confusion matrices for structure separated by place of articulation are presented in Tables 4-7. Perception errors for stimuli with guttural consonants involved proportionally more segment insertion errors than for stimuli with non-gutturals, particularly among native listeners, as well as a greater number of identifications of consonants as vowels and loss of these consonants following vowels.

Non-native listeners had lower accuracy for identifying structure than native listeners; the difference was larger for sequences with gutturals (69.8% and 84.6% respectively, $p < 0.0001$) than for sequences without gutturals (83.1% and 88% respectively, $p = 0.00067$), but both reached significance.

Non-guttural consonants were identified with significantly higher accuracy than guttural consonants in both listener groups, though the difference was much larger among non-native listeners. Native listeners identified structure with 88% accuracy for words without gutturals and 84.6% for words with gutturals (p value = 0.025), and non-native listeners identified them with 83.1% and 69.8% accuracy, respectively ($p < 0.0001$).

Table 4: Native Listeners' Guttural Structural Identifications

	ʔVC	ʔVRC	ʔVRəC	ʔVRV
ʔVC	**97**	2	1	1
ʔVRC	9	**80**	10	2
ʔVRəC	0	13	**77**	10

Table 5: Native Listeners' Non-Guttural Structural Identifications

	ʔVC	ʔVRC	ʔVRəC	ʔVRV
ʔVC	**97**	1	0	1
ʔVRC	8	**86**	5	1
ʔVRəC	0	18	**79**	4

Non-native listeners frequently did not perceive a final consonant following another consonant or perceived a vowel between the two consonants, particularly when the second consonant was guttural. The difference between the number of errors with guttural and non-

Table 6: Non-native Listeners' Guttural Structural Identifications

	ʔVC	ʔVRC	ʔVRəC	ʔVRV
ʔVC	**92**	6	2	0
ʔVRC	25	**54**	18	3
ʔVRəC	2	20	**63**	15

Table 7: Non-native Listeners' Non-Guttural Structural Identifications

	ʔVC	ʔVRC	ʔVRəC	ʔVRV
ʔVC	**95**	3	1	1
ʔVRC	17	**73**	11	0
ʔVRəC	3	20	**73**	4

guttural consonants, compared with a χ^2 test, is significant: for vowel insertion $p < 0.0001$, and for consonant deletion $p < 0.0001$.

For native listeners there was not a significant difference in the number of consonant deletion errors in sequences with gutturals and sequences without gutturals ($p = 0.45$). The difference in relative frequency of vowel deletion errors between sequences with gutturals and without gutturals approaches significance ($p = 0.058$).

3.2 Segment Confusions

The patterns of confusions among segments and the directions of these confusions indicate perceptual similarity and potentially also which acoustic cues are being used to make identifications. The largest number of confusions among non-native listeners involved errors in place of articulation, while confusions among native listeners were more evenly balanced among place, manner, and voicing. The pattern of confusions among native listeners varied more between the blocks with simple vowel environment and the blocks with mixed structures.

Tables 8-9 present the matrix of confusions between consonants present in the stimuli. For space reasons, only a selection of segments is included: the consonants which appear in all environments, all of the vowels, and /r/. Responses for ∅ include only cases in which options were available with and without a segment i.e. with ʔVR and ʔVRV forms in the mixed-structure blocks. Cells with a value of 0 have been marked with a dash, for ease of reading. Responses have been pooled across all phonetic contexts.

The results for confusions are influenced by the selection options; errors were restricted by what competing options with similar were present in the array and in the phonological inventory of Arabic. The consonants included in the experiment were chosen to include all of the guttural consonants and a selection of non-guttural consonants for comparison. Confusions for a larger number of consonants for English have been presented in a number of previous studies, e.g. Miller and Nicely 1955.

Based on results of previous work on native vs. non-native speakers' patterns of misperception, it is expected that native Arabic listeners would be more sensitive to distinctions between sounds in their language, and thus make identifications faster and with a smaller percent of errors than other listeners, while native English listeners who have studied Arabic as a second language will be less sensitive to the distinctions between Arabic sounds, making decisions more slowly, less accurately, and with a potentially different set of typical errors. It is accordingly also expected that the non-native listeners will have the highest accuracy for phonemes which are also present within English.

This hypothesis is borne out, as presented in Table 10; while non-native listeners perform with significantly lower accuracy than native listeners for many of the consonants, the

Table 8: Native Listeners' Segment Identifications

	a	i	u	t	d	s	r	k	q	χ	ʁ	ħ	ʕ	ʔ	h	∅	other
a	**69**	6	-	-	-	-	-	-	3	-	9	-	-	-	12	-	-
i	-	**83**	-	-	2	-	-	-	-	-	4	-	-	2	2	4	4
u	-	-	**85**	-	-	-	-	-	-	-	6	-	-	-	6	2	-
t	-	-	-	**68**	2	1	-	12	3	3	3	-	-	2	1	3	-
d	-	-	-	4	**78**	-	-	1	-	2	4	-	-	1	-	4	4
s	-	-	-	3	-	**85**	-	1	-	3	1	-	-	-	-	-	6
r	-	-	-	-	-	-	**82**	-	1	-	10	-	3	3	-	-	2
k	-	-	-	6	1	1	-	**66**	8	4	3	-	1	3	2	2	3
q	-	-	-	2	-	-	-	6	**71**	5	3	-	1	4	-	-	5
χ	-	-	-	1	1	1	-	-	5	**77**	3	2	-	-	4	2	2
ʁ	-	-	-	-	2	-	4	-	1	1	**79**	-	2	-	1	3	5
ħ	-	-	-	-	1	-	-	1	3	3	-	**74**	6	3	5	4	-
ʕ	1	-	-	-	-	-	-	1	4	2	1	1	**67**	16	3	4	-
ʔ	1	-	-	1	1	-	-	-	2	1	1	-	2	**77**	4	6	2
h	1	-	-	2	1	1	-	1	3	7	2	16	2	5	**52**	6	1
∅	-	-	-	-	2	-	-	1	1	1	1	-	1	2	6	**85**	-

Table 9: Non-native Listeners' Segment Identifications

	a	i	u	t	d	s	r	k	q	χ	ʁ	ħ	ʕ	ʔ	h	∅	other
a	**71**	-	2	-	-	2	-	-	-	-	6	3	5	9	2	2	-
i	1	**88**	-	-	-	1	-	-	1	-	1	-	2	-	-	-	6
u	-	-	**87**	1	-	-	-	-	-	-	3	-	-	-	4	3	3
t	-	-	-	**55**	1	2	-	10	4	5	1	2	1	6	1	4	8
d	-	-	-	4	**65**	-	1	1	-	1	4	-	-	2	1	8	13
s	-	-	-	4	1	**73**	-	-	1	1	1	-	-	1	-	1	15
r	-	-	-	-	-	-	**76**	-	1	1	12	-	-	2	1	-	5
k	-	-	-	8	2	-	-	**46**	18	9	1	3	1	5	1	4	2
q	-	-	-	3	1	-	-	14	**49**	9	7	2	1	5	-	2	3
χ	-	-	-	3	1	1	-	4	4	**62**	3	8	1	4	3	3	1
ʁ	-	-	1	-	3	-	7	-	1	3	**59**	2	3	3	2	5	9
ħ	1	-	-	3	1	-	1	4	4	8	2	**36**	3	8	16	8	4
ʕ	1	-	-	-	-	-	1	2	5	3	2	5	**35**	31	3	10	-
ʔ	1	-	-	3	-	-	-	2	2	1	2	5	8	**59**	4	12	1
h	1	-	-	3	1	-	-	2	2	7	3	24	3	16	**21**	11	4
∅	-	-	-	-	-	-	-	-	-	1	2	2	2	5	3	-	**85**

difference is very large for all of the Arabic consonants which are not present within English. There are also several other sounds that are present within English that were identified with significantly lower accuracy among non-native than native listeners, some of which likely show this difference because there is no sound in English that is acoustically similar, whereas Arabic has a phonological neighbor; e.g. although [s] exists in English, Arabic also has [sˤ], which shares many acoustic features with [s]. Other confusions may be due to attention to different cues; for example, in English the contrast between [t] and [d] in many environments is aspiration, while in the dialects of Arabic represented in the stimuli, the difference is consistently voicing.

Table 10: Native and Non-native Listeners' Segment Identification Accuracy Compared

	Native Accuracy	Non-native Accuracy	p-value
w	.91	.92	0.61
t	.68	.55	0.00026***
d	.78	.65	0.00072***
s	.85	.73	< 0.0001***
l	.61	.64	0.57
r	.82	.76	0.17
j	.87	.82	0.14
k	.66	.46	< 0.0001***
q	.71	.49	< 0.0001***
χ	.77	.62	< 0.0001***
ʁ	.79	.59	< 0.0001***
ħ	.74	.36	< 0.0001***
ʕ	.67	.35	< 0.0001***
ʔ	.77	.59	< 0.0001***
h	.52	.21	< 0.0001***

The lower accuracy of non-native listeners for identifying particular sounds suggests that they are less sensitive to the cues used to identify these sounds, such that the presence of noise inhibits their ability to distinguish them. However, at least some aspects of their strategy seems to be similar to native listeners, as they show some of the same patterns in influence of phonetic environment. Cues from the vowel are particularly important for identifying consonant place of articulation, though the formant transitions for guttural consonants are less familiar to native speakers of English.

Native listeners' segment identifications increased in accuracy over the consistent-structure blocks, primarily over the first three blocks, from 73.7% to 79.4%, though this did not reach significance ($p = 0.078$); they also increased in accuracy over the mixed-structure blocks, from 66.8% to 71.9% ($p = 0.034$). Non-native listeners' mean segment accuracy increased over the consistent-structure blocks, from 54.3% to 66.7% ($p < 0.0001$); however, non-native listeners' mean accuracy did not change over mixed-structure blocks. This may indicate that they reached a plateau of familiarizing with the speakers' idiosyncrasies and learning to compensate for the masking noise, or were unable to adapt their expectations to the differences in pronunciation in the new phonetic contexts of the second section of blocks.

3.2.1 Segment Identification Accuracy by Phonetic Environment

There was only a significant effect of the consonant's position within a syllable on accuracy during the first task among native listeners; the trend among non-native listeners followed the inverse pattern. Among native listeners, segment identification accuracy was higher in codas (81.2%) than in onsets (77.1%); the difference is significant ($p = 0.021$). Among non-native listeners, accuracy was higher in onsets (61.5%) than in codas (60.2%), but the difference is not significant ($p = 0.4$). Patterns of responses and accuracy in onsets and codas also varied by individual segment.

Consonants were more accurately identified in blocks with consistent structure than in blocks with mixed structure. Among native listeners, the mean segment identification accuracy in consistent-structure blocks was 79.1% and in mixed-structure blocks was 70% ($p < 0.0001$). For non-native listeners, the mean segment identification accuracy in consistent-structure blocks was 60.8% and in mixed-structure blocks was 51.1% ($p < 0.0001$).

However, this difference in accuracy was not consistently the case for all consonants. For example, [q] was identified more accurately in the mixed-environment contexts than in simple contexts, with 53.1% vs. 43.8% accuracy among non-native listeners ($p = 0.047$) and 78.6% vs. 60.7% accuracy among native listeners ($p = 0.0022$). The explanation for this may lie in production differences, as the speakers tended to produce a longer and louder release burst for this sound in clusters than in simple codas.

The vowel environment also influenced accuracy of identification of segments. For native listeners, the mean accuracy in simple syllables with [a:] was 81.6%, with [i:] was 80.6%, and with [u:] was 75.2%. The difference between [u:] and each of the other vowels was significant ($p = 0.0042$ and $p = 0.017$, respectively). For non-native listeners, the mean accuracy in simple syllables with [a:] was 63.2%, with [i:] was 62%, and with [u:] was 57%. As with native listeners, the difference between [u:] and each of the other vowels was significant ($p = 0.0025$ and $p = 0.015$, respectively).

However, in blocks of mixed structure, the effect of the vowels differed. For native listeners, the mean segment identification accuracy in complex structures with [a:] was 61.6%, with [i:] was 70%, and with [u:] was 75.2%. All of these differences reach significance: for [a:] vs. [i:], $p = 0.00071$; for [u:] vs. [i:], $p = 0.031$; for [a:] vs. [u:], $p < 0.0001$. Among non-native listeners, the mean accuracy in complex structures with [a:] was 45.2%, with [i:] was 53%, and with [u:] was 55.1%; for [a:] vs. [i], $p = 0.00011$; for [u:] vs. [i:], $p = 0.3$; for [a:] vs. [u:], $p < 0.0001$.

Not all consonants were equally influenced by the phonetic environment. [ʕ] exhibited the greatest difference in accuracy for native listeners between simple and complex environments, with 89.8% accuracy in simple blocks and 51.5% accuracy in mixed-structure blocks ($p < 0.0001$); [t] was identified with 88% and 55.3% accuracy, respectively ($p < 0.0001$). This large effect of environment for these segments may suggest that cues in the adjacent vowel are important for identifying these consonants. Stops in general showed the largest effect of environment on accuracy of identification. The manner of articulation of [ʕ] in Arabic is debated; it is most often described as a fricative, but has also been called a stop or a glide. Within these stimuli, it was most frequently realized as a voiceless stop, which is consistent with this effect of environment.

The segments for which non-native listeners showed the greatest difference in accuracy between the simple environments and the mixed-structure environments were [k] (60.3% and 36.5%, respectively, $p < 0.0001$) and [ʕ] (56.3% vs. 21.1%, respectively, $p < 0.0001$). The similar pattern in different environments between native and non-native listeners for this pharyngeal may indicate that the cues which non-native listeners are trying to identify are the same ones which the native listeners are attending to, even though non-native listeners do so with lower accuracy.

3.2.2 Effect of Consonant Place on Segment Identification Accuracy

The accuracy of segment identification differed by the region of the consonantal constriction. Non-native listeners had significantly lower accuracy in identifying guttural consonants than consonants with other places of articulation; in simple syllables, mean accuracy was 49.9% for gutturals and 74.1% for other consonants, $p < 0.0001$. For native listeners, this difference was also apparent in simple syllables: mean accuracy was 74.8% for gutturals and 84.5% for other consonants, $p < 0.0001$.

In mixed-structure blocks, the region of consonantal constriction influenced accuracy for non-native listeners but not for native listeners. For non-native listeners, mean accuracy for gutturals was 43.5% and for other consonants was 63.5% ($p < 0.0001$); for native listeners, mean accuracy for gutturals was 68.8% and for other consonants was 71% ($p = 0.28$).

Native listeners' lower accuracy identifying gutturals than other sounds in simple environments seems to be largely due to lower accuracy of identifications of other consonants in mixed-structure blocks, while guttural consonants are less strongly influenced by phonetic environment; perhaps the lack of difference in accuracy for guttural and non-guttural consonants in mixed-structure environments is because vowel transitions are important for many of the other consonants, many of which were stops, while most gutturals can be identified by their own features. Most of the gutturals were identified with similar accuracy in simple and complex environments, with the exception of [ʕ].

3.2.3 Accuracy by Feature

There were strong patterns in the relative accuracy of identification of different features across sounds. However, some of the sounds do not fall into clear categories due to the combinations of features present in them or the presence of several realizations of the phoneme in different environments; e.g. [ʔ] is sometimes realized a creaky vowel, and the pharyngeal transcribed as [ʕ] appears both as a voiceless stop and a voiced fricative.

Manner of articulation was maintained in consistent-structure blocks with 88.7% accuracy by native listeners and with 79.2% accuracy by non-native listeners. In mixed structure blocks, manner was maintained with 80.8% accuracy by native listeners and with 70.5% accuracy by non-native listeners.

Voicing was maintained in consistent-structure blocks with 95.5% accuracy by native listeners and with 89.7% accuracy by non-native listeners. In mixed-structure blocks, voicing was maintained with 84.7% accuracy by native listeners and with 77.9% accuracy by non-native listeners. Within the mixed-structure blocks, both groups of listeners made significantly more errors identifying voiced segments as voiceless than the converse: native listeners had 81.6% accuracy identifying voiced consonants as voiced and 89.4% accuracy identifying voiceless consonants as voiceless ($p < 0.0001$); non-native listeners had 71.1% accuracy identifying voicing of voiced consonants and 84.2% accuracy identifying voicing of voiceless consonants ($p < 0.0001$). This effect may be the result of attributing voicing to the preceding liquid. Within the simple structure blocks, there was no significant difference among non-native listeners (90% accuracy for voiced consonants, 89.4% accuracy for voiceless consonants, $p = 0.53$); among native listeners the difference was marginally significant (96.5% accuracy for voiced consonants, 94.6% accuracy for voiceless consonants, $p = 0.042$).

Place of articulation was maintained in consistent-structure blocks with 84% accuracy by native listeners and with 73.3% accuracy by non-native listeners. In mixed-structure blocks, place of articulation was maintained with 73.5% accuracy by native listeners and with 59% accuracy by non-native listeners. The pattern of place identifications for glottals is similar to that of other places of articulation, which suggests that at least within Arabic, glottals do indeed have a place feature, instead of being placeless, as has been suggested for glottals in some phonological systems (e.g. Rose 1996).

Different places of articulation were identified with different accuracy. Among non-native listeners, the place of articulation which was maintained with the lowest accuracy was pharyngeal (39.3%), with low accuracy also for the glottals (49.8%), uvulars (66.4%), and velars (45.8%). Velars and uvulars were most often confused with each other. Pharyngeals were most often confused with glottals (28.8%), followed by uvulars (12.2%). Glottals were most often confused with pharyngeals (20.2%).

Among native listeners, the lowest accuracy of place identification was for velars (65.7%), followed by glottals (69.5%); velars were most often confused with uvulars (14.7%) and dentals (10.9%), and glottals were most often confused with pharyngeals (10.1%) and uvulars (7.9%). In both listener groups, glottals were the consonants most frequently not perceived in clusters and after vowels.

3.2.4 Response Time

Response times showed a number of patterns in relation to overall accuracy, as well as in relation to accuracy by participant and accuracy by segment. Non-native listeners made identifications more rapidly than native listeners, with a mean response time of 3.5 s vs. 3.7 s. Non-native listeners' response times decreased over time, with starting latencies longer than those of native listeners; native listeners' response times changed less during the experiment.

In both listener groups, the mean response time for segmentally accurate responses was faster than the mean response time for inaccurate responses: in consistent structure blocks, native listeners' mean latency for accurate responses was 2.4 s and for inaccurate responses was 3.7 s, $p < 0.0001$ (in mixed structure blocks, means were 4.4 s and 5.4 s, respectively; $p < 0.0001$); non-native speakers' mean accurate response time was 2 s and the mean inaccurate response time was 3.3 s, $p < 0.0001$ (in mixed structure blocks, means were 4.2 s and 4.7 s, respectively; $p < 0.0001$).

Figures 1-2 present plots of accuracy vs. response time among native listeners and non-native listeners, respectively, by segment. Response times and proportion of accurate identifications were z-scored by participant, to compensate for large individual variation in raw speed and accuracy of identifications. These calculations include only data from consistent-structure blocks, as certain consonants did not appear in all environments.

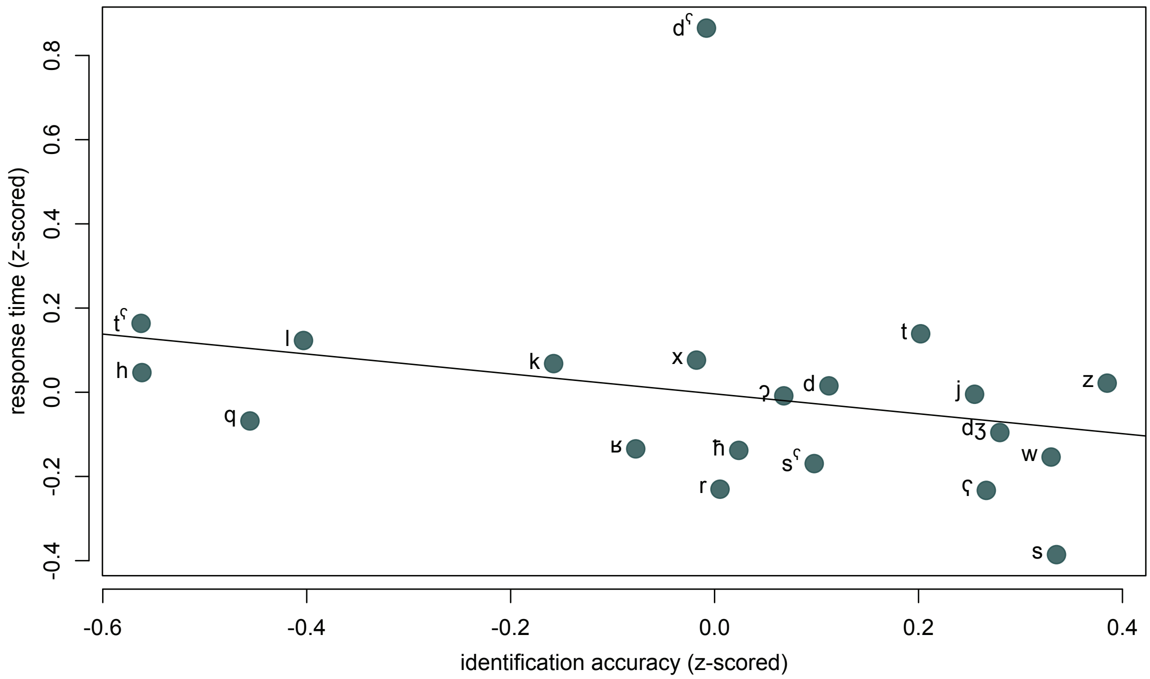

Figure 1: Native Listeners' Accuracy vs. Response Time, z-scored by Listener

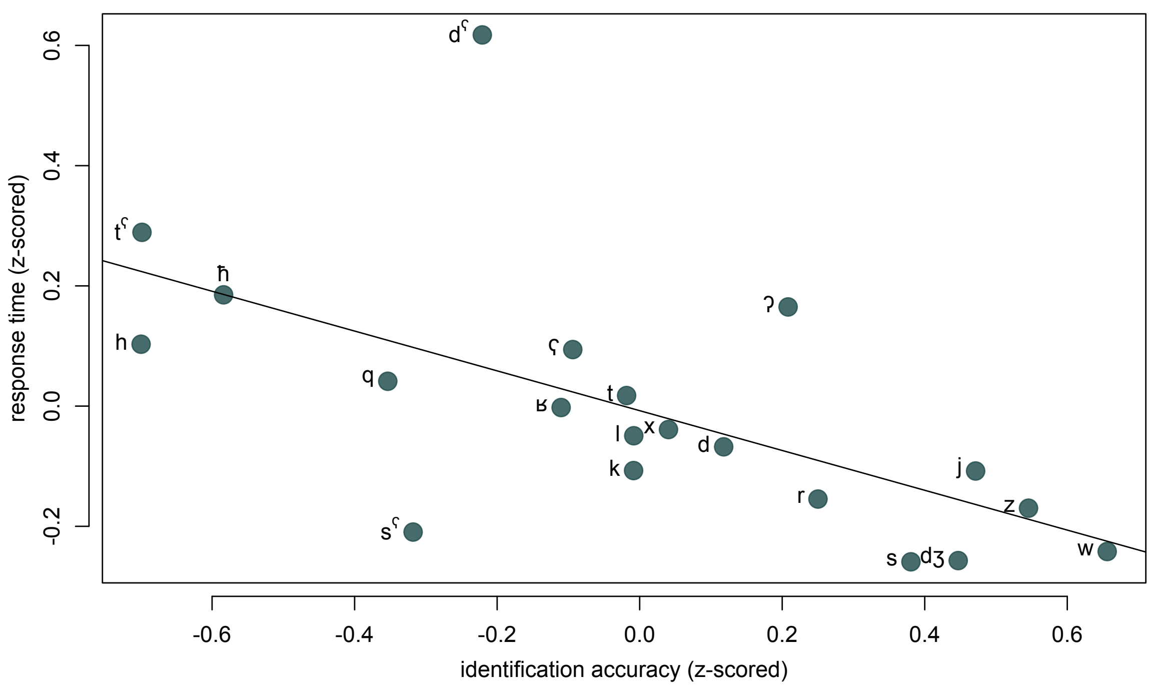

Figure 2: Non-native Listeners' Accuracy vs. Response Time, z-scored by Listener

The mean response times for segments had a strong inverse correlation with the mean accuracy of identification for each segment, pooled across all trials in consistent-structure blocks; among non-native listeners: R = -0.62 (p = 0.00046). This pattern of correlation was weaker among native listeners: R = -0.28 (p = 0.21).

For both listener groups, [dˤ] stands out as an outlier to the general pattern. This is the result of the Iraqi speaker's pronunciation, in which this sound was fronted, producing a segment which did not align with listeners' expectations, as there were no interdental or labial obstruents present among the selection options. The long response time seems to reflect listeners' lack of confidence in identifying this sound; several listeners asked about this sound after the task was over, wondering whether all of the sounds present in the stimuli were also provided as options in the button array. The moderate accuracy can be attributed to the same lack of strong competing selection options.

In the other direction, [sˤ] was identified relatively rapidly but with low accuracy among non-native listeners, which may suggest a low awareness among listeners of the confusability between this segment and [s], or at least little awareness of the presence of features in the stimulus that would make them wonder whether a token might be uvularized, such that they were confident in their identifications of this segment as [s] instead of deliberating about the choice. In contrast, trials in which [sˤ] was the selected response had a very long response latency. Uvularized consonants were relatively infrequent choices, within both listener groups but particularly in the non-native listener group.

Among native listeners, there was a positive correlation between each participant's mean response time and mean accuracy of identifications, R = 0.38; among non-native listeners, the correlation was negative: R = -0.34. However, neither of these correlations reached significance (p = 0.28 and p = 0.17, respectively).

3.2.5 Discriminability

In additions to the patterns of accuracy for each segment, there were patterns in the frequency with which each segment was selected as a response. Because some segments were selected as responses much more frequently than others, the percents of accurate responses are not all equally meaningful. The d' statistic was calculated for a sample of guttural and non-guttural sounds for native and non-native listeners, to evaluate the discriminability of these sounds, based on comparing the number of correct and incorrect identifications of each sound to the number of times stimuli without that sound were identified as having it or not; results are given in Table 11.

The discriminability of glottal and pharyngeal sounds was much lower among non-native listeners than among native listeners; some of these consonants were misidentified more frequently than they were correctly identified,[3] resulting in very low d' values.

Most non-guttural sounds showed greater similarity in discriminability among native and non-native listeners, though native listeners consistently had higher d' values across consonants, all even for consonants which are also present in English and which non-native listeners identified with high accuracy. Discriminability also depended on the number of other selection options which shared many features with a sound: the lower discriminability

[3] However, they were still significantly above chance levels of identification, which in a set of 20 selection options would be 5% accuracy.

of [k] compared to other oral obstruents among is largely due to the high number of confusions between [k] and [q].

Table 11: D' Scores for a Selection of Segment Confusions

	Native Listeners	Non-native Listeners
h	2.03	1.1
ʔ	2.64	1.74
ʕ	2.7	1.7
ħ	2.84	1.39
ʁ	2.67	2.07
χ	2.75	2.12
q	2.46	1.77
k	2.51	1.78
t	2.67	2.05
d	2.94	2.44
s	3.54	2.81
w	3.77	3.73

3.2.6 Patterns of Directionality

The patterns of consonant confusion among native listeners showed strong directionality for many consonants. Directionality is examined in Table 12, which gives the p-values produced by a χ^2 test for accurate identifications and confusions within each pair of sounds; the sounds included present a sample of confusions of guttural and non-guttural consonants. Where there were significantly more errors in one direction than the other for a pair of sounds, it was consistently the same direction for both native listeners and non-native listeners.

Table 12: P-values for Contingency Table Directionality, Selected Sound Pairs

	Native Listeners	Nonnative Listeners	directionality
h∼ħ	0.00097***	0.049*	h > ħ
ʕ ∼ʔ	0.69	< 0.0001***	ʔ < ʕ
χ∼ħ	0.5	0.48	none
χ∼h	0.47	0.049*	χ < h
ʁ∼ʕ	0.21	0.77	none
ʁ∼l	0.0003***	0.018*	ʁ < l
q∼tˤ	0.11	0.00018***	q < tˤ
t∼tˤ	0.023*	0.0059**	t < tˤ
s∼sˤ	0.015*	< 0.0001***	s < sˤ
t∼k	0.73	0.5	none
l∼d[4]	< 0.0001***	< 0.0001***	l > d

[4]Confusion between these sounds only occurs in syllable onsets.

One of the strong cases of directionality represented in several pairs of consonants was uvularized consonants identified as their non-uvularized counterparts. Other patterns of directionality were less consistent; between the pharyngeals and the glottals, the voiceless segments had a significant direction of glottals identified as pharyngeals in both listener groups, while the direction of the voiced segments, which was only significant among the non-native listeners, was of pharyngeals identified as glottals. None of the pairs of sounds have significantly different mean response times for each direction of confusion, which suggests that the errors in each direction are the result of the same decision process.

Several of these patterns go against the generally posited directionality of diachronic place of articulation change among guttural consonants, that the most likely pathways of change are from uvulars to pharyngeals, from pharyngeals to glottals, and glottals being lost (Kümmel 2007; Simpson 2002). However, this is not incompatible with the observed sound changes, as Simpson (2002) has observed that these characterizations of directionality hold most strongly when the sounds being produced by the change are not already present within a language's phonological inventory.

3.3 Effects of Individual Speaker and Listener

3.3.1 Effects of Speaker

The speaker had effects on accuracy of responses. As there were only two speakers represented in the stimuli, it is not clear whether differences between the speakers were based on dialect differences or idiosyncratic differences. The interaction between speaker and listener dialect for segment identification, with Levantine listeners performing with significantly higher accuracy on stimuli from the Palestinian (Levantine) speaker, may suggest that the differences were at least in part based on dialectal differences. However, non-Levantine listeners' greater accuracy in structure for stimuli produced by the Palestinian speaker suggests that features of this speaker's productions not related to dialect were important.

Native listeners as a group did not have significantly higher structure identification accuracy with stimuli from the Iraqi speaker (83.2% accuracy) or the Palestinian speaker (86% accuracy); the p-value for the difference was 0.076. There was also no significant interaction between listener dialect and speaker dialect. Native Levantine listeners identified structure in stimuli produced by the Palestinian speaker with only slightly higher accuracy (86.9%) than stimuli from the Iraqi speaker (85.5%), $p = 0.48$. Among non-Levantine listeners, the difference was greater but also not significant (84.9% vs. 80.4%, respectively; $p = 0.07$).

There was no significant difference in accuracy of structure identification among non-native listeners between stimuli from each speaker (73.1% accuracy and 75% accuracy, respectively, $p = 0.2$). All of the non-native listeners except one had learned the Levantine dialect, so there was not enough data to investigate potential interaction between speaker dialect and dialect studied on accuracy of identifications made by non-native listeners.

Native listeners performed with significantly higher segment identification accuracy with stimuli from the Palestinian speaker (76% accuracy) than the Iraqi speaker (72.1% accuracy); the p-value for the difference was 0.0034. This difference was largely due to significant interaction of speaker and listener dialect on segment identification: the mean segment accuracy of Levantine listeners with the Palestinian stimuli was 75.9% and with the Iraqi stimuli was 70.8% ($p = 0.0055$); the mean segment accuracy of non-Levantine listeners with the Palestinian stimuli was 83% and with the Iraqi stimuli was 78% ($p = 0.18$).

The non-native listeners also performed with higher segment identification accuracy with stimuli from the Palestinian speaker (57.7% accuracy) than the Iraqi speaker (54.1% accuracy); the p-value for the difference was 0.0022. Differences between the two speakers did not hold for all segments, however; for example, [ʔ] as produced by the Iraqi speaker was identified with significantly higher accuracy both among native listeners (82.9% accuracy vs. 71.2% accuracy, $p = 0.023$) and among non-native listeners (70.8% accuracy vs. 46.2% accuracy, $p < 0.0001$).

3.3.2 Effects of Listener

In mixed-structure blocks, both listener groups had a large range of individual participants' mean accuracy, though the range was larger among non-native listeners (36.1%–70%) than native listeners (53.6%–82.5%); in consistent-structure blocks, native listeners had a smaller range (67.7%–88.4%) than in the mixed-structure blocks, though non-native listeners had an equally large range (40.7%–76.2%). The correlation between mean segment identification accuracy by listener in simple contexts and mixed-structure contexts was 0.85 for non-native listeners and 0.65 for native listeners. The lower correlation among native listeners is likely because environment had a greater influence, while for non-native listeners, errors were more based on listeners' familiarity with the features of Arabic consonants.

The variation among listeners' structure identification accuracy was greater among non-native listeners than among native listeners; among the native listeners, mean accuracy for structure identification ranged from 77.3% to 94.9%, while for non-native listeners it ranged from 55.5% to 91.5%. The correlation of accuracy in structure identifications with segment identification accuracy from simple contexts, by listener, was 0.73 ($p < 0.0001$) and with segment identification accuracy in mixed-structure blocks was 0.84 ($p < 0.0001$) for non-native listeners; among native listeners, the correlation was 0.52 with accuracy from simple contexts ($p = 0.1$) and 0.77 with mixed-structure blocks ($p = 0.0016$).

An individual listener's accuracy could vary widely among different segments, but there was a positive correlation for sound identification accuracy across many pairs of sounds, with more consistently positive correlations among non-native listeners (mean correlation 0.27, $p < 0.0001$) than native listeners (mean correlation 0.19, $p < 0.0001$). This indicates that among non-native listeners, some listeners were more able to accurately identify Arabic consonants than other listeners, though there was also variation for individual sounds.

There were some noteworthy patterns of high correlations among the native listeners; all of the correlations between accuracy for uvularized sounds were positive and at least moderately high. There is also a cluster of positive correlations among the velars and uvulars.

Among the non-native listeners, almost all of the uvulars and pharyngeals had very high correlations, though the correlations between the accuracy of identifications of the glottals and accuracy of any of these sounds was lower and less consistent; accuracy of identifications of the glottal stop were not strongly correlated with accuracy of identifications of any other sound. There was also a strong correlation among accuracy of identifications of the uvularized consonants with each other and with the uvulars [q] and [ʁ], though not [χ], which may be due to the more forward position of this consonant relative to the other uvulars.

Listener dialect did not have any overall significant effect on segment identification accuracy; the mean segment identification accuracy of Levantine listeners was 73.4% and the accuracy of non-Levantine listeners was 74.7% ($p = 0.31$).

There was a significant difference in overall structure accuracy based on listener dialect; the mean structure identification accuracy of Levantine listeners was 86.2% and the accuracy of non-Levantine listeners was 82.6% ($p = 0.026$). Higher accuracy of structure identifications by Levantine listeners is consistent with the phonetic permissibility of consonant clusters within Levantine dialects, while many other dialects do not permit them. This trend suggests that native speakers of Arabic may be influenced by knowledge of phonotactic constraints on structure that exist within their dialect, even when listening to Modern Standard Arabic.

3.4 Effect of Noise Type

Within the experiment, each listener was either in a condition of pink noise (weighted towards lower frequencies) or blue noise (weighted towards higher frequencies). Both native and non-native listener groups had a higher mean accuracy of identifications in the higher frequency noise than in the lower frequency noise.

Structural identification among native listeners was 79.1% in low frequency noise and 88.9% in high frequency noise ($p < 0.0001$); among non-native listeners structural accuracy was 71.5% in low frequency noise and 76.5% in high frequency noise ($p = 0.00061$).

Both listener groups identified segments with better accuracy in higher frequency noise than lower frequency noise: among native listeners, accuracy was 59.8% in low frequency noise and 76.1% in high frequency noise ($p < 0.0001$); among non-native listeners, accuracy was 47.4% in low frequency noise and 54.8% in high frequency noise ($p < 0.0001$). This interacted with the region of the consonantal constriction.

Table 13: Accuracy by Noise Type and Consonant Place for Native Listeners (percent accurate)

	blue	pink
guttural	77	64
non-guttural	82	74

Table 14: Accuracy by Noise Type and Consonant Place for Non-native Listeners (percent accurate)

	blue	pink
guttural	51	41
non-guttural	70	68

Low frequency noise hindered identification more for gutturals than non-gutturals. Among native listeners, the difference in accuracy in high frequency vs. low frequency noise for gutturals is highly significant ($p < 0.0001$); the difference is smaller but still significant for non-gutturals ($p < 0.0001$). Among non-native listeners the difference in accuracy in high frequency vs. low frequency noise for gutturals is highly significant ($p < 0.0001$); the difference for non-gutturals was not significant ($p = 0.17$). The difference in accuracy between gutturals and non-gutturals was significant in both listener groups in both noise types.

3.5 Nonce Word vs. Real Word Stimuli

While most of the stimuli were nonce words, some of them are real lexical items within Arabic. Previous work has found that phonemes are identified more accurately in meaningful words than in nonce words (Boothroyd and Nittrouer 1988); differences have also been found in how native and non-native listeners are influenced by whether stimuli are real words; native listeners are more likely to make use of lexical information in identifying sounds in context (e.g. Mattys et al. 2010).

Among the native listeners, there was no significant difference in accuracy of identifications of real words (75.3%) than non-words (72.6%), $p = 0.16$. However, among the non-native listeners, there was a significantly higher accuracy of identification of real words (62.4%) than non-words (54.2%), $p < 0.0001$. This effect can largely be attributed to the higher relative frequency of real-word responses. Among non-native listeners the ratio of stimuli heard to responses clicked for real words was about 14:15, while the ratio for non-words was about 33:32. For the number of responses compared to the number of stimuli in each category, the χ^2 statistic $= 4.3$ ($p = 0.038$). Among the native listeners there was no significant difference between the relative number of responses in each category.

4 General Discussion

Patterns of confusions varied based on a number of factors; most errors were based on acoustic features, either intrinsic to the consonant or due to the obscuring effect of the masking noise or phonetic environment, while certain patterns of responses were due to language background and phonological knowledge.

4.1 Acoustics

Most common confusions were based on confusion of a single feature, sometimes two; among non-native listeners, the feature which was most commonly misidentified was place, particularly among consonants with a place of articulation not found in English.

The primary cues which listeners seem to be using for many of the consonants are the formant transitions in the neighboring vowels. These cues are obscured by environments in which there is no vowel adjacent to the consonant; in these environments many consonants were identified with much lower accuracy than in vowel-adjacent environments. The vowel cues are also partially obscured by low frequency noise, which is strongest at the same frequencies as the formants; this would explain why accuracy was lower for listeners who heard the sounds presented in low-frequency noise.

For guttural consonants, there are also pseudoformants present in the consonants themselves, which is likely why the phonetic environment does not have as strong an effect on accuracy of gutturals as for other consonants among native listeners. The low frequency noise obscures these cues, particularly F1, which is one of the clearest features associated with the identity of guttural consonants; this leaves only weaker cues for identification, such as duration and spectrum of the frication. However, native listeners are less influenced by noise type and phonetic environment than non-native listeners are, suggesting that they are better able to shift their attention to different cues for identification depending on which parts of the signal are obscured.

There are also differences in identification which result from acoustic differences present in the stimuli; part of the process of identifying stimuli is learning the cues being provided by individual speakers, who have subtle idiosyncrasies and sometimes very noticeable idiosyncratic features that can be used for identification. Native listeners more rapidly accommodate to the features of the speakers whom they are listening to, as shown by their rapid increase in identification accuracy over the first three blocks of the first task and again in the new environments of the second task; non-native listeners' accuracy of identifications also increased, but more slowly and only during the task with simple syllable structures, suggesting that they are less able to acclimate to the speech characteristics of individual speakers across phonetic environments.

4.2 Phonological Knowledge

Listeners have a bias towards perceiving sounds which are not present in their native phonological system as sounds which are. For example, voiced pharyngeal consonants are often heard as a low vowel by individuals who do not have pharyngeals in their native language (Bessell 1992); the vowel [ɑ] is identified as one of the outcomes of Arabic pharyngeals among bilingual speakers of Jaffa Palestinian Arabic and Modern Hebrew, perhaps also originating in misperception by late bilinguals (Horesh 2014).

Subconscious phonological knowledge from one's native language interacts with conscious knowledge of the phonological system of a second language. Thus, non-native listeners do not always make errors of identifying sounds as those which are present within their system if they are aware that there are additional sounds within a foreign system. However, the acoustic features which they have associated with an orthographic symbol may not be a reliable representation of the sound which that symbol represents for native listeners, which complicates interpretation of non-native listeners' responses in identification tasks. For example, for many non-native listeners, the symbol meant to represent [ħ] seems to be more salient than [h] as a voiceless fricatives with low peak frequency, but it is not certain that its phonetic representation for them is a pharyngeal.

We might expect that the accuracy of identifications of a sound will be related to its frequency within a language. Speech errors have been associated with phoneme frequency (e.g. Levitt and Healy 1985), but accuracy of responses also depends on the consonants present in the system, how acoustically similar they are, and thus how much competition there is for identification. While not frequently reported within misperception studies, the correlation between phoneme frequency and identification accuracy can be calculated from the reported results; in some studies there was a weak correlation (e.g. $R = 0.37$ in Cutler et al. 2004), and not in others (e.g. $R = 0.06$ in Miller and Nicely 1955).

Based on the number of distinct occurrences of each consonant in Greenberg's (1950) collection of 3775 verbal roots, the order by frequency of the Arabic consonants used in this experiment is: r w l ʕ j d ħ q s dʒ ʔ k h χ tˤ sˤ z t ʁ dˤ. The frequency of these segments has a positive correlation with accuracy of identifications among non-native listeners, which approaches significance ($R = 0.35$, $p = 0.11$), and a positive but small correlation with accuracy among native listeners ($R = 0.12$, $p = 0.61$). The stronger correlation among non-native listeners may be the result of greater practice with the more frequent consonants in a variety of environments, which may improve individuals' mental categories of these new sounds, whereas native listeners have clear categories for all of these sounds.

Non-native listeners also seem to be influenced by their knowledge of the lexicon, more frequently selecting response options which are real words; this may reflect a strategy used by non-native listeners in conversation also, using knowledge of the lexicon to guess the identity of words when they are uncertain about the phonetic form they have heard. This strategy may be particularly common among non-native speakers of a language with a phonological inventory very different than their native language, such as English speakers using Arabic.

5 Conclusion

Native listeners identified stimuli more accurately than non-native listeners both in structure and segment identity; this difference was strongest for the guttural consonants, but was also present for other consonants. There is much more variability in the identification accuracies among individual non-native listeners than among native listeners, probably reflecting the wide range of how established the Arabic phonological categories are for non-native speakers.

The higher accuracy of identifications in higher frequency noise than in lower frequency noise suggests that the pseudoformants of guttural consonants as well as the vowel transitions for both gutturals and non-gutturals are the most important cues used for identification, particularly of place. The similar effect of noise on both native and non-native listeners suggests that both groups are listening for the same main cues; the smaller effect of noise types and different phonetic environments on native listeners suggests that they have more flexibility in attending to different acoustic cues, based on which parts of the signal are accessible in different phonetic environments and noise contexts, while native listeners may be more limited in what cues they have learned to attend to.

Patterns of misperception provide data which may be useful in considering possible pathways of diachronic sound change. Directional asymmetries of misperception in experimental tasks parallel historical changes, because they are caused by some of the same processes (Garrett and Johnson 2013; Blevins and Garrett 2004). Data from misperception studies are not sufficient to make absolute conclusions about sound change, but may provide a useful line of additional evidence in reconstructing changes involving sounds for which there is little typological data.

6 References

Abu-Melhim, Abdel-Rahman. 1995. Code-Switching and Linguistic Accommodation in Arabic. In Mushira Eid, ed., *Perspectives on Arabic Linguistics VII*, pp. 231-250. Philadelphia: John Benjamins Publishing.

Alwan, Abeer. 1989. Perceptual Cues for Place of Articulation for the Voiced Pharyngeal and Uvular Consonants. *Journal of the Acoustical Society of America* 86(2): 549-556.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we Know about What we have Never Heard: Evidence from Perceptual Illusions. *Cognition* 104: 591-630.

Bessell, Nicola Jane. 1992. *Towards a Phonetic and Phonological Typology of Post-Velar Articulation.* Ph.D. dissertation, University of British Columbia.

Blevins, Juliette and Andrew Garrett. 2004. The Evolution of Metathesis. In Bruce Hayes, Robert Kirchner, and Donca Steriade, eds., *Phonetically Based Phonology*, pp. 117-156. Cambridge: Cambridge University Press.

Boothroyd, Arthur and Susan Nittrouer. 1988. Mathematical Treatment of Context Effects in Phoneme and Word Recognition. *Journal of the Acoustical Society of America* 84: 101-114.

Bradlow, Ann, Cynthia Clopper, Rajk Smiljanic, and Mary Ann Walter. 2010. A Perceptual Phonetic Similarity Space for Languages: Evidence from Five Native Language Listener Groups. *Speech Communication* 52: 930-942.

Chang, Steve, Madelain Plauché, and John Ohala. 2001. Markedness and Consonant Confusion Asymmetries. In Elizabeth Hume and Keith Johnson, eds., *The Role of Speech Perception in Phonology*, pp. 79-101. San Diego: Academic Press.

Cutler, Anne, Andrea Weber, and Nicole Cooper. 2004. Patterns of English Phoneme Confusions by Native and Non-Native Listeners. *Journal of the Acoustical Society of America* 116(6):3668-3678.

Dubno, Judy and Harry Levitt. 1981. Predicting Consonant Confusions from Acoustic Analysis. *Journal of the Acoustical Society of America* 69(1): 249-261.

Dupoux, Emmanuel, Erika Parlato, Sonia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. Where do Illusory Vowels Come From? *Journal of Memory and Language* 64: 199-210.

van Engen, Kristin and Ann Bradlow. 2007. Sentence Recognition in Native- and Foreign-Language Multi-Talker Background Noise. *Journal of the Acoustical Society of America* 121(1): 519-526.

Esling, John. 2005. There are no Back Vowels: The Laryngeal Articulator Model. *The Canadian Journal of Linguistics* 50: 13-44.

Esling, John, Katherine Fraser, and Jimmy Harris. 2005. Glottal Stop, Glottalized Resonants, and Pharyngeals: A Reinterpretation with Evidence from a Laryngoscopic Study of Nuuchahnulth (Nootka). *Journal of Phonetics* 33: 383-410.

Ghazeli, Salem. 1977. *Back Consonants and Backing Coarticulation in Arabic.* Ph.D. dissertation, University of Texas at Austin.

Ghowail, Thanaa. 1987. *The Acoustic Phonetic Study of the Two Pharyngeals /ḥ, ʕ/ and the Two Laryngeals /ʔ, h/ in Arabic.* Ph.D. dissertation, Indiana University.

Horesh, Uri. 2014. *Phonological Outcomes of Language Contact in the Palestinian Arabic Dialect of Jaffa.* Ph.D. dissertation, University of Essex.

Kishon-Rabin, Liat and Judith Rosenhouse. 2000. Speech Perception Test for Arabic-Speaking Children. *Audiology* 39(5): 269-277.

Kümmel, Martin. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion.* Wiesbaden: Reichert Verlag.

Lecumberri, M.L Garcia and Martin Cooke. 2006. Effect of Masker Type on Native and Non-Native Consonant Perception in Noise. *Journal of the Acoustical Society of America* 119(4): 2445-2454.

Levitt, Andrea and Alice Healy. 1985. The Role of Phoneme Frequency, Similarity, and Availability in the Experimental Elicitation of Speech Errors. *Journal of Memory and Language* 24: 717-733.

Mattys, Sven, Lucy Carroll, Carrie Li, and Sonia Chan. 2010. Effects of Energetic and Informational Masking on Speech Segmentation by Native and Non-Native Speakers. *Speech Communication* 52: 887-899.

McCarthy, John. 1994. The Phonetics and Phonology of Pharyngeals. In Patricia Keating, ed., *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, pp. 191-233. Cambridge: Cambridge University Press.

Meyer, Julien, Laure Dentel, and Fanny Meunier. 2013. Speech Recognition in Natural Background Noise. *PLOS ONE* 8(1): e79279.

Miller, George and Patricia Nicely. 1955. An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America* 27(2): 338-352.

Phatak, Sandeep and Jont B. Allen. 2007. Consonant and Vowel Confusions in Speech-Weighted Noise. *Journal of the Acoustical Society of America* 121(4): 2312-2326.

Redford, Melissa and Randy Diehl. 1999. The Relative Distinctiveness of Initial and Final Consonants in CVC Syllables. *Journal of the Acoustical Society of America* 106(3): 1555-1565.

Rose, Sharon. 1996. Variable Laryngeals and Vowel Lowering. *Phonology* 13(1): 73-117.

Ryding, Karin. 2005. *A Reference Grammar of Modern Standard Arabic*. New York: Cambridge University Press.

Schlittenlacher, Josef, Takeo Hashimoto, Hugo Fastl, Seiichiro Namba, Sonoko Kuwano, and Shigeko Hatano. 2011. Loudness of Pink Noise and Stationary Technical Sounds. *Inter-noise 2011, Osaka, Japan, September 4-7.*

Shafiro, Valeriy, Erika Levy, Reem Khamis-Dakwar, and Anatoliy Kharkhurin. 2012. Perceptual Confusions of American-English Vowels and Consonants by Native Arabic Bilinguals. *Language and Speech* 56(2): 145-161.

Shahin, Kimary. 2002. *Postvelar Harmony*. Philadelphia: John Benjamins Publishing Co.

Singh, Sadanand and John Black. 1966. Study of Twenty-Six Intervocalic Consonants as Spoken and Recognized by Four Language Groups. *Journal of the Acoustical Society of America* 39(2): 372-387.

Simpson, Andrew. 2003. Gutturals in Diachronic Perspective: The Case of Pharyngeal Merger and Loss in Semitic and Beyond. Ms. thesis, University of California, Berkeley.

Soli, Sigfrid and Phipps Arabie. 1979. Auditory versus Phonetic Accounts of Observed Confusions between Consonant Phonemes. *Journal of the Acoustical Society of America* 66(1): 46-59.

Wang, Marilyn and Robert Bilger. 1973. Consonant Confusions in Noise: A Study of Perceptual Features. *Journal of the Acoustical Society of America* 54(5): 1248-1266.

Zeroual, Chakir, John Esling, and Philip Hoole. 2011. EMA, Endoscopic, Ultrasound, and Acoustic Study of Two Secondary Articulations in Moroccan Arabic. In Zeki Majeed Hassan and Barry Heselwood, eds., *Instrumental Studies in Arabic Phonetics*, pp. 277-297. Philadelphia: John Benjamins Publishing.