811.111’342.2
https://doi.org/10.18485/bells.2022.14.1

Mahdi Duris*
Iowa State University
Ames, Iowa, USA

THE ACOUSTIC INTELLIGIBILITY OF CONSONANTS IN SAUDI SPOKEN ENGLISH**

Abstract
This study presents an instrumental phonetic account of the intelligibility of Saudi Spoken English (SSE) consonants. Few studies have investigated the spoken consonants of highly proficient EFL teachers in Saudi Arabia. This research informs on how intelligible SSE consonants are perceived by General American English (GAE) listeners using the Koffi (2021) intelligibility framework. Traditionally, intelligibility has been measured by having listeners transcribe speakers’ utterances. How well the speech is transcribed demonstrates a certain level of intelligibility. Koffi (2021) has proposed an acoustic approach to measuring consonant intelligibility using acoustic thresholds of Just Noticeable Differences (JND) combined with considerations for Relative Functional Load (RFL). An analysis of 23 segments spoken by 32 Saudi EFL teachers using acoustic correlates for intensity, duration, F2, and F3 inform the results. The quantitative results based on 1,280 tokens suggest that Saudi speakers of English are perceived as intelligible by GAE listeners when specifically analyzing their consonant production. Missing L1 segments [p] and [g], and substituting segments [f] for [v] does not impact intelligibility. Only the female participants did not distinguish their [ɹ] from [l]. Findings confirm that Saudi teachers of English can be intelligible in the segmental production of consonants.

* Email address: mduris@iastate.edu

** This paper was presented at the Sixth Belgrade International Meeting of English Phoneticians (BIMEP 2022), 25–26 March 2022, Faculty of Philology, University of Belgrade.

1. Introduction

With a paradigm shift in pronunciation principles with Levis' seminal paper in 2005, intelligibility has been widely accepted as a more desired goal than sounding native. This shift has also confirmed that pronunciation is a matter of both the speaker and the listener, removing the learning burden from speakers only. Intelligibility can be defined at the lexical level as the decoding of specific spoken words by listeners (Levis, 2018). Traditionally in second language (L2) research (Munro & Derwing, 1995; Kennedy & Trofimovich, 2008), this has been accomplished by having listeners transcribe speakers' utterances. How well the speech is transcribed demonstrates a certain level of intelligibility. This same focus on word-based features (Levis, 2018, p. 24) at the segmental level has been heavily studied, specifically for speakers of Japanese English with the /l/ and /ɹ/ pair. Intelligibility studies focused on L2 pronunciation at the segmental level have also been extensively studied in acoustic phonetics since 1999.

Intelligibility in acoustic phonetics draws directly from work in the late 1940s directed by physicists, engineers, and psychoacoustic researchers striving to make one human communication invention as intelligible as possible: the modern telephone. Most notable was the work conducted by Fletcher for Bell Laboratories between the 1920s and the 1960s, delivering scientific measurements for "auditory thresholds, intensity discrimination, frequency discrimination, tone-on-tone masking, tone-in-noise masking, the critical band, the phon scale of loudness, and the articulation index" (Yost, 2015, p. 49). Based on these discoveries, the field of phonetics changed from impressionistic assessment of sounds to methodical scientific measurements when the first spectrographs appeared in 1952 and then became widely available in the US in the 1980s. In 1995, another revolution in phonetics came to life with the release of Praat by Boersma and Weenink. Praat gave the power to measure acoustic sounds instrumentally by simply using a desktop computer.

Since the paradigm shift towards intelligibility in L2 research, only a few studies have used the word-based decoding abilities of acoustic phonetics to inform about L2 speech intelligibility. Such studies all relate to the framework developed by Koffi (2021), who developed a quantitative method of analysis to measure the intelligibility of L2 pronunciation both at the segmental and suprasegmental levels. Specifically, since 2012, Koffi has measured over 12,000 speech tokens from 67 non-native speakers to derive the first acoustic phonetic approach to measuring intelligibility. He does so by combining the works of Fletcher (1940, 1953) in using the Critical Band Theory (CBT), Just Noticeable Differences (JNDs) for relevant acoustic correlates classified by natural class and Catford's (1987) Relative Functional Load (RFL).

The present study focuses on a specific context that has not yet received much attention in academic research, mainly the acoustic phonetic characteristics of Saudi Spoken English (SSE) by Saudi EFL teachers. Previous word-based feature research (Duris, 2021) focused on vowel intelligibility and showed Saudi English teachers having highly intelligible vowels. This study will continue exploring the intelligibility of these teachers by focusing on their consonants. This continuation is important to grow a solid body of findings to inform about the complete intelligibility of Saudi Spoken English. Furthermore, this study and future ones from the author intend to build a complete acoustic phonetic profile of SSE for both segmental and suprasegmentals. Responding to Zielinski's (2015) call, this paper intends to inform further on "features of pronunciation [segmental vs. suprasegmental] as part of an integrated and interactive system, where the production of one can influence the other" (p. 402). Additionally, this study intends to help inform L2 teachers on how to approach intelligibility, specifically when dealing with pronunciation pedagogy in a Saudi EFL context.

2. Literature Review

As this article is set to inform the complete acoustic correlates of consonants in Saudi Spoken English, distinguishing the characteristics of General American English (GAE) and Modern Standard Arabic (MSA) consonants is essential. GAE and MSA share most sounds that have the same natural class. However, some distinctions may contribute to intelligibility barriers. The first part of this literature review will briefly account for the consonant charts for both GAE and MSA and provide both languages' major consonant characteristic features. This will help in pinpointing phonetic similarities and differences between the two languages. The second part of this chapter will detail the framework developed by Koffi (2021), which will be used in this study to expand on the acoustic correlates of consonants for SSE. It will also detail past studies that have used this framework and how findings from this research will inform the speech intelligibility of Saudi Spoken English consonants.

2.1 Consonant Characteristics of GAE and MSA

This first part of the chapter focuses on GAE and MSA consonant characteristics. As highlighted previously, understanding the similarities and differences between the L1 (Modern Standard Arabic) and the target L2 (GAE) helps understand which speech features may interfere with intelligibility. Graph 1 below presents the GAE chart of consonants for all three features discussed.

			Place of Articulation (POA)						
			Bilabial	Labiodental	Interdental	Alveolar	Palatal	Velar	Glottal
Manner of Articulation (MOA)	Stop	– voice	p <pat>			t <tap>		k<cat>	
		+ voice	b <bat>			d <dapper>		g <gap>	
	Fricative	– voice		f <fat>	θ <thatch>	s <sat>	ʃ <sharp>		h <hat>
		+ voice		v <vat>	ð <that>	z <zap>	ʒ <genre>[3]		
	Affricate	– voice					tʃ <chap>		
		+ voice					dʒ <jab>		
	Nasal	+ voice	m <mat>			n <nap>		ŋ <fang>	
	Liquid	+ voice				l <lap>	ɹ <rat>		
	Glide	+ voice					j <yap>	w <what>	

Graph 1: GAE Consonant Chart (Koffi, 2021)

GAE includes a total of 24 consonants. For this study, the following consonants (boxed in red) are not explored: the semi-vowel [j], the palatal fricative [ʒ], and the two palatal affricates [tʃ] and [dʒ], which are not present in the corpus used for this study to measure SSE consonants. To compare these target consonants of L2 for Saudi learners, the Modern Standard Arabic consonants from Alghamdi (2015) are presented in Graph 2.

	Bilabial شفتاني	Labiodental شفوي أسناني	Interental بين أسناني	Alveodental لثوي أسناني	Alveopalatal غاري لثوي	Palatal غاري	Velar طبقي	Lab-velar شفوي طبقي	Uvular لهوي	Pharyngeal حلقي	Glottal حنجري
Nasal أنفي	m م			n ن							
Stop شديد	b ب			t ت d د			k ك		q ق		ʔ ء
Emphatic Stop*				tˤط dˤض							
Fricative رخو		f ف	θ ث ð ذ	s س z ز	ʃ ش				χ خ ʁ غ	ħ ح ʕ ع	h هـ
Emphatic fricative**			ðˤ ظ	sˤص							
Affricate مزجي					dʒ ج						
Glide ليني						j ي		w و			
Lateral جانبي				l ل							
Trill تكراري				r ر							

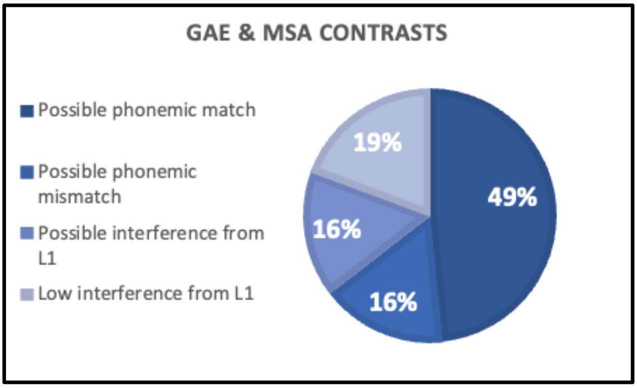
Graph 2: Modern Standard Arabic Consonant Chart
(Alghamdi, 2015)

Modern Standard Arabic includes 28 consonants (Ibrahim et al., 2020) with additional manner and place features not encountered in GAE. To better assess the similarities and differences between the target L2 (GAE) and the L1 (MSA), consonants from both languages are combined into a GAE and MSA consonant chart. Changes are made to accommodate both languages, as seen in Graph 3 below.

Manner of Articulation (MOA)		Place of Articulation (POA)																	
		Bilabial		Labiodental		Interdental		Alveolar		Palatal		Velar		Uvular		Pharyngeal		Glottal	
		GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA	GAE	MSA
Stop	[- voice]	p <pat>	–					t <tap>	t <ta>			k <cat>	k <kaf>	--	q <qaf>			--	ʔ <hamzah>
	[+ voice]	b <bat>	b <ba>					d <dapper>	d <dal>			g <gap>	–						
Fricative	[- voice]			f <fat>	f <fa>	θ <thatch>	θ <tha>	s <sat>	s <seen>	ʃ <sharp>	ʃ <sheen>			--	x <kha>	--	ħ <ha>	h <hat>	h <ha>
	[+ voice]			v <vat>	–	ð <that>	ð <dhal>	z <zap>	z <zay>					--	ɣ <ghayn>	--	ʕ <'ayn>		
Nasal	[- voice]																		
	[+ voice]	m <mat>	m <meem>					n <nap>	n <noon>			ŋ <fang>	–						
Liquid	[- voice]																		
	[+ voice]							l <lap>	l <lam>	ɹ <rat>	–								
	[+ voice]							--	r <ra>										
Glide	[- voice]											w <what>	w <waow>						
	[+ voice]																		
Emphatic Stop	[- voice]							--	tˤ <ta>										
	[+ voice]							--	dˤ <dad>										
Emphatic Fricative	[- voice]							--	sˤ <sad>										
	[+ voice]					--	ðˤ <dha>												

Graph 3: GAE and MSA Consonant Comparison Chart

Consonants highlighted in red in Graph 3 show a possible phonemic mismatch due to the sound being only available in the target language (L2). Inversely, all consonants highlighted in green show a possible match as the sound is available in both GAE and MSA. A third category highlighted for comparison are consonants in yellow. These consonants occur only in MSA; however, they have a similar counterpart in the place of articulation that occurs in GAE and MSA. Lastly, consonants highlighted in blue only occur in MSA and have no equivalency in GAE in one of the nodes. The possible phonemic match represents 49% of the shared consonants. Only 16% of consonants may cause a possible mismatch due to no equivalence in the L1. However, some possible interference from consonants may occur at the same level as the low interference may occur at 16%, as shown in Graph 4.



Graph 4: GAE and MSA Phonemic Contrasts

As mentioned in the introduction, a larger number of studies have used the Koffi (2021) framework to assess the intelligibility of vowels, while few of them have expanded into consonants. This study intends to fill this gap while expanding on existing data for the intelligibility of Saudi Spoken English vowels. This study will also respond to the methodological consideration set out by Flege (1987) to consider "intergroup differences" (p. 288). Flege (1987) recommends gathering measurements from groups of at least 6 to 12 participants to "observe a systematic phonetic difference between groups" (p. 288). This study provides such a number for both female and male participants. The following section will detail the Koffi (2021) framework for intelligibility.

2.2 Koffi's (2021) Framework for Intelligibility

This research study uses the Koffi (2021) framework to assess intelligibility for L2 Englishes. This framework is the central tenet in Koffi's (2021) book titled *Relevant Acoustic Phonetics of L2 Englishes.* The book offers the first literature that combines the seminal works of acousticians like Fletcher, modern phoneticians like Ladefoged, and intelligibility linguists like Levis to the Second Language Acquisition field. For over nine years, Koffi analyzed 12,000 tokens of L2 Englishes and developed the first acoustic phonetic driven intelligibility theory. The Acoustic and Masking Intelligibility (AMI) condition posits that:

> "Segments that are acoustically close may mask each other with only a minimal risk to intelligibility unless their relative functional loads dictate otherwise" (Koffi, 2021, p. 55).

The first part of this section will describe the Koffi (2021) framework in detail, followed by a descriptive account of how it will be used to determine the intelligibility of SSE consonants.

The Koffi framework (2021, pp. 37–55) gathers the following principles: correlate hierarchy, Critical Band Theory (CBT), Just Noticeable Differences (JNDs), Acoustic Masking, and Relative Functional Load (RFL) to give a complete assessment of L2 English intelligibility. The correlate hierarchy (pp. 37–40) expands on how acoustic correlates for formants, duration, and intensity, used to show relevance for specific segments, need to be weighted to assess intelligibility. In other words, "some correlates are more relevant than others for specific classes of sounds" (p. 37). Harvey Fletcher, also known as the father of stereophonic sound, developed the Critical Band theory as part of his findings in 1953. He identifies the frequency bands commonly known as formants (F0, F1, F2, F3) in acoustic phonetics. CBT is used in the framework to pinpoint the limits to which each formant can be perceived. The third principle in the Koffi framework for intelligibility relates the Just Noticeable Differences. JNDs are also known as acoustic thresholds, which Koffi (2021) describes as thresholds "which segments, and natural classes of segments are optimally perceived" (p. 46). Vast research has been conducted around JNDs for every formant in English and other languages, which Koffi describes in detail (2021, p. 40). JNDs for F1 and F2 have been researched by Scharf (1961), Mermelstein (1978), Hawks (1994), and Labov et al. (2006). For F3 and F4, JNDs have been validated by Scharf (1961). For VOT, the findings for JNDs have been obtained by Lisker and Abramson (1964), Byrd and Toben (2010), Fant (1960), Hirsh (1959), Abel (1972), Miller (1981), Phillips et al. (1994), Stevens (2000), Lehiste (1970) and Quené (2004), and for intensity, the JND is confirmed by Hansen (2001). Combined with the CBT, it is now possible to relate the acoustic correlates with specific formants with specific thresholds for human speech perception. Graph 5 summarizes these principles.

No.	Segments/Suprasegments	Acoustic Correlates	JND Thresholds
	Vowels		
1.	Vowels	F1	> 60 Hz
2.	Vowels	F2	≥ 200 Hz
3.	Vowels	F3	≥ 400 Hz
	Consonants		
1.	Stops	Voice Onset Time (VOT)	≥ 26, 34, 42 ms
2.	Fricatives and affricates	Intensity	≥ 3 dB
3.	Nasals	F2 for [m] and [n]	≥ 200 Hz
4.	Nasals	F3 for [n] and [ŋ]	≥ 400 Hz
5.	Approximants	F3	≥ 400 Hz
6.	Voicing ratios	Length in milliseconds	40/60
	Suprasegmentals		
1.	Stress	F0/Pitch	≥ 1 Hz
2.	Intensity	Intensity	≥ 3 dB
3.	Duration	Signals less than 200 ms	≥ 10 ms
4.	Duration	Signals between 200–300 ms	≥ 20 ms
5.	Duration	Signals over 300 ms	≥ 40 ms

Graph 5: Correlates and JNDs for Segmentals and Suprasegmentals (Koffi, 2021, p. 38)

The fourth element is acoustic masking which Fletcher developed in 1953. Masking refers to the sounds that can impose their features on other sounds for speech intelligibility. As Koffi (2021) describes it, “if two sounds overlap significantly in relevant acoustic properties, then masking can cause intelligibility to be compromised” (p. 47). The final principle of the framework is the Relative Functional Load. RFL is specifically suited for intelligibility because it prioritizes the sounds that have the biggest impact on frequency in the language. Levis (2018) contends that "more frequent phonemes are likely to play a larger role in intelligibility than less frequent ones" (p. 82). Koffi's framework explanation credits Catford (1987) for compiling the RFL values, which were then organized for consonants and vowels by Koffi (2021), available in Appendix A and B. Based on errors and the RFL, Koffi makes a direct relationship whereas "if a person substitutes Segment A for Segment B, the degree of intelligibility is directly proportional with their RFL" (2021, p. 48). This is seen in Graph 6, which gives a final assessment of intelligibility after considering correlate hierarchy, CBT, JNDs, and Acoustic Masking.

No.	RFL	Intelligibility Rating
1.	0–24%	Good intelligibility
2.	25–49%	Fair intelligibility
3.	50–74%	Mediocre intelligibility
4.	75–100%	Poor intelligibility

Graph 6: Correlations between RFL and Intelligibility
(Koffi, 2021, p. 50)

In sum, the Koffi framework gathers many principles in various fields dealing with speech production and perception to provide a robust and quantitative method to determine the intelligibility of L2 speech by way of acoustic phonetic analysis. The following section will detail how this framework has been used to determine the intelligibility of consonants for L2 Englishes. Given the theoretical thresholds of consonants seen in Graph 5, consonants for SSE will be analyzed by their stops (VOT), fricatives, nasals, and approximants. For each acoustic correlate, the specific methodology used in previous studies will also be the guiding model for this study. As such, the research questions for this study are:

1. Can SSE consonants produced by Saudi EFL teachers be perceived intelligibly by GAE listeners considering the lack of some consonant segments in MSA? (RQ1)

2. If intelligibility is compromised, which segments are used to compensate for this? (RQ2)
   Do these substitutions also affect intelligibility when used? (RQ3)

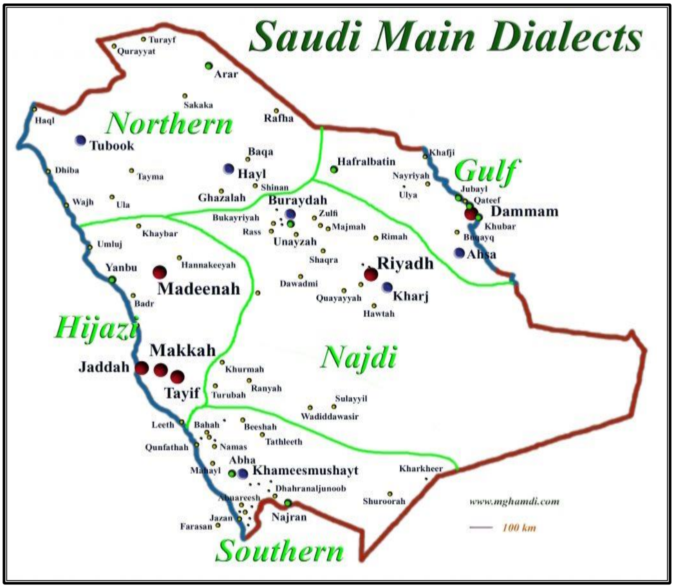
3. Methodology

3.1 Participants

For this empirical study, 32 Saudi adults were recruited (23 females and 9 males). As in most acoustic phonetic analyses of a homogenous group, the participants are separated into their biological differences. Kent & Read (2002, p. 194) confirm that male vocal tracts are generally longer than female ones, leading to lower formant frequencies.

The mean age for the female Saudi group is 32 years old, and all of them were born in Saudi Arabia except for one born in Canada. All females reside in Riyadh and are EFL teachers at the world's largest female-only university. Most of these female participants (69%) ascribe their Modern Standard Arabic dialect to Najdi, while the second shared dialect is Hijazi (17%). More than half of this group (56%) self-reported having lived outside Saudi Arabia, in an Inner Circle country. A majority shared that they attribute "entertainment" (English spoken in movies, TV shows, and the internet) as a major contributing factor to their L2 fluency. A complete sociometric report with linguistic data is available in Appendix C.

A total of 9 Saudi males also participated in this study. As shown in Appendix D, this group's average age is 34 years old, they are EFL instructors in higher education institutions, and most of them speak a Najdi dialect of MSA. Graph 7 shows a dialect map of Saudi Arabia highlighting the five dialects used in the Kingdom of Saudi Arabia (KSA).



Graph 7: Dialects of Modern Standard Arabic in KSA[1]

3.2 Data Collection Instruments

For all participants, the acoustic data was captured using a SONY ICD-UX560F (2018-12) voice recorder with stereo samples formatted in MP3 (sample rate of 44.1 kHz). A fixed microphone integrated into a headphone was used to keep an equal distance between the mouth and the microphone. The microphone is a Cardioid (unidirectional) with a calibrated frequency response between 50 and 20,000 Hz. The audio samples were converted from MP3 format down to a WAV mono file, keeping the sampling rate at 44.1 KHz. The data captured was analyzed using the computer program Praat (Boersma & Weenink, 2022) to measure the acoustic correlates of intensity, duration, F2, and F3.

1 Provided by Dr. Mansour Alghamdi on March 28th, 2020 in Riyadh, KSA

3.3 Procedures

To extract the data for this study, the elicitation paragraph used is similar to the one used for previous consonant studies mentioned in the introduction (Zhang, 2014; Koffi, 2015; Koffi & Ribeiro, 2016; and Koffi, 2019). The paragraph is from George Mason University's, Speech Accent Archive text (Weinberger, 2015). Weinberger (2015) explains that most of the segment sounds of English are present in this text. Koffi (2021) describes that this elicitation text contains all common English sounds "except the vowel [ʊ], the semi-vowel [j], the palatal fricative [ʒ], and the palatal affricate [dʒ]" (p. 66):

> ***Please call Stella****. Ask her to* ***bring these*** *things with* ***her*** *from the store:* ***Six*** *good spoons of fresh snow* ***peas****,* ***five thick*** *slabs of blue cheese, and maybe a foot-long sandwich as a snack for her brother Bob. We also* ***need*** *a small* ***plastic*** *snake, the little yellow book, a rubber duck, and a paper Ipad.* ***She*** *should not forget the dog video game and the* ***big toy*** *frog for the* ***kids****. She must leave the faked gun at home, but she may bring the ten sea turtles, the mat that my mom bought, and the black rug. She can scoop these things into three* ***red*** *bags and two old backpacks.* ***We*** *will* ***go meet*** *her, Sue, Jake, and Jenny,* ***Wednesday*** *at the very last train station.*

Participants were asked to read the above text to capture speech sounds as they would be uttered in continuous form. Duris (2021) mentions that this approach gives two benefits: context and frequency. The participant is unaware of which words are to be analyzed and focuses on reading words with clear and familiar context. Secondly, the elicitation paragraph uses high frequency words that most L2 learners use. One drawback of using connected speech instead of words in isolation for a segmental analysis of consonants is the lack of word boundaries. In their 2015 research, Alameen and Levis define "connected speech," citing Lass (1984) as "the processes that words undergo when their border sounds are blended with neighboring sounds" (p. 160). These processes are so common that they proposed six categories (linking, deletion, insertion, modification, reduction, and multiple) based on the literature (Alameen & Levis, 2015, p. 161).

3.4 Analysis

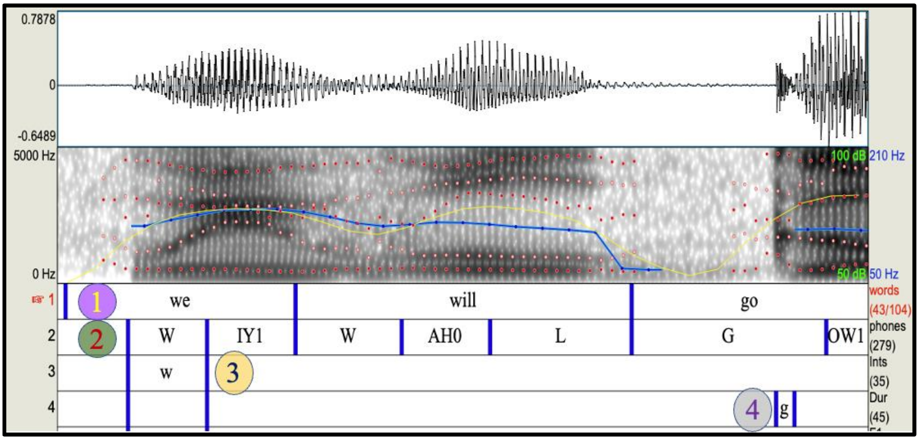
Each natural class was assigned specific words from the elicitation paragraph to analyze and measure the key consonants under investigation. Table 1 provides a visual representation of all segments that were analyzed for this study.

Stops					
peas	**big**	**toy**	**(W)day**	**kids**	**go**
[p]	[b]	[t]	[d]	[k]	[g]
Fricatives					
six	**please**	**five**	**five**	**thick**	**these**
[s]	[z]	[f]	[v]	[θ]	[ð]
her	she	plastic	(W)day		
[h]	[ʃ]	[t̪]	[d̪]		
Nasals					
meet	**need**	**bring**			
[m]	[n]	[ŋ]			
Approximants					
call	**Stella**	red	we		
[ɫ]	[l]	[ɹ]	[w]		

Table 1: Target Words for each Segments Analyzed by Natural Class

Segments highlighted in red represent those sounds that may cause intelligibility issues since they are not present as a phonemic category in MSA, according to Table 1. Additionally, for fricatives, two segments were added to accommodate for the increasing substitutions of [θ] for [t̪] and [ð] for [d̪] in GAE (Koffi, 2021). Once all tokens were assigned, automated separation of the sounds was programmed using the Montreal Forced Aligner (MFA) tool inside Praat. For this study, the MFA tool was used to eliminate the redundancy of manually creating phone delimitations for all participants. Since this data was manually verified for the tokens used, there are no adverse effects on the reliability of MFA on the data for this study. Next, to ensure data validity, all the tokens were reviewed manually inside Praat to ensure that the MFA tool correctly separated each target sound.

With each target sound isolated, a manual analysis of VOT was conducted for each stop token and every participant (192 tokens). Next, using Praat, all fricative, nasal, and approximant tokens were manually reviewed and labeled with a unique identifier to automate the data extraction using a Praat script. Once all remaining tokens were manually delimited and labeled in Praat, each participant's annotated TextGrid files were run through a modified Praat script to extract intensity, duration, F2, and F3 (1,088 tokens). Graph 8 shows a completed TextGrid analysis for participant KSAF17.



Graph 8: Praat Analysis Items for Participant KSAF17

Graph 8 details the automated and manual process completed for each participant. For KSAF17, this Praat window is focused on the utterance "we will go." Label 1 is a tier completed automatically by the MFA tool without manual intervention. Label 2 details the segment level delimitation using the MFA tool. The voiced velar approximant [w] is focused, which was manually adjusted with label 3. Label 3 shows the identifier manually inserted, confirming the segment's measurable area. Lastly, label 4 shows the manual VOT delimitation for stop [g], which is different from the MFA alignment on tiers 1 and 2. The Praat script used to extract measurements for the fricatives, nasals, and approximants is a modified one created by Dr. Elvira-Garcia. Specific parameters were determined to measure the participants' data accurately, mainly for the differences in vocal tract length between the female and male speakers.

In summary, the methodology used to determine the intelligibility of Saudi Spoken English consonants adheres closely to one followed in previous studies (Koffi, 2015 and Koffi, 2021). Most of the methodology from this work has also been preserved in this study with few modifications. These modifications include using the Montreal Forced Aligner (McAuliffe et al., 2017) and a Praat script (by Dr. Elvira-Garcia) to automate the extraction of measurements. This was done to automate some tedious pointing and clicking, which Praat is less user-friendly when dealing with large data sets. All of the consonant findings presented in the results section following have been carefully reviewed manually to preserve reliability and validity.

4. Results

The results section presents a full analysis of the key consonants of English as spoken by female and male Saudi EFL teachers and an intelligibility assessment of those segments is provided. Under investigation are the acoustic characteristics of SSE for stops (VOT), fricatives, nasals, and approximants. The following results will be sectioned into those four categories of consonants, whereas each section will give results for female participants first, then their male counterparts. This results section will depart from the traditional acoustic phonetic practice of providing all measured tokens in tables to limit the word count. Appendix E and F provide all detailed measurements. All in all, this section gives the first instrumental analysis for acoustic correlates in SSE consonants (1,280 tokens) for 32 participants.

4.1 The Stops of Saudi Spoken English

As seen previously in Graph 3, stops are also present in Modern Standard Arabic with four voiceless stops and only two voiced stops. Most favorable to unintelligibility is the lack of the voiceless bilabial [p] for Saudi speakers of English. The Relative Functional Load (Appendix A) described by

Catford (1987) shows that [p] in English carries a high load in word-initial position (98%). The second missing segment in MSA, the voiced velar [g], carries a lesser load at 49% for a word-initial position.

The key results for stops will be assessed for intelligibility using the previously discussed Koffi (2021) framework. Graph 5 details that the acoustic correlate for stop segments is Voice Onset Time and that certain segments have different temporal thresholds in GAE. Those thresholds (JND) used for stops inform any masking of segments and drive the intelligibility assessment of those consonants. Koffi (2021) proposes the following "JND for intelligibility" (p. 113) of bilabial, alveolar, and velar stops. For bilabial stops, "voiceless bilabial stops do not mask their voiced counterparts if their VOT is ≥ 26 ms and voicing during the hold period is less than 40%. If not, masking may occur." For alveolar stops, "voiceless alveolar stops do not mask their voiced counterparts if their VOT is ≥ 34 ms and voicing during the hold period is less than 40%. If not, masking may occur." Lastly, "voiceless velar stops do not mask their voiced counterparts if their VOT is ≥ 42 ms and voicing during the hold period is less than 40%. If not, masking may occur." With these intelligibility conditions in place, female SSE measured stops are summarized in Table 2.

Word	peas	big	toy	(W)day	kids	go
Segment	**[p]**	**[b]**	**[t]**	**[d]**	**[k]**	**[g]**
Female SSE SD	29.1	85.6	20.7	8.7	10.6	26.5
Female SSE Mean	**69**	**-57**	**82**	**18**	**58**	**14**
Lisker & Abramson	28	7/-65	39	9/-56	43	17/-45
JND (VOT)	≥ 26 ms		≥ 34 ms		≥ 42 ms	

Table 2: Female Saudi Spoken English VOTs (in ms)

The Lisker and Abramson (1964) measurements for GAE stops are provided to contrast the expected VOT values for listeners. For these 23 female participants, 138 VOT tokens were manually measured, resulting in the following observations. The mean VOT of SSE bilabial stops falls within the same thresholds as those of GAE expected values. The [p] segment was produced with a VOT of 69 ms on average. This shows that these female Saudi EFL teachers can confidently produce a missing segment in their L1.

It is done by lengthening their [p] by more than 40 ms than GAE speakers. The JND threshold is also attained (≥ 26 ms), and no intelligibility issues should result since no masking occurs. For the alveolar stops present in MSA, female participants produced them in line with GAE speakers. Similarly, SSE velar stops are clearly produced within the known measurements for GAE, causing no intelligibility issues. The mean value for segment [g] is 14 ms, which is within the measured VOT for GAE speakers, according to Lisker and Abramson (1964). In summary, the 23 female participants are highly intelligible when producing English stops. They can compensate for two missing segments in their L1 ([p] and [g]). A feature of SSE female stops is the lengthening of all voiced segments by at least 15 ms compared to their GAE counterparts.

For their nine male colleagues, 54 tokens were manually measured to account for the characteristics of their SSE stops. Table 3 compiles their VOTs.

Word	peas	big	toy	(W)day	kids	go
Segment	**[p]**	**[b]**	**[t]**	**[d]**	**[k]**	**[g]**
Male SSE SD	25.0	65.0	25.1	9.6	15.3	50.7
Male SSE Mean	**45**	**-34**	**60**	**22**	**58**	**-17**
Lisker & Abramson	28	7/-65	39	9/-56	43	17/-45
JND (VOT)	≥ 26 ms		≥ 34 ms		≥ 42 ms	

Table 3: Male Saudi Spoken English VOTs

Like their female counterparts, all nine of these Saudi educators can produce all of the stops for GAE. Segments [p] and [b] are within the comparable VOT means with GAE for the bilabial segments. The [p] segment is not available in the L1; however, they produce a [p] slightly lengthened by 17 ms. For an assessment of intelligibility, on average, these participants fulfill the conditions laid out earlier, with their voiceless stop above the JND of ≥ 26 ms. Some individual cases do not agree with the group. For example, participant KSAM3 produced a short [p] at 12 ms. Since the RFL for [p] is 98%, this is deemed "poor intelligibility" according to Graph 6. The participants produced the missing MSA segment [g] within the GAE range [g]. This is contrasted by a consequent variability with a high standard deviation at 50.7 ms. To conclude, all nine participants show that they can produce the missing stops from their L1 comparable to GAE stops and above the JND for intelligibility. The high numbers in standard deviation for 5 out of 6 stops show that individual variability exists, and intelligibility problems could occur in SSE stops.

4.2 The Fricatives of Saudi Spoken English

The inventory of fricatives in MSA is larger than that of GAE. With 11 segments in MSA against eight for American English speakers, the voiced labiodental [v] is not a phoneme in Arabic. Fricatives should not be a challenge for Saudi speakers of English in general; as Graph 3 shows, they have 80% of possible phonemic match. Some considerations for RFL and intelligibility are needed. Appendix A shows that [v] and [f] in word-final is has a small load at 9%. This is good for Saudi speakers of English, showing that intelligibility will not be a frequent effect for them. According to the Koffi (2021) framework, the acoustic correlate for fricatives is intensity (in dB), and its Just Noticeable Difference threshold is ≥ 3 dB (Hansen, 2001). The consideration of intelligibility for fricatives is summarized as such by Koffi (2015):

> "Fricatives can be substituted one for another without interfering with intelligibility if they agree in place of articulation and voicing, and if the intensity distance between the two is ≤ 3 dB" (p. 8)

To evaluate the possible perception of GAE listeners, the data set for fricatives and all other remaining consonants will be that of 10 GAE participants (5 females and 5 males). These 10 participants have also read the same elicitation text, and their data is available on the Speech Accent Archive. Table 4 gives information on 5 female GAE fricative measures.

Word	six	please	five	five	thick	these	her	she	plastic	(W)day
Segment	**[s]**	**[z]**	**[f]**	**[v̥]**	**[θ]**	**[ð]**	**[h]**	**[ʃ]**	**[t̪]**	**[d̪]**
CA 32 F	67	64	58	59	65	72	60	68	62	67
GA 330 F	70	64	48	64	59	62	53	60	56	61
NY 6 F	74	75	70	58	58	64	67	68	65	65
OR 184 F	72	74	66	64	63	undef	74	72	71	70

TX 286 F	74	84	66	67	63	73	70	72	71	71
Female GAE Mean	**71**	**72**	**62**	**62**	**62**	**68**	**65**	**68**	**65**	**67**
Female GAE SD	3.0	8.3	8.6	3.7	3.0	5.4	8.3	4.8	6.6	4.0

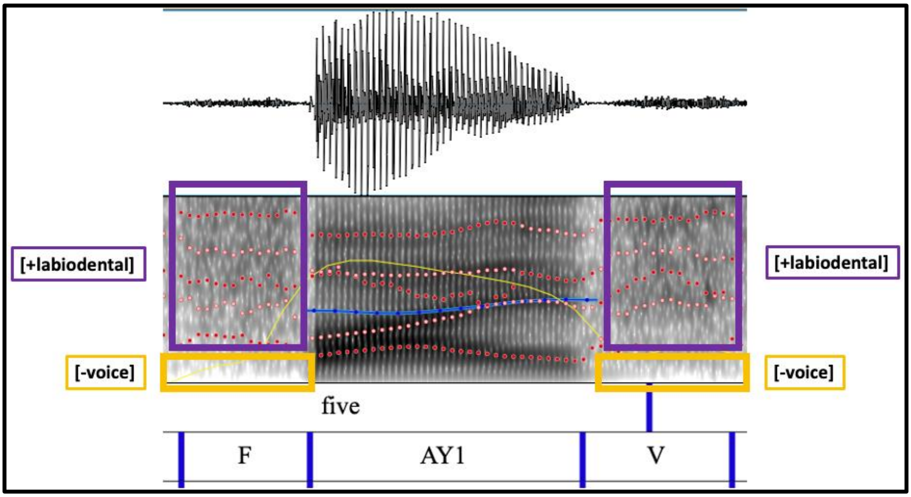
Table 4: Female GAE Fricatives (intensity in dB)

The SSE fricatives mean measurements for 23 female Saudi EFL teachers are detailed below in Table 5. A total of 10 segments were instrumentally measured for intensity for a total of 230 tokens.

Word	six	please	five	five	thick	these	her	she	plastic	(W)day
Segment	**[s]**	**[z]**	**[f]**	**[v̥]**	**[θ]**	**[ð]**	**[h]**	**[ʃ]**	**[t̪]**	**[d̪]**
Female SSE SD	5.1	5.7	5.4	4.6	7.0	7.5	3.8	5.2	8.6	8.5
Female SSE Mean	**64**	**67**	**52**	**56**	**50**	**51**	**57**	**72**	**55**	**54**
GAE Mean	71	72	62	62	62	68	65	68	65	67

Table 5: Female Saudi Spoken English Fricative Measurements
(intensity in dB)

Data from the female participants confirm, overall, that intensity plays a part in discriminating between the different types of fricatives. Only segments [θ] and [ð] show a perceptually similar intensity. This could be a group-specific finding since the participants are inconsistent in producing these sounds with a deviation of 7.0 and 7.5, respectively. Results for segment [v] show that the fricative threshold is respected with a perceptual difference between [f] at 52 dB and [v] at 56 dB. However, while marking the spectrograms for this segment, it was noted that all female participants (23) produced [v] for [f]. Graph 9 shows the spectrogram of KSAF19 for an SSE [f] segment in word-initial and [v] segment in word-final. [v] for [f].



Graph 9: Spectrogram of Word "five" for Participant KSAF19

Ladefoged (2005) informs about the characteristics of the [f] segment where "the noise is made by air being forced through a narrow gap" (p. 56), which is apparent in the purple window in Graph 9 for [f] and [v]. For voicing, it is described as a "voiceless fricative" (Ladefoged, 2005, p. 56) which is also apparent in the bottom portion of the orange window. With an RFL at 9%, this substitution is likely to cause little effect on intelligibility. The male measurements will validate if this is a particular feature of SSE.

Focus now is given to the male participants. Consistent with the female teachers, an assessment of intensity as an acoustic correlate is covered, along with the [f] and [v] pair. The GAE mean provided is from the Speech Accent Archive, which gives measurements for five male participants. The male SSE fricatives are presented in Table 6, which accounts for 90 spoken tokens.

Word	six	please	five	five	thick	these	her	she	plastic	(W)day
Segment	**[s]**	**[z]**	**[f]**	**[v̥]**	**[θ]**	**[ð]**	**[h]**	**[ʃ]**	**[t̪]**	**[d̪]**
Male SSE SD	8.7	6.6	4.6	3.8	2.3	8.7	7.4	6.6	6.2	7.2
Male SSE Mean	**64**	**67**	**56**	**55**	**55**	**54**	**60**	**70**	**57**	**63**
GAE Mean	69	71	63	61	62	67	66	66	63	68

Table 6: Male Saudi Spoken English Fricative Measurements
(intensity in dB)

The intensity factor for these Saudi teachers shows mixed results. Segments [s], [z], [h], [ʃ], [t̪], and [d̪] show a perceptual distance greater than 3 dB while the acoustic distance for [f], [v], [θ], and [ð] is under 3 dB. For the segments of interest, similar to the female teachers, nine out of nine participants have used [f] in place of the word-final [v]. Since 32 participants made this substitution, we can relate this result to Flege & Bohn's (2021) findings that "L1 phonetic categories in learners do "interfere and sometimes block" the creation of L2 categories" (p. 23).

To conclude, the high correspondence of MSA fricatives to GAE ones works in favor of Saudi speakers of English. All 32 participants produced the segment [v] as [f], confirming findings from Koffi's (2021) work on Arabic fricatives. An impressionistic review of segments is still needed to detect fricatives using thresholds. This point will be explored in future studies. As for intelligibility considerations, since [v] and [f] have a small RFL load (9%), this segment is deemed to have "good intelligibility" according to Graph 6.

4.3 The Nasals of Saudi Spoken English

Nasals in Modern Standard Arabic share two of three nasals in GAE. Alghamdi (2015) confirms that segment [m] is a voiced bilabial while [n] is a voiced alveolar in MSA. The voiced velar [ŋ] will be the focus of this section to highlight how Saudi EFL teachers produce this segment and if intelligibility can be a problem. Catford's (1987) RFL findings for the voiced velar [ŋ] in word-final point to an 18% load when replaced by [n] and 14% for [m]. Like fricatives, any substitutions in nasals have a low impact on intelligibility in Saudi Spoken English. Research conducted by Koffi (2021, pp. 179-180) on the nasals of GAE, using the same read speech data, point to intensity as a potential acoustic correlate for distinguishing between [m] and [n]. Duration, on the other hand, is an acoustic correlate between [m] and [n], and [n] and [ŋ] only. The JND threshold for intensity is ≥ 3dB and for duration ≥ 10 ms, according to Koffi (2021). Formants are also a cue to nasal intelligibility with their respective JNDs, as shown in Graph 5. In brief, segments [m] and [n] will be perceived as similar between an L1 and an L2, so long as their acoustic distance is ≤ 200 Hz for F2. For segment [ŋ], the threshold looks at F3 as the strongest cue, and the respective JND is ≤ 400 Hz.

The measurements for SSE nasals are presented next. First, the findings for intensity and duration are provided along with measurements for F2 and F3. The results presented for both female and male SSE nasals are derived from the measurements of 276 tokens. Table 7 shows intensity and duration results for female participants.

	Intensity (dB)				**Duration (ms)**		
Word	meet	need	bring		meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**		**[m]**	**[n]**	**[ŋ]**
Female SSE SD	3.5	4.3	5.2		21.7	22.0	39.3
Female SSE Mean	**66**	**68**	**64**		**91**	**64**	**111**
Female GAE Mean	70	73	70		86	60	95

Table 7: Female SSE Intensity and Duration of Nasals

For this group, intensity seems to be an acoustic cue to differentiate between SSE [n] and [ŋ] with an acoustic distance above the JND threshold (≥ 3 dB). A stronger cue is duration with all three segments perceptually above the 10 ms threshold. Table 8 details the measurements for F2, and F3.

	F2 (Hz)				F3 (Hz)		
Word	meet	need	bring		meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**		**[m]**	**[n]**	**[ŋ]**
Female SSE SD	346.2	465.9	414.3		99.2	370.1	126.7
Female SSE Mean	**1494**	**1708**	**1487**		**2615**	**2775**	**2678**
Female GAE Mean	1520	1838	1968		2646	2865	2768

Table 8: Female SSE Formant Measurements of Nasals

Based on the JND thresholds discussed previously, the F2 of SSE segment [m] (1494 Hz) is perceived as similar than the GAE [m] (1520 Hz) with only 26 Hz between them. The same occurs for segment [n]. The SSE alveolar is within the 200 Hz JND of GAE segment [n]. Lastly, for segment [ŋ], the participants produced this missing L1 sound within the 400 Hz JND of GAE, making it an intelligible segment. The results for their male counterparts are presented next. Table 9 details the intensity and duration for male SSE nasals.

	Intensity (dB)				Duration (ms)		
Word	meet	need	bring		meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**		**[m]**	**[n]**	**[ŋ]**
Male SSE SD	4.3	2.4	4.2		33.0	18.4	55.4
Male SSE Mean	**66**	**72**	**67**		**87**	**71**	**144**
Male GAE Mean	69	73	72		82	67	79

Table 9: Male SSE Intensity and Duration of Nasals

For male Saudi teachers, intensity is a cue to discriminate between [m] and [n], and [n] and [ŋ]. More robust is duration, whereas all three segments have more than 10 ms between them. Table 10 looks at the F2 and F3 for male SSE nasals.

	F2 (Hz)				F3 (Hz)		
Word	meet	need	bring		meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**		**[m]**	**[n]**	**[ŋ]**
Male SSE SD	200.3	344.0	547.6		171.4	231.8	216.0
Male SSE Mean	**1540**	**1609**	**1634**		**2453**	**2434**	**2477**
Male GAE Mean	1458	1536	1381		2528	2542	2372

Table 10: Male SSE Formant Measurements of Nasals

Similar to the female SSE measurements, all results for the male group show that segments [m], [n], and [ŋ] are intelligible to GAE listeners. The F2 values for SSE [m] and [n] are within 200 Hz of the GAE group. Segment [ŋ] is within the 400 Hz threshold, making nasals produced by male participants intelligible to GAE listeners. In conclusion, although segment [ŋ] is not present in MSA, both female and male participants produce it without substitution and are intelligible to GAE listeners. The most relevant acoustic cue for SSE nasals is duration over intensity.

4.4 The Approximants of Saudi Spoken English

The approximants of GAE are separated into liquids and glides. Leaving [j] aside, since it is not investigated here, two out of three segments are a phonemic match between GAE and MSA ([l] and [m]). The MSA segment [r] is a voiced alveolar trill, while the GAE segment [ɹ] is a voiced palatal approximant. This segment will be a focus for this section. One possible consequence to intelligibility when considering RFL occurs between [l] and [ɹ] in word-initial, with a high load at 83%. However, since the MSA segment [l] is similar to that of GAE and MSA's [r] is trilled, this RFL consideration may not be necessary. The acoustic cues for approximants in GAE show that intensity is not a factor, duration only discriminates between [ɹ] and [w], and F3 is a strong indicator between [l] and [ɹ] (Koffi, 2021, pp. 206–208). The JND threshold used by Koffi (2021) for liquids considers that "[r] masks [l] if its F3 is ≥ 2,600 Hz, unless it is trilled" (p. 208). Since trilling is possible as an L1 transfer feature for SSE, determining the JND of when trilling occurs for [r] is important. Ladefoged (2003) contends that the degree of trilling (in Hz) can be determined by dividing the absolute duration (1000 ms) over the relative duration of the segment. The threshold for trilling has also been determined ≥ 22 Hz.

	Intensity (dB)					Duration (ms)					F3 (Hz)			
Word	call	Stella	red	we		call	Stella	red	we		call	Stella	red	we
Seg-ment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**		**[ɫ]**	**[l]**	**[ɹ]**	**[w]**		**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
Female SSE SD	4.4	3.4	9.0	4.9		14.4	25.0	15.1	22.0		339.6	288.7	347.4	232.0
Female SSE Mean	**64**	**65**	**58**	**65**		**40**	**72**	**76**	**60**		**3154**	**3087**	**2689**	**2877**
Female GAE Mean	73	71	71	72		39	68	52	49		2863	3067	2281	2608

Table 11: Female SSE Intensity, Duration and F3 Measurements of Approximants

Data of approximants (276 tokens) produced by female Saudi EFL teachers are presented with findings for intensity, duration and F3 measurements of SSE segments [ɫ], [l], [ɹ], and [w] in Table 11.

Female participants do not perceptually distinguish their approximants by way of intensity. Only the SSE segment [ɹ] seems to have a significant intensity difference; however, the high standard deviation of 9 dB shows inconsistency in the way it is produced by this group. Similarly, with duration, the SSE segment [ɹ] is lengthened by 24 ms compared to the GAE segment, making this perceptually different. The degree of potential vibration is 13.15 Hz (1000/76), way below the threshold to qualify as a trilled [ɹ]. With F3 the most salient formant to determine intelligibility of [ɹ] for L2 speakers, Table 11 shows that the 23 female participants produce the segment above 2600 Hz, which according to the JND indicates that [ɹ] is masking [l] since these participants to not trill their [ɹ] segment. Furthermore, according to Graph 5 seen previously, the JND for approximants is ≥ 400 Hz. Any acoustic distance for F3 measurements below this JND between SSE and GAE is deemed perceptually similar. SSE segments [ɫ], [l], and [w] show no intelligibility issues, while SSE segment [ɹ] may be a contributor to poor intelligibility and could be confused with [l] by GAE listeners. To see if this is also the case for the male participants, Table 12 shows the intensity and duration, and F3 for male SSE approximants.

Results show that intensity is a dependable acoustic correlate in male SSE approximants for distinguishing between [ɫ] and [w] only. Duration is an acoustic cue for distinguishing between [l] and [ɹ] only. Furthermore, similar to their female colleagues, the SSE segment [ɹ] is lengthened (by 28 ms) and could be a strategy to accommodate for this new L2 category. The F3 value of segment [ɹ] is below the JND threshold for masking with [l]. Overall, all four SSE segments are perceptually intelligible since the acoustic distance is less than 400 Hz for similar segments in GAE. Overall, both female and male participants use duration as a robust acoustic correlate to distinguish their approximants. Additionally, the MSA segments present in GAE transfer without causing intelligibility issues. Only the female SSE [ɹ] segment has been shown to potentially cause masking with [l]. If so, the RFL shows a load of 83% and causes poor intelligibility. This was not observed with the male participants.

	Intensity (dB)					Duration (ms)					F3 (Hz)			
Word	call	Stella	red	we		call	Stella	red	we		call	Stella	red	we
Segment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**		**[ɫ]**	**[l]**	**[ɹ]**	**[w]**		**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
Male SSE SD	8.1	4.8	9.2	5.1		39.6	23.2	50.8	30.0		170.3	108.8	453.6	232.5
Male SSE Mean	**63**	**66**	**65**	**68**		**68**	**71**	**81**	**73**		**2587**	**2567**	**2411**	**2523**
Male GAE Mean	76	71	70	72		56	56	53	53		2450	2587	2170	2413

Table 12: Male SSE Intensity, Duration and F3 of Approximants

5. Discussion and Conclusion

This empirical research offers a rare glimpse into the segmental features of English as spoken by Saudi EFL teachers. To this day, research that has looked into the L2 speech of Saudi speakers of English overwhelmingly focuses on younger participants in an ESL setting. This work on consonants further informs on the segmental features and intelligibility of Saudi Spoken English after being done so for vowels (Duris, 2021) for a population working as EFL teachers in Saudi Arabia's most prestigious universities. The methodology used for this research is drawn from several phoneticians and acousticians in perceptual thresholds for General American English and the pronunciation work on intelligibility from Levis (2005). In 2012, Koffi started research to combine such findings in acoustic phonetics and intelligibility for L2 Englishes. Since then, many works have been published using Koffi's framework for intelligibility, and this research is the first one dedicated to consonants of Saudi L2 English. When looking into L2 segmentals from a bottom-up approach, the L1 has a large impact on the production and perception of new sounds (Flege & Bohn, 2021).

Regarding RQ1, the study divides consonants into four main groups to answer this research question, focusing on stops, fricatives, nasals, and approximants. For the stops, the missing segments in MSA are the voiceless bilabial [p] and the voiced velar [g]. All other stops have a possible phonemic match since their features (glottal state, place, and manner) are similar. For both female and male participants, results show that all six segments of stops are produced intelligibly, including the missing L1 MSA segments [p] and [g]. Furthermore, the instrumental data from the 32 participants show that lengthening is an apparent strategy used to produce segment [p]. The female SSE [p] is lengthened by 15 ms compared to known GAE thresholds. The male segment [p] is lengthened by 17 ms compared to GAE. For fricatives, the missing segment in MSA is only the voiced labiodental [v]. All 32 participants produced the L2 target segment [v] as a voiceless labiodental [f]. The perceptual difference is only present for the female participants while the male participants do not distinguish their segment [f] and [v] using intensity. This may cause an intelligibility interference further covered in RQ2 and RQ3. For nasals, both groups of Saudi EFL teachers could produce intelligibly the only missing GAE nasal in their L1, the voiced velar [ŋ]. Duration is the strongest acoustic correlate to distinguish nasals when analyzing SSE. Some differences occurred between the two groups with the last consonant group, approximants. In MSA, the rhotic segment is different because it is trilled. This did not affect the male participants, who produced intelligibly the voiced palatal [ɹ] within the same acoustic distance as GAE known literature data. However, this is different for the female EFL teachers, who show a potential intelligibility barrier for segment [ɹ], which can mask segment [l]. This is further developed in RQ2 and RQ3 next.

Two specific segments in SSE could not be perceived intelligibly based on the acoustic correlates and associated thresholds. The male participants' fricative segment [v] was substituted by the voiceless labiodental [f]. When considering RFL for such change, Appendix B points to a 9% load for a word-final replacement. This substitution does not affect intelligibility as a whole, and it will be perceived as "good intelligibility," according to Graph 6. For the approximant segment [ɹ], the acoustic measurements have shown that the female participants do not distinguish their SSE [ɹ] from their SSE [l] segment. The substitution used by these speakers does not refer to their L1. It is conclusive from the measurements that female SSE [ɹ] is not trilled; however, since F3 is the acoustic correlate used to distinguish these two segments, a lack of lip rounding and protrusion may be at play. It is also important to highlight that the threshold for this segment is ≥ 2,600 Hz. The female participants passed this threshold by only 89 Hz, with a total standard deviation for this segment at 347 Hz. Human errors in measurements should be considered. With a confusion of [ɹ] for [l], potential GAE listeners would not distinguish the segments clearly. This masking affects intelligibility since the RFL for this pair carries an 83% load, which is deemed "poor intelligibility" (Graph 6).

To conclude, Saudi Spoken English consonants would almost all be perceived intelligibly by typical GAE listeners in this study. With a total of five segments not present in their L1, these Saudi EFL teachers show a high level of intelligibility. Even when a substitution occurs ([f] for [v]), the low relative functional load for this pair in English proves beneficial for these L2 speakers.

References

Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *The Journal of the Acoustical Society of America, 51*(4B), 1219–1223.

Alameen, G., & Levis, J. M. (2015). Connected speech. *The handbook of English pronunciation*, 157–174.

Alghamdi, M. (2015). Arabic Phonetics and Phonology, Al-Toubah Bookshop. Riyadh, Kingdom of Saudi Arabia.

Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.05, retrieved 5 January 2022 from http://www.praat.org/

Byrd, D., & Mintz, T. H. (2010). *Discovering speech, words, and mind*. John Wiley & Sons.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. ed. by J. Morley, 87-100. *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, DC: Teachers of English to Speakers of Other Languages.

Duris, M. (2021). "Vowel Intelligibility Analysis of Female Saudi Spoken English," *Linguistic Portfolios*: Vol. 10 , Article 3.

Fant, G. (1998). Acoustical Analysis of Speech. *In: Handbook of Acoustics*, ed. by Malcom J. Crocker, 1245–1254. New York: A Wiley-Interscience Publication, John Wiley and Sons, Inc.

Flege, J. E. (1987). The instrumental study of L2 speech production: Some methodological considerations. *Language Learning*, *37*(2), 285–296.

Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge: Cambridge University Press. doi:10.1017/9781108886901.002.

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics* 12: 47–65.

Fletcher, H. (1953). *Speech and Hearing in Communication*. New York: D. Van Nostrand Company, Inc.

Hansen, C. H. (2001). Fundamentals of acoustics. *Occupational Exposure to Noise: Evaluation, Prevention and Control. World Health Organization*, 23–52.

Hawks, J. W. (1994). Difference limens for formant patterns of vowel sounds. *The Journal of the Acoustical Society of America*, *95*(2), 1074–1084.

Hirsh, I. J. (1959). Auditory perception of temporal order. *The Journal of the Acoustical Society of America, 31*(6), 759–767.

Ibrahim, A. B., Seddiq, Y. M., Meftah, A. H., Alghamdi, M., Selouani, S. A., Qamhan, M. A. & Alshebeili, S. A. (2020). Optimizing arabic speech distinctive phonetic features and phoneme recognition using genetic algorithm. *IEEE Access*, *8*, 200395–200411.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*, 459–489.

Kent, R. D. & Read, C. (2002). *The acoustic analysis of speech*. Singular.

Koffi, E. (2015). "The Pronunciation of Voiceless TH in Seven Varieties of L2 Englishes: Focus on Intelligibility," *Linguistic Portfolios*: Vol. 4, Article 2.

Koffi, E. (2019). "An Acoustic Phonetic Account of the Confusion between [l] and [n] by some Chinese Speakers," *Linguistic Portfolios*: Vol. 8, Article 6.

Koffi, E. (2021). Relevant acoustic phonetics of L2 English: Focus on intelligibility. CRC Press.

Koffi, E. & Ribeiro, L. D. (2016). "An Acoustic Phonetic Portfolio of a Portuguese-Accented English Idiolect," *Linguistic Portfolios*: Vol. 5, Article 8.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Wiley-Blackwell.

Ladefoged, P. (2005). *Vowels and consonants* (Vol. 1). Wiley-Blackwell.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: The MIT Press.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly*, *39*(3), 369–377.

Levis, J. M. (2018). Intelligibility, oral communication, and the teaching of pronunciation. Cambridge University Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Interspeech* (Vol. 2017, pp. 498-502).

Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *The Journal of the Acoustical Society of America*, *63*(2), 572–580.

Miller, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, *38*(1–3), 159–180.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*,(1), 73–97.

Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (1994). Auditory duration discrimination in young and elderly listeners with normal hearing. *Journal of the American Academy of Audiology*, *5*(3), 210–215.

Quené, H. (2004). What is the Just Noticeable Difference for tempo in speech?. *LOT Occasional Series*, *2*, 149–158.

Scharf, B. (1961). Complex sounds and critical bands. *Psychological Bulletin*, *58*(3), 205.

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

Weinberger, S. (2015). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu

Yost, W. A. (2015). Psychoacoustics: A brief historical overview. *Acoustics Today*, *11*(3), 46–53.

Zhang, B. (2014). "An Acoustic Phonetic Portfolio of a Chinese-Accented English Idiolect," *Linguistic Portfolios*: Vol. 3, Article 11.

Zielinski, B. (2015). 22 The Segmental/Suprasegmental Debate. *The handbook of English pronunciation*, 397.

Appendix A: Relative Functional Load Percentages (Koffi, 2021, pp. 49-50)

No.	Word Initial	Percentage	Word Final	Percentage	Words	Vowel Phonemes'	Percentage
1.	k/h	100	d/z	100	bit/bat	/ɪ/ vs. /æ/	100
2.	p/b	98	d/l	76	beet/bit	/i/ vs. /ɪ/	95
3.	p/k	92	n/l	75	bought/boat	/ɔ/ or /ɑ/ vs. /o/	88
4.	p/t	87	t/d	72	bit/but	/ɪ/ vs. /ʌ/	85
5.	p/h	85	d/n	69	bit/bait	/ɪ/ vs. /e/	80
6.	s/h	85	l/z	66	cat/cot	/æ/ vs. /ɔ/ or /ɑ/	76
7.	l/r	83	t/k	65	cat/cut	/æ/ vs. /ʌ/	68
8.	b/d	82	t/z	61	cot/cut	/ɔ/ or /ɑ/ vs. /ʌ/	65
9.	t/k	81	l/n	58	caught/curt	/ɑ/ or /ɔ/ vs. /ʌ̢ /	64
10.	t/s	81	t/s	57	coat/curt	/o/ vs. /ʌ̢ /	63
11.	d/l	79	p/t	43	bit/bet	/ɪ/ vs. /ɛ/	54
12.	p/f	77	p/k	42.5	bet/bait	/ɛ/ vs. /e/	53
13.	b/w	76	m/n	42	bet/bat	/ɛ/ vs. /æ/	53
14.	d/r	75	s/z	38	coat/coot	/o/ vs /u/	51
15.	h/zero	74	t/tʃ	31	cat/cart	/æ/ vs. /ɑ˞ /	51
16.	t/d	73	k/g	29	beet/boot	/i/ vs. /ʊ/	50
17.	b/g	71	*t/θ	27	bet/but	/ɛ/ vs. /ʌ/	50
18.	f/h	69	k/tʃ	26	bought/boot	/ɔ/ or /ɑ/ vs. /ʊ/	50
19.	f/s	64	b/d	24	hit/hurt	/ɪ/ vs. /ʌ̢ /	49
20.	n/l	61	d/g	23	beat/beard	/i/ vs. /iɚ/	47
21.	m/n	59	v/z	22	pet/pot	/ɛ/ vs. /ɑ/	45
22.	d/g	56	d/dʒ	22	hard/hide	/ɑ˞ / vs. /ɑɪ/	44
23.	ʃ/h	55	b/m	21	bet/bite	/ɛ/ vs. /ɑɪ/	43
24.	s/ʃ	53	g/ŋ	21	cart/caught	/ɑ˞ / vs. /ɑ/ or /ɔ/	43
25.	d/n	53	b/g	20	cart/cur	/ɑ˞ / vs. /ʌ̢ /	41
26.	k/g	50	n/ŋ	18	boat/bout	/o/ vs. /ɑʊ/	40.5
27.	g/w	49	p/f	17	cut/curt	/ʌ/ vs. /ʌ̢ /	40
28.	n/r	41	s/θ	17	cut/cart	/ʌ/ vs. /ɑ˞ /	38
29.	t/tʃ	39	dʒ/z	16	Kay/care	/e/ vs. /ɛ˞ /	35
30.	d/dʒ	39	m/v	16	cart/cot	/ɑ˞ / vs. /ɑ/ or /ɔ/	31.5
31.	s/tʃ	37	ŋ/l	15	*here/hair	/iɚ/ vs. /ɛ˞ /	30
32.	g/dʒ	31	p/b	14	light/lout	/ɑɪ / vs. /ɑʊ/	30
33.	b/v	29	m/ŋ	14	*cot/caught	/ɔ/ vs. /ɑ/	26

Appendix B: RFL continued

No.	Word Initial	Percentage	Word Final	Percentage	Words	Vowel Phonemes	Percentage
34.	*w/hw[10]	27	g/dʒ	13	fire/fair	/ɑɪɚ/ vs. /ɛ˞/	25
35.	*ʃ/tʃ	26	*tʃ/ʃ	12	her/here	/ɚ/ vs. /iɚ/	24
36.	*f/v	23	*f/v	9	buy/boy	/ɑɪ/ vs. /ɔɪ/	24
37.	*v/w	22	*f/θ	9	car/cow	/ɑ/ vs. /ɑʊ/	23
38.	dʒ/dr	21	tʃ/dʒ	8	her/hair	/ɚ/ vs. /ɛ˞/	21
39.	s/θ	21	b/v	7	*tire/tower	/ɑɪɚ/ vs. /ɑʊɚ/	19
40.	dʒ/j	20.5	s/ʃ	7	box/books	/ɑ/ or /ɔ/ vs. /ʊ/	18
41.	*d/ð	19	z/ð	7	*paw/pore	/ɔ/ vs /ɔ˞/	15
42.	*tʃ/dʒ	19	*θ/ð	6	pill/pull	/ɪ/ vs. /ʊ/	13.5
43.	*t/θ	18	*d/ð	5	pull/pole	/ʊ/ vs. /o/	12
44.	tʃ/tr	16	v/ð	1	bid/beard	/ɪ/ vs. / iɚ/	11
45.	*f/θ	15			bad/beard	/æ/ vs./ iɚ/	10
46.	*f/hw	13			*pin/pen	/ĩ/ vs. /ẽ/	9
47.	*v/ð	11			*put/putt	/ʊ/ vs. /ʌ/	9
48.	*kw/hw	8			bad/beard	/æ/ vs. /ɛɚ/	8
49.	d/z	7			*pull/pool	/ʊ/ vs. /u/	7[11]
50.	*s/z	6			*sure/shore	/uɚ/ vs. /ɔ˞/	5
51.	*tw/kw	5			pooh/poor	/u/ vs. /uɚ/	5
52.	v/z	2			*cam/calm	/æ/ vs. /ɑ/	4.5
53.	*θ/ð	1			purr/poor	/ʌ/ vs. /uɚ/	4.5
54.	*z/ð	1			good/gourd	/ʊ/ vs. /ʌ/	1

Appendix C: Linguistic Profile Data for Female Saudi Participants

Participant	Age	Country of Birth	City of Birth	Cities lived in KSA	Dialect ascription	Inner Circle life	If yes, age outside KSA	If yes, English used?	Age of 1st Spoken English	Age of 1st Eng class	Major Contrib. to fluency
KSAF1	35	KSA	Jeddah	Jeddah Riyadh	Hijazi	UK / Canada	Adult	Yes	12	13	Entertainment
KSAF2	30	KSA	Riyadh	Riyadh	Najdi	-	-	-	12	11	NS Interaction
KSAF3	27	KSA	Riyadh	Riyadh	Najdi	-	-	-	7	7	Entertainment/NS Interaction
KSAF4	31	KSA	Riyadh	Riyadh	Najdi	UK / USA	Child/Adult	Yes	12	12	Entertainment
KSAF5	35	KSA	Riyadh	Riyadh	Najdi	-	-	-	5	11	Entertainment
KSAF6	37	KSA	Riyadh	Riyadh	Najdi	USA	Child	Yes	7	3	Inner Cirlce Childhood
KSAF7	34	KSA	Riyadh	Riyadh	Southern	-	-	-	11	11	Sibling
KSAF8	31	KSA	Riyadh	Riyadh	Najdi	-	-	-	12	16	Entertainment/NS Interaction
KSAF9	45	KSA	Riyadh	Riyadh	Najdi	UK	Adult	Yes	18	17	NS Interaction
KSAF10	29	KSA	Medina	Medina Riyadh	Hijazi	-	-	-	6	6	Entertainment
KSAF11	29	KSA	Riyadh	Riyadh	Najdi	USA	Child	Yes	8	5	School
KSAF12	35	KSA	Riyadh	Riyadh	Najdi	USA	Child	No	6	6	Inner Cirlce Childhood
KSAF13	30	KSA	Taif	Taif Riyadh	Hijazi	USA	Child	Yes	6	6	Inner Cirlce Childhood
KSAF14	34	KSA	Jubail	Dammam Riyadh	Gulf	USA	Child	Yes	13	13	NS Interaction
KSAF15	35	KSA	Riyadh	Riyadh	Hijazi	-	-	-	12	12	Entertainment
KSAF16	30	KSA	Riyadh	Riyadh	Najdi	Australia	Adult	Yes	11	11	Entertainment
KSAF17	41	KSA	Riyadh	Riyadh	Najdi	Canada	Child	Yes	13	13	Entertainment/NS Interaction
KSAF18	34	KSA	Riyadh	Riyadh	Najdi	-	-	-	6	12	Entertainment/NS Interaction
KSAF19	37	KSA	Riyadh	Riyadh	Najdi	USA	Adult	Yes	13	13	NS Interaction
KSAF20	31	KSA	Riyadh	Jubail Jeddah Hail Riyadh	Northern	-	-	-	12	12	School
KSAF21	19	KSA	Ottowa, CA	Riyadh	Najdi	Canada	Child	Yes	5	5	Inner Cirlce Childhood
KSAF22	25	KSA	Jeddah	Jeddah Riyadh Taif Tabuk	Najdi	-	-	-	13	9	Entertainment/NS Interaction
KSAF23	30	KSA	Riyadh	Riyadh	Najdi	UK/US	Adult	Yes	7	7	Tutoring

Appendix D: Linguistic Profile Data for Male Saudi Participants

Participant	Age	Country of Birth	City of Birth	Cities lived in KSA	Dialect ascription	Inner Circle life	If yes, age outside KSA	if yes, English used?	Age of 1st Spoken English	Age of 1st Eng class	Major Contrib. to fluency
KSAM1	30	KSA	Riyadh	Riyadh	Najdi	-	-	-	21	15	NS Interaction
KSAM2	28	USA	Michigan	Riyadh	Najdi	USA	Child/Adult	Yes	5	5	Inner Circle Childhood
KSAM3	28	KSA	Makkah	Makkah Riyadh	Najdi	Canada	Adult	Yes	5	9	Entertainment
KSAM4	31	KSA	Buraydah	Buraydah Riyadh	Najdi	USA	Adult	Yes	16	15	Entertainment/NS Interaction
KSAM5	38	KSA	Riyadh	Riyadh	Najdi	-	-	-	19	13	School
KSAM6	39	KSA	Riyadh	Riyadh	Najdi	-	-	-	16	13	Entertainment
KSAM7	53	KSA	Medina	Riyadh	Najdi	-	-	-	19	13	NS Interaction
KSAM8	32	KSA	Riyadh	Riyadh	Northern	UK/US	Adult	Yes	11	11	NS Interaction
KSAM9	28	KSA	Riyadh	Hafir Batin	Najdi	USA	Adult	Yes	24	17	NS Interaction

Appendix E: Female SSE Analyzed Tokens

Female Saudi Spoken English VOTs (in ms)

Word	peas	big	toy	(W)day	kids	go
Segment	**[p]**	**[b]**	**[t]**	**[d]**	**[k]**	**[g]**
KSAF1	31	-86	54	11	53	16
KSAF2	73	7	63	22	49	22
KSAF3	23	-47	93	13	62	28
KSAF4	9	-155	90	14	58	16
KSAF5	61	11	85	4	66	-64
KSAF6	91	9	62	30	52	17
KSAF7	71	9	100	6	53	16
KSAF8	117	17	62	10	48	19
KSAF9	68	-108	96	18	60	14
KSAF10	63	153	97	9	55	31
KSAF11	93	14	102	29	90	37
KSAF12	51	16	36	37	56	-70
KSAF13	34	-122	102	17	59	39
KSAF14	43	-108	80	14	55	16
KSAF15	64	-106	79	14	51	19
KSAF16	70	-131	79	16	47	15
KSAF17	47	9	63	15	41	10
KSAF18	100	23	117	27	62	20
KSAF19	91	-108	66	23	55	16
KSAF20	97	-94	76	34	56	25
KSAF21	108	-183	120	19	57	19
KSAF22	78	-125	69	24	57	31
KSAF23	107	-201	92	19	82	19
Female SSE SD	29.1	85.6	20.7	8.7	10.6	26.5
Female SSE Mean	**69**	**-57**	**82**	**18**	**58**	**14**
Lisker & Abramson	28	7/-65	39	9/-56	43	17/-45
JND (VOT)	≥ 26 ms		≥ 34 ms		≥ 42 ms	

Female Saudi Spoken English Fricative Measurements (intensity in dB)

Word	six	please	five	five	thick	these	her	she	plastic	(W)day
Segment	**[s]**	**[z]**	**[f]**	**[v]**	**[θ]**	**[ð]**	**[h]**	**[ʃ]**	**[t̪]**	**[d̪]**
KSAF1	61	69	49	56	51	61	55	73	38	48
KSAF2	61	65	49	56	47	53	54	78	56	62
KSAF3	63	64	61	61	52	55	60	67	36	36
KSAF4	73	74	50	59	62	56	64	75	61	37
KSAF5	68	66	51	57	50	52	57	75	57	47
KSAF6	62	55	55	56	51	60	59	68	62	50
KSAF7	72	69	63	67	48	64	63	79	62	61
KSAF8	61	66	48	49	45	38	51	67	48	42
KSAF9	65	72	57	61	62	43	52	83	52	54
KSAF10	67	68	52	61	51	46	57	69	60	64
KSAF11	64	67	56	53	50	50	55	67	48	55
KSAF12	57	64	53	53	47	39	58	62	57	61
KSAF13	72	71	50	54	38	40	59	73	64	64
KSAF14	69	72	53	62	59	52	56	68	39	48
KSAF15	63	62	46	56	57	45	53	79	53	56
KSAF16	63	64	53	57	54	47	61	72	52	61
KSAF17	61	68	45	45	57	51	62	71	62	55
KSAF18	59	70	55	56	47	46	53	74	64	54
KSAF19	65	70	57	59	60	61	62	76	63	62
KSAF20	66	72	45	53	43	53	57	75	56	52
KSAF21	53	55	40	53	40	61	53	66	64	55
KSAF22	68	74	56	55	51	46	62	78	64	68
KSAF23	57	55	50	58	38	50	55	72	55	52
Female SSE SD	5.1	5.7	5.4	4.6	7.0	7.5	3.8	5.2	8.6	8.5
Female SSE Mean	**64**	**67**	**52**	**56**	**50**	**51**	**57**	**72**	**55**	**54**
GAE Mean	71	72	62	62	62	68	65	68	65	67
JND Fricatives ≤ 3dB										

Female SSE Intensity and Duration of Nasals

	Intensity (dB)			Duration (ms)		
Word	meet	need	bring	meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**
KSAF1	67	70	64	117	50	121
KSAF2	66	71	64	70	49	106
KSAF3	66	58	66	109	82	98
KSAF4	66	73	67	103	55	41
KSAF5	69	72	65	88	47	84
KSAF6	69	68	66	122	38	97
KSAF7	70	70	58	107	56	77
KSAF8	59	59	60	132	72	119
KSAF9	68	68	63	81	133	58
KSAF10	72	68	62	94	50	85
KSAF11	63	66	60	114	71	140
KSAF12	66	66	59	62	50	200
KSAF13	63	61	65	56	71	115
KSAF14	70	73	71	92	50	117
KSAF15	61	68	65	56	59	136
KSAF16	67	70	65	86	80	141
KSAF17	70	74	71	108	63	174
KSAF18	65	65	65	64	80	170
KSAF19	68	71	66	76	80	127
KSAF20	63	65	45	69	36	43
KSAF21	68	66	65	102	90	85
KSAF22	61	72	67	90	35	101
KSAF23	72	68	64	90	80	108
Female SSE SD	3.5	4.3	5.2	21.7	22.0	39.3
Female SSE Mean	**66**	**68**	**64**	**91**	**64**	**111**
Female GAE Mean	70	73	70	86	60	95

Female SSE Formant Measurements of Nasals

	F1 (Hz)			F2 (Hz)			F3 (Hz)		
Word	meet	need	bring	meet	need	bring	meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**
KSAF1	380	391	485	1531	1339	1512	2524	2343	2609
KSAF2	438	341	437	2169	1749	2291	2741	3012	2809
KSAF3	460	421	529	1540	1557	1537	2618	2878	2653
KSAF4	501	430	607	1153	1795	817	2574	2755	2797
KSAF5	403	378	457	1547	1967	1251	2659	2411	2782
KSAF6	305	429	309	1601	1154	1408	2622	2354	2523
KSAF7	465	398	455	2436	2304	1967	2885	2919	2878
KSAF8	391	484	416	1446	1094	1314	2606	2673	2710
KSAF9	440	477	466	1362	2504	1121	2487	3447	2701
KSAF10	329	191	595	1340	1151	2159	2477	2575	2609
KSAF11	352	401	421	1470	1437	1707	2630	2425	2715
KSAF12	461	569	356	1306	1167	728	2625	2564	2604
KSAF13	321	461	388	1698	1518	1532	2764	2905	2824
KSAF14	462	541	421	1375	1449	1266	2606	2333	2468
KSAF15	392	468	506	1158	2142	1469	2533	2647	2670
KSAF16	394	326	408	1613	1649	1545	2595	2552	2602
KSAF17	442	358	478	1882	1307	1535	2581	2647	2558
KSAF18	521	438	532	1279	1031	968	2545	2516	2702
KSAF19	451	694	495	1004	2087	1036	2527	3272	2499
KSAF20	432	483	531	1225	2350	1822	2535	3704	2938
KSAF21	354	347	396	1426	2258	1817	2560	3269	2492
KSAF22	415	432	489	1830	1964	2091	2772	2951	2768
KSAF23	329	482	428	973	2302	1313	2668	2662	2687
Female SSE SD	59.1	97.3	71.4	346.2	465.9	414.3	99.2	370.1	126.7
Female SSE Mean	**410**	**432**	**461**	**1494**	**1708**	**1487**	**2615**	**2775**	**2678**

Female SSE Intensity and Duration of Approximants

	Intensity (dB)				Duration (ms)			
Word	call	Stella	red	we	call	Stella	red	we
Segment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
KSAF1	71	62	46	61	32	63	95	87
KSAF2	66	72	62	61	46	49	55	49
KSAF3	68	70	59	70	26	87	75	43
KSAF4	61	66	63	70	40	91	75	96
KSAF5	70	67	77	66	22	62	45	95
KSAF6	61	63	60	67	30	66	68	69
KSAF7	67	66	58	66	20	52	102	57
KSAF8	61	63	42	63	51	76	72	62
KSAF9	59	68	60	56	45	54	103	82
KSAF10	65	65	65	68	34	48	90	51
KSAF11	61	62	48	62	32	51	60	68
KSAF12	58	65	59	58	27	62	60	85
KSAF13	68	68	41	69	38	119	70	35
KSAF14	63	68	70	71	50	73	88	90
KSAF15	57	58	56	65	26	29	95	66
KSAF16	61	62	52	70	30	63	64	41
KSAF17	62	62	50	74	61	80	80	45
KSAF18	61	66	60	58	55	66	64	79
KSAF19	67	69	72	66	76	114	80	28
KSAF20	71	61	55	67	57	130	79	37
KSAF21	60	70	64	57	23	50	79	34
KSAF22	70	64	56	61	44	98	87	42
KSAF23	68	64	64	65	47	71	70	37
Female SSE SD	4.4	3.4	9.0	4.9	14.4	25.0	15.1	22.0
Female SSE Mean	**64**	**65**	**58**	**65**	**40**	**72**	**76**	**60**
Female GAE Mean	73	71	71	72	39	68	52	49

Female SSE Formant Measurements of Approximants

	F1 (Hz)				F2 (Hz)				F3 (Hz)			
Word	call	Stella	red	we	call	Stella	red	we	call	Stella	red	we
Segment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
KSAF1	507	509	556	528	1293	1168	2224	923	3242	3038	2785	3005
KSAF2	600	619	422	527	1204	1240	1251	1749	3609	3313	2412	2816
KSAF3	491	496	564	503	1349	1269	1468	1248	3427	2881	2534	3062
KSAF4	767	563	510	559	2197	1134	1272	914	3460	3230	2467	2920
KSAF5	759	542	522	643	1545	1539	1459	1540	3407	3286	2557	3330
KSAF6	536	509	492	374	1158	1272	1470	945	3098	3006	2641	2586
KSAF7	476	531	882	456	1758	1434	2626	1040	3140	3296	3546	2780
KSAF8	437	530	417	425	1258	1282	1643	1308	2865	2899	2761	2472
KSAF9	593	476	452	472	1668	1653	1232	1973	2757	3058	2338	3180
KSAF10	496	486	472	454	2517	1183	1451	1169	3337	3297	2305	2570
KSAF11	630	470	806	430	1482	1288	1950	1274	3044	3171	2999	3174
KSAF12	732	451	465	554	1081	1560	1762	1976	2063	2742	2685	3268
KSAF13	601	555	376	429	999	955	1659	1372	3435	2337	2550	2997
KSAF14	588	517	437	447	1640	1496	1223	998	3234	2596	2424	2944
KSAF15	690	550	643	451	1666	1141	997	942	3119	2947	2533	2691
KSAF16	625	417	423	474	1338	1662	1947	1052	3535	3064	2623	3046
KSAF17	498	658	956	510	1208	1376	2104	1314	3246	3452	3303	2898
KSAF18	442	525	2157	688	1382	1345	2406	1431	2739	2928	3504	2671
KSAF19	500	479	469	421	1614	1175	1372	1814	3408	3389	2402	2722
KSAF20	601	387	746	574	1495	1682	1961	1144	3026	3403	2769	2670
KSAF21	743	470	466	463	1300	1103	1383	1622	3387	3414	2585	2867
KSAF22	461	463	544	430	1401	1306	1821	1208	2935	3385	2766	2668
KSAF23	585	392	539	400	1550	1551	1439	1400	3032	2859	2366	2834
Female SSE SD	103.3	64.3	368.4	77.0	344.0	200.0	415.4	328.9	339.6	288.7	347.4	232.0
Female SSE Mean	**581**	**504**	**622**	**487**	**1483**	**1340**	**1657**	**1320**	**3154**	**3087**	**2689**	**2877**
Female GAE Mean	561	611	429	443	1375	1129	1544	1578	2863	3067	2281	2608

Appendix F: Male SSE Analyzed Tokens

Male Saudi Spoken English VOTs

Word	peas	big	toy	(W)day	kids	go
Segment	**[p]**	**[b]**	**[t]**	**[d]**	**[k]**	**[g]**
KSAM1	50	-55	73	8	53	-41
KSAM2	79	21	50	38	77	-52
KSAM3	12	-69	45	22	48	24
KSAM4	42	10	77	14	64	28
KSAM5	34	12	*[4]	19	71	14
KSAM6	46	9	111	24	66	24
KSAM7	23	-178	43	33	37	-79
KSAM8	32	8	37	24	35	28
KSAM9	89	-60	46	13	70	-98
Male SSE SD	25.0	65.0	25.1	9.6	15.3	50.7
Male SSE Mean	**45**	**-34**	**60**	**22**	**58**	**-17**
Lisker & Abramson	28	7/-65	39	9/-56	43	17/-45
JND (VOT)	≥ 26 ms		≥ 34 ms		≥ 42 ms	

Male Saudi Spoken English Fricative Measurements (intensity in dB)

Word	six	please	five	five	thick	these	her	she	plastic	(W)day
Segment	[s]	[z]	[f]	[v]	[θ]	[ð]	[h]	[ʃ]	[t̪]	[d̪]
KSAM1	63	76	57	50	60	61	62	73	64	72
KSAM2	46	62	56	53	54	48	53	64	55	66
KSAM3	61	60	56	55	55	50	53	74	53	59
KSAM4	62	68	55	57	55	60	58	62	53	63
KSAM5	63	70	55	61	54	58	63	69	59	62
KSAM6	72	63	62	57	57	46	68	77	64	68
KSAM7	78	78	63	58	54	71	71	78	66	72
KSAM8	66	62	55	51	52	45	64	73	52	54
KSAM9	63	62	47	51	54	49	49	60	49	52
Male SSE SD	8.7	6.6	4.6	3.8	2.3	8.7	7.4	6.6	6.2	7.2
Male SSE Mean	**64**	**67**	**56**	**55**	**55**	**54**	**60**	**70**	**57**	**63**
GAE Mean	69	71	63	61	62	67	66	66	63	68
JND Fricatives ≤ 3dB										

Male SSE Intensity and Duration of Nasals

	Intensity (dB)			Duration (ms)		
Word	meet	need	bring	meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**
KSAM1	68	68	71	149	52	150
KSAM2	59	74	61	54	50	100
KSAM3	67	71	61	91	70	137
KSAM4	66	70	66	117	69	189
KSAM5	69	70	69	43	114	104
KSAM6	60	73	72	87	69	113
KSAM7	73	76	70	90	70	100
KSAM8	65	71	67	94	75	270
KSAM9	66	71	63	56	71	134
Male SSE SD	4.3	2.4	4.2	33.0	18.4	55.4
Male SSE Mean	**66**	**72**	**67**	**87**	**71**	**144**
Male GAE Mean	69	73	72	82	67	79

Male SSE Formant Measurements of Nasals

	F1 (Hz)			F2 (Hz)			F3 (Hz)		
Word	meet	need	bring	meet	need	bring	meet	need	bring
Segment	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**	**[m]**	**[n]**	**[ŋ]**
KSAM1	300	305	309	1397	1382	1474	2421	2512	2274
KSAM2	502	418	478	1475	1472	1417	2367	2480	2534
KSAM3	451	410	489	2022	1741	2479	2511	2497	2565
KSAM4	328	306	348	1446	1593	1362	2530	2646	2479
KSAM5	342	420	605	1398	1082	1697	2480	1942	2840
KSAM6	386	759	399	1536	1712	769	2747	2351	2171
KSAM7	464	509	498	1445	1558	1518	2095	2217	2232
KSAM8	319	330	484	1679	2362	2501	2494	2621	2551
KSAM9	475	333	401	1459	1579	1489	2436	2638	2651
Male SSE SD	77.4	143.5	90.1	200.3	344.0	547.6	171.4	231.8	216.0
Male SSE Mean	**396**	**421**	**446**	**1540**	**1609**	**1634**	**2453**	**2434**	**2477**
Male GAE Mean	281	275	309	1458	1536	1381	2528	2542	2372

Male SSE Intensity and Duration of Approximants

	Intensity (dB)				Duration (ms)			
Word	call	Stella	red	we	call	Stella	red	we
Segment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
KSAM1	69	57	45	68	103	100	201	51
KSAM2	49	69	64	69	127	67	30	35
KSAM3	70	68	66	73	27	34	80	64
KSAM4	62	64	68	58	36	93	75	132
KSAM5	67	69	64	63	40	51	98	83
KSAM6	64	61	80	75	51	62	45	84
KSAM7	69	73	65	70	27	62	50	44
KSAM8	64	66	70	69	107	104	54	96
KSAM9	49	65	62	70	95	66	96	65
Male SSE SD	8.1	4.8	9.2	5.1	39.6	23.2	50.8	30.0
Male SSE Mean	**63**	**66**	**65**	**68**	**68**	**71**	**81**	**73**
Male GAE Mean	76	71	70	72	56	56	53	53

Male SSE Formant Measurements of Approximants

	F1 (Hz)				F2 (Hz)				F3 (Hz)			
Word	call	Stella	red	we	call	Stella	red	we	call	Stella	red	we
Segment	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**	**[ɫ]**	**[l]**	**[ɹ]**	**[w]**
KSAM1	470	376	824	438	1143	1184	1551	1378	2481	2595	2379	2439
KSAM2	768	508	507	377	1729	1041	1582	1555	2761	2599	2522	2282
KSAM3	554	504	482	455	988	1176	1176	916	2760	2632	1940	2504
KSAM4	546	505	448	495	1171	1243	1314	1457	2607	2385	2134	2553
KSAM5	531	542	557	466	1361	1155	1012	1165	2414	2436	2146	2533
KSAM6	519	382	488	562	1111	1633	1711	975	2364	2584	2386	2249
KSAM7	531	482	369	606	976	1018	1679	1967	2414	2759	2334	3061
KSAM8	443	398	944	372	1602	2293	1595	779	2784	2534	2326	2558
KSAM9	559	193	1471	441	1215	1766	2618	924	2696	2582	3529	2531
Male SSE SD	91.6	108.9	351.4	77.4	261.6	424.4	455.8	385.2	170.3	108.8	453.6	232.5
Male SSE Mean	**547**	**432**	**677**	**468**	**1255**	**1390**	**1582**	**1235**	**2587**	**2567**	**2411**	**2523**
Male GAE Mean	588	529	295	375	1090	1087	1515	1499	2450	2587	2170	2413

Received: 1 October 2022
Accepted for publication: 31 October 2022