



Universiteit  
Leiden  
The Netherlands

## Segmental analysis of speech intelligibility problems among Sudanese listeners of English

Tajeldin Ali, E.M.; Heuven, V.J.J.P. van; Acar A., Robertson P.

### Citation

Tajeldin Ali, E. M., & Heuven, V. J. J. P. van. (2009). Segmental analysis of speech intelligibility problems among Sudanese listeners of English. *Journal Of English As An International Language*, 4, 129-165. Retrieved from <https://hdl.handle.net/1887/16049>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/16049>

**Note:** To cite this publication please use the final published version (if applicable).



**Title**

Segmental analysis of speech intelligibility problems among Sudanese listeners of English

**Authors**

Ezzeldin Mahmoud Tajeldin Ali and Vincent J. van Heuven

**Biodata**

- Ezzeldin Mahmoud Tajeldin Ali, Ph.D candidate, Phonetics Laboratory, Leiden University,
  - Language lecturer, Dept of English, Gadarif University, Eastern Sudan
  - Areas of interest English phonetics and phonology, contrastive linguistics.
- Member of EFL syllabus design committee, Faculty of Education, Gadarif University.

[M.T.A.Ezzeldin@hum.leidenuniv.nl](mailto:M.T.A.Ezzeldin@hum.leidenuniv.nl)

Vincent J. van Heuven is professor of Experimental Linguistics and Phonetics at the Leiden University Centre for Linguistics, and former director of the institute. He is a member of the Royal Netherlands Academy of Arts and Sciences, secretary of the Permanent Council of the IPA, and associate editor of *Phonetica*.

**Abstract**

This paper aims to investigate the problems of speech intelligibility of Sudanese university learners of English. The whole work was done on the basis of segmental analysis of vowels, consonants, and consonant clusters of English so as to explore the types of perception errors made in the areas under concern. Ten Sudanese learners of English (both male and female) were selected for the experiments. The subjects were asked to listen to four lists of words that include vowels, single and cluster consonants which work in an integrative way, and a list of SPIN

sentences (SPIN = Speech Perception in Noise test, developed by Kalikov, Stevens and Elliot, 1977). The single-item stimuli were constructed on the basis of the Modified Rhyme Test (MRT) but with a few potential improvements. It is less time consuming as the number of the stimuli is reduced. Moreover, the MRT provides reliable results even with small groups of 10 to 20. The obtained information can be analyzed by confusion matrices that will in turn show how different phonemes are misidentified. Thus, the MRT helps localize the learning difficulties.

Errors were committed by Sudanese listeners at vowel, consonant, and cluster levels, in addition to SPIN sentences. But more errors were made in the perception of vowels, coda consonants, clusters of English, and SPIN sentences. English vowels proved to be the most difficult area of perception to the listeners, more so than the single and cluster consonants because the students are not familiar with a large number of vowels. Listeners use their L1 perceptual strategies, and fall back on L1 inventory when L2 knowledge is lacking.

**Keywords:** Sudanese Arabic, foreign accent, second language acquisition, speech intelligibility, SPIN test, Modified Rhyme Test, contrastive analysis, native-language interference, transfer, wrong implementation, communication breakdown, phonemic awareness, acoustic cues, phoneme inventory, basic sound knowledge.

## **Introduction**

This paper aims to present experimental evidence for the causes of speech intelligibility problems which face Sudanese university listeners of English. The study was done on the basis of segmental analysis of vowels, single consonants, and consonant clusters of English. It explores the types of perception errors made in the areas under concern, accounting for issues like how vowels, consonants, and clusters of English manifest themselves as perception problems, and what the major causes of such problems are. The paper also attempts to account for how the experimental subjects in this study deal with the influence of consonants on vowels as an example of the ways in which speech sounds interact in different phonetic environments. That is, listeners need to know that in some environments, the vowel /i:/ e.g., in *beat*, *beep* should not be realized precisely the same as /i:/ in *peat* or *keep*, which often reduces the intelligibility of a foreign learner of English (Allen and Miller 1999). Moreover, given that pronunciation plays a

prominent linguistic role in accounting for speech intelligibility between L1 and L2 speech participants, the study examined how the differences of the phonetic, and phonological properties across languages add to the problems of the speech perception. For example, when L2 norms are lacking learners usually fall back on habits of their mother-tongue. Finally, this issue is discussed into four sections where each section integrates with the others in a way as to provide coherence between the components of such sections.

## **Literature Review**

### **Speech perception problems**

To our knowledge, a very few reports have been provided about the perception problems of English speech among Sudanese listeners. The perception of the English vowels proved to be difficult for the Sudanese university listeners. In this concern, the listeners cannot discriminate between /e/ and /ei/ in words like *let, shade, make, rate*, etc. Moreover, the English tense and lax vowels /ɪ, i:/, and /ʊ, u:/ are frequently confused in words such as *beat/bit, sit/seat*. Listeners also fail to deal with vowels such as *pot, put, pert cut* etc. This is probably because their L1 (Arabic) lacks central vowels (Brett 2004).

Munro (1993) states that such types of errors occur due to the wrong realization of the English vowel categories which occur when listeners use their L1 perceptual strategies for the perception of English vowels. The English consonantal sounds also form problems for our listeners. For instance, there are interchangeable substitutions of [s, θ] in words like *sick/thick*, and *sink/think*, and of [ð, z] in words like *then/zen, zone, that*, etc. Similar errors are made in the perception of the English approximants /r, l, w/. The sound /w/ is often heard as /r, l/ as in *rent /lent/went*. It is probably due to similarity in the manners of articulation between these two approximants. This type of substitution error reveals a kind of linguistic development where there is a phonological rule merging /r/ with /w/. This rule normally appears in the child's linguistic development as temporary rule which is replaced later by appropriate one. It reinforces the potential that two different phonological representations are often possible for the same sound (Hyman 1975: 22-23). Literature on EFL learners shows that differences in phonetic and phonological implementation in a learner's mother-tongue can often result in misperception of the speech

sounds of L2. For example, they can make it difficult for such listeners/speakers to correctly identify the phonemes produced by the native speakers of L2. An acoustic matter such as the VOT often presents an element of difference which leads to misperception. Italian-English bilinguals identify the voiced English stops /b, d, g/ as voiceless /p, t, k/. This problem is attributed to the assumption that Italian voiced stops are pre-voiced which requires that glottal pulsing starts before the articulation of the consonant, whereas it is totally the opposite in English (Rasmussen 2007: 4-32). Arabic native speakers learning English experience a similar problem but only in the implementation of English /p, b/ which is also due to pre-voicing property in Arabic inventory (Flege 1980). However, this is not the most serious speech perception element, many other factors lead to perception problems of English consonants such as /ʃ, dʒ, f, v, g/ especially among Arabic speaking learners of English.

Similarly, Sudanese listeners of English have difficulty in the recognition of the English cluster items. In fact clusters like that of English, are totally absent from the Arabic consonant inventory. Probably this makes the learning of English clusters difficult for our listeners. For instance, clusters like /nt/ is heard as /mt/, /pl/ as /bl/, or /dl/, /ts/, /tz/, /pr/ as /pr/, /dr/ as /gr/, /θr/ as /θr/, etc. Cluster items like /nt/ are heard as /mt/, /pl/ as /bl/, or /dl/, /ts/, /tz/, /pr/ as /pr/, /dr/ as /gr/, /θr/ as /θr/, etc. These types of confusion can be referred to several factors. Similarities between the members of sonorant consonant clusters often motivate phonological change which triggers perceptual confusion. Seo (2003: 50-60) argues that segments' positional restriction motivates phonological alternations on similar consonant clusters which result in poor speech perception. An account of speech perception of some cross-linguistic patterning provides correct predictions that homorganic C/liquid sequences are more likely to undergo phonological change than heterorganic C/liquid sequences in a given language. Findings of cross language investigations of 31 world languages from different language families show that nasal/liquid, obstruent/liquid clusters (or sonorant/sonorant and obstruent/sonorant sequences) of homorganic sequences like lp, rk, pl, kr and /pr/, /br/ and /nt/, /lt/, etc. are more vulnerable to phonological change than that of heterorganic sequences.

However, compared with heterorganic consonants, homorganic consonants have an additional shared acoustic property, e.g. vowel formant transitions for the same place of articulation, assuming that they are adjacent to a vowel. Thus, the two sounds in a homorganic C/liquid sequence can be considered as being phonetically more similar to each other than those in a heterorganic C/liquid sequence. Moreover, phonological change can also occur due to the absence of contexts with appropriate phonetic cues: e.g. velar-to-alveolar shift is interpreted as a repair strategy. According to Kawasaki (1982) if two sounds in a sequence are acoustically and auditorily similar, the degree of distinctiveness of the two sounds would be diminished and thus they would be subject to modification. However, such types of perception problems are widely spread among the Sudanese listeners of English which necessitate investigation.

### **Linguistic Background: The phonemic inventories of English and Arabic languages**

Vowels: Important information in this context is that the first language of our subjects is Arabic, a language which has a small inventory of vowel sounds. It maintains a classical triangular Proto-Semitic (PS) vocalism which is represented as / i, u, a / . In Classical Arabic (CA) and in Modern Standard Arabic (MSA), such vowels are geminated to give long vowels. However, many dialects in MSA have developed other vowels (Kaye 1997:188-204, Munro 1993: 41-43). Moreover, Arabic short vowels are normally not represented in letters at all, but indicated by special marking (diacritics) that have an essential morph-phonemic function in the root structure of the Arabic words.<sup>1</sup>

For example, Arabic verbal roots such as *drs*, *ktb*, and *hml* are interspersed with diacritics; *darasa* ‘he studied’, *kataba* ‘he wrote’, *hamala* ‘he carried’, respectively-- a process that reveals a non-concatenative morphological system of a deep “underlying” phonological analysis (Kenstowics 1994: 394-405, Frisch 1996, Nwesri, Tahaghoghi and Scholer 2006). Thus, Arabic vowels show correspondence to only similar English vowels. Munro (1993) stated that Arabic classical PS vowels / i, u, a/ stand for lax/short vowels /ɪ, ʊ, ʌ/, whilst their geminated forms plus the newly developed vowels /e, o/ are realized as tense/long vowels /i:, u:, a:, e:, ɔ:/. The Sudanese Arabic vowel inventory has adopted the MSA inventory, but it contrasts /e/ and /e:/. The long vowels are shortened in word-final position, i.e., the long vowel /a:/ is reduced here to

[a] (Raimy 1997, Munro 1993). In comparison to the Arabic vowel inventory, the Received Pronunciation (RP) English vowel system is complex. It consists of twenty vowel phonemes, i.e. twelve monophthongs and eight diphthongs. The RP vowel system becomes more complicated with durational variation, especially due to a tense vs. lax opposition in the monophthongs. Among the most common phonemic features of RP there is a widespread loss of /uə/ and merger of /ɔ:/ in words like *sure*, although other words may retain /uə/, e.g. *poor*. There is no longer a distinction between /ɔə/ for speakers with /ɔ:/ e.g. in words like *paw*, *port*, and *talk*, etc. Thus, some words such as *sure* are pronounced as /ʃɔ:/ *shore*, but *poor* as / puə/. In the majority of accents now the phoneme /u:/ is commonly used in words like *suit*, and *enthusiasm*, etc. (Trudgill and Hanan 2001: 101-112). Finally, RP is considered a practical accent for EFL learners to achieve successful communication (Collins and Mees 1981).

Consonants: The first language of the subjects is Arabic, a language with at least 28 consonantal sounds. These are the obstruents /b, t, d, k, f, s, z, n, m, ʃ, ð, θ, dʒ/, approximants /w, j/, trill /r/, and the back consonants glottal /ʔ, h/, velar /ɣ, x, k/, uvular/q/ and pharyngeal /ħ, ʕ/, plus the emphatic stops and fricatives /tˤ, dˤ, ðˤ, zˤ, sˤ/ (Huthaily 2003, Allan 1997: 188-189, Laufer 1988: 1197-1198, Amayreh and Dyson 1998). Important information is that /g/ is not part of the Arabic consonant inventory, but in Sudanese Arabic (SA) the uvular /q/ is always replaced by /g/. Moreover, the /g/ sound is often used by Bedouins in the place of /q/, which reveals that the latter is the original phoneme (Karouri 1996:27-30). English, the target language, has 24 consonants /p, b, t, d, k, g, f, s, z, n, m, ʃ, ð, l, w, v, dʒ, ʒ, ŋ/ and an approximant /r/. In principle, some kind of similarities exist between English and Arabic consonants where some sounds are shared (Suhana 2001), e.g. /s, n, t, d, k, z, b, etc./. However, many of the English and Arabic consonants show categorical phonemic differences in the place and manner of articulation, context, and acoustic features of the phonemes which may hinder the perception of L2 consonant sounds. In this way, it often becomes difficult to make a clear division between similar consonant sounds that can result in positive transfer, and those which are phonologically marked differently and can cause

negative transfer. These factors are expected to make the perception of English consonants more difficult for our listeners.

Clusters: Initial two (CCVC) and three-segment (CCCVC) clusters are common in English but do not have corresponding equivalents in Arabic. Arabic language has a syllable system that usually follows the CVCV pattern which does not permit two consecutive consonants nor four consecutive vowels (Nwesri et al. 2006). For instance, /pr, pl, gr, θr, θw, sp/, and three-segment initial consonant clusters such as /spr, skr, str, spl/ are entirely absent in Arabic. Furthermore, in contrast to Arabic which has no words ending in two or three-segment clusters, English has 78 three-segment clusters and fourteen four-segment clusters occurring at the end of words. Consonant clusters of English predominate in word final position which is attributed to the addition of the [s, z, t, d] morphs that indicate tense and number. Furthermore, the three-element clusters are considered the most complex type of consonantal onsets permitted in English due to their linguistic structure, which has been found to contribute to unintelligibility (McLeod, Doon and Reed 2001, Gierut and Champion 2001). These factors, combined, make Arabic-speakers learning English face a challenge with the perception of consonant clusters.

## **Method**

Intelligibility tests used: Intelligible speech is defined as speech that is understood by native speakers (Munro et al. 2006: 112-114). This means that speech intelligibility is principally a hearer-based construct that depends on interaction in an appropriate context involving the comprehension of the message between the listener and the speaker. It is also possible to refer to speech intelligibility as any successful communication that involves both native and non-native speakers of English, because the final goal of such speech is understandability. Since listeners of this study are expected to have an incorrect conception of English speech sounds, focus will be on examining vowels, consonants, and consonant clusters, in part, because they form the basic sound knowledge of the English language, the mastery of which is required for perfect learning of speech. And second, because the assessment of whether speech is intelligible or not is attributed to segmental factors, more than 50% of speech intelligibility is accounted for on the basis of speech sounds (Pascoe 2005: 5-6, Luchini 2005 ).

The Rhyme Modified Test (MRT) was used in the experiments. The MRT is considered to be the most accurate and reliable measure of intelligibility (Logan, Greene, and Pisoni 1989). Speech intelligibility measures involve word identification tasks in a closed-set of four items, where the listeners are asked to select the response they think the speaker intended. The score is the number of correctly responded to items. Test items normally target phonemes, multi-phonemes, or words. Phonemes refer to vowels and single consonants, whilst multi-phonemes refer to cluster consonants. The formal assessments of phonemes and multi-phonemes interpret the responses as either intelligible or unintelligible; put in figures, a score of (close to) 100% is interpreted as completely intelligible performance (Lafon 1966). Word intelligibility, on the other hand, was determined on the basis of final words embedded in short redundant SPIN sentences. SPIN is an abbreviation of ‘Speech Perception in Noise’ Test (Kalikow, Stevens and Elliott 1977, Wang 2007, Wang and Van Heuven 2007). It is a perception test that measures listeners perception abilities. Measurement is based on a recognition task of twenty-five words embedded in meaningful and highly predictable sentences, as in She wore her broken arm in a sling (target word underlined). Listeners write down the final word that they think they heard in each sentence. This part of the SPIN test proved to be efficient at assessing speech recognition abilities (Rhebergen and Versfeld 2005). Although the listeners’ performance is primarily quantified in terms of number of whole words correctly recognized, partially correct answers are also important since they give information about the perception of phonemes in onset, nucleus and coda position.

L2 listeners: The subjects of the study were ten Sudanese university English students in the Department of English at El Gadarif University in the Sudan. The subjects involved in these experiments specialized in English language teaching (TEFL). They had studied for six semesters when they participated in the listening test. During the period of study, which extends for four years, students attended three courses in the field of pronunciation; these are (i) an introduction to phonetics, (ii) phonology, and (iii) practical phonetics, delivered in three consequent semesters. They also attended two classes on English listening skills, which usually take place in semesters one and three. English is treated as a foreign language (not a second language), the learning of which starts in the fifth year of primary school and continues at secondary schools for three years. English lessons obtained during these stages vary between 5 and 6 hours per week; English is

treated as a school subject that provides basic principles of the language in a traditional way of language teaching.

Overall structure of the test battery: The experimental stimuli include four tests. These are (i) a vowel test which is composed of minimal quartets including short and long vowels as well as diphthongs, (ii) single consonants in either onset or coda position and (iii) consonant clusters in onset or coda position. These target sounds were embedded in meaningful C\*VC\* words (where C\* stands for one to three consonants). (iv) The fourth test comprised 25 sentences taken from the high-predictability set included in the SPIN (Speech Perception in Noise) test (Kalikow et al. 1977). These are short everyday sentences in which the sentence-final target word is made highly predictable from the earlier words in the sentence, as in *She wore her broken arm in a sling* (target word underlined). Word stimuli in the first three tests were embedded in a fixed carrier sentence [say...again], which insured a fixed intonation with a rise-fall accent on the target word. The vowel and the single consonant tests contained items on each individual vowel or consonant phoneme in the R.P inventory.<sup>2</sup>

Moreover, the consonant test targeted all the consonants in onset position and in coda position. For the cluster test, the number of test items had to be limited as the total inventory of onset and coda clusters is very large; including all the clusters would have been too demanding on the subjects. Nine onset and eight coda clusters were selected that represent problems to Sudanese-Arabic learners of English (Allen 1997: 188-189, Patil 2006: 88-131).

All items in the tests were chosen such that they occurred in dense lexical neighborhoods, i.e. there should be many words in English that differ from the test item only in the target sounds. For instance, the vowel /ɪ/ was tested in the word *pit*, since the /p\_t/ consonant frame can also be filled in by many other vowels, as in *peat, pet, pat, pot, part, port, put, putt* and *pout*. These so-called lexical neighbors, differing from the target word in only the identity of the test sound, make up the pool of possible distracters (alternatives) in the construction of the MRT test. When selecting the three distracters needed for each test items we preferably selected lexical neighbors that differ from the target in only one distinctive feature. For the target *pit*, we selected alternatives with vowels that differed from /ɪ/ in just one vowel feature, i.e. *pet* (differing in

height), *put* (differing in backness) and *pot*. The latter alternative differs from the target in both height and backness; we preferred this to the one-feature difference in *peat* (or *Pete*) as we decided to exclude proper names and low-frequency alternatives as much as possible which may show a larger decrement in recognition than high-frequency words. The full set of test items is included in the Appendix.

Tests materials: The stimulus sentences were typed on sheets of paper (one sheet for each test) and then read by a male native speaker of R.P. English. Recordings took place in a sound-treated room. The speaker's voice was digitally recorded (44.1 KHz, 16 bits) through a high-quality swan-neck Sennheiser HSP4 microphone. The speaker was instructed to inhale before uttering the next sentence so that clear recording is achieved. The target words were excerpted from their spoken context using a high-resolution digital waveform editor Praat (Boersma and Weenink 1996). Target words were cut at zero-crossings to avoid clicks at onset and offset. Target words and SPIN sentences were then recorded onto Audio CD in seven tracks. The first track contained two practice trials for the vowel test, and was followed by track 2 which contained the 19 test vowel items. Tracks 3 and 4 contained the practice and test trials for the single consonant tests, and tracks 5 and 6 contained the cluster items. Track 7 comprised the 25 SPIN sentences with no practice items. In the single consonant and cluster tests trials targeting onsets preceded the items targeting codas. Other than that, the order of the trials within each part of the test battery was random. Trials were separated by a 5-second silent interval. After every tenth trial a short beep was recorded, to help the listeners keep track on their answer sheets.

Test procedure: The stimuli were presented over loudspeakers in a small classroom that seated ten listeners. Subjects were given standardized written instructions and received a set of answer sheets that listed four alternatives for each test item. They were instructed for each trial to decide which of the four possibilities listed on their answer sheet they had just heard on the CD. They had to tick exactly one box for each trial and were told to gamble in case of doubt. Alternatives were listed in conventional English orthography. In the final test (SPIN), subjects were instructed to write down only the last word of each sentence that was presented to them. There were short breaks between tests and between presenting the practice items and test trials. Subjects could ask for clarification during these breaks in case the written instructions were not clear to them.

We will now present the results of the test battery in four sections, one for each test. Each section will first outline the structural differences between the sounds in the source language (Sudanese Arabic - SA) and in the target language (RP English). Such comparisons may help understand why certain English sounds are difficult for Sudanese learners and others are not.

### **Results and discussion**

In this part, I present the results and the discussion of four sections separately which include vowels, consonants, clusters and SPIN sentences of English.

English Vowels Results: Figure 1 shows the rates of vowel perception errors made by the Sudanese listeners. It provides means, and standard deviations of the whole performance of the subjects concerned.

As it appears from the figure, listeners show a complete failure in the recognition of the short vowel /ʌ/ and the long vowel /ɑ:/. These are followed by high rate of misperception of the lax/short English vowels /ɪ/ and /ʊ/, /e/ and /ɔ/. Similarly, tense /long vowels /ɜ:/, u:/, and diphthongs like /eə/, /uə/, /eɪ/, /aɪ/, /ɪə/ and /aʊ/ also proved to be problematic. However, listeners show no errors in perceiving the two diphthongs /ɔɪ/ and /ɔ:/, while there are few errors committed in the perception of the short vowel /æ/.

Furthermore, table 1 enables us to view the picture more clearly. It shows the confusion matrix of the correct responses, and the areas confused by the Sudanese listeners in the perception of English vowels. The diagonal line running across the table contains the correct scores whilst the spots scattered around it represent the problem areas.

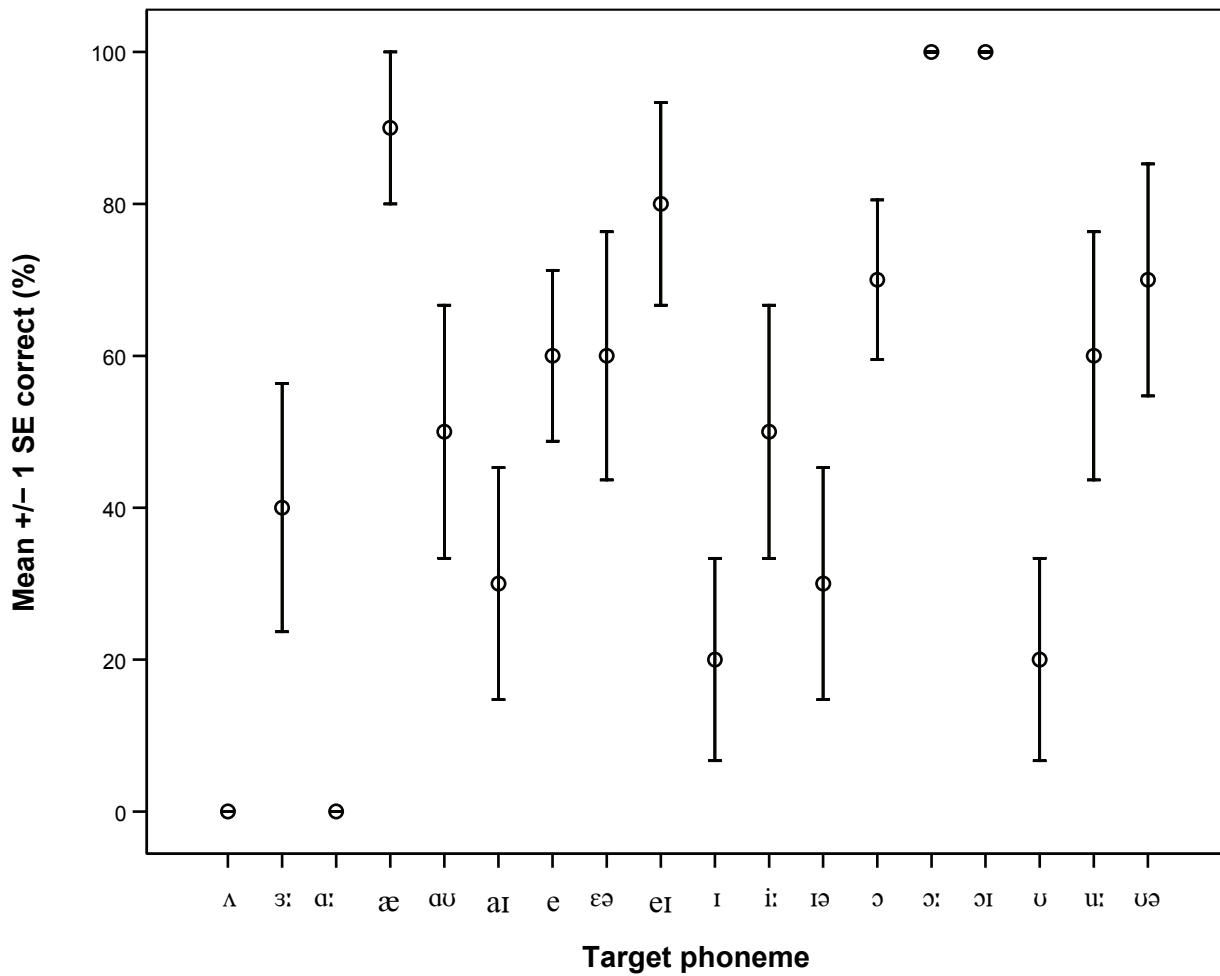


Figure 1. Percentage of English vowels correctly identified by ten Sudanese listeners. Error bars represent +/-1 Standard Error of the mean.

Table 1. Confusion matrix of 20 English stimulus vowels and diphthongs (in the rows) perceived by ten Sudanese-Arabic listeners (in the columns). Correct responses are on the main diagonal, indicated in bold face. Confusions ( $\geq 30\%$ ) are in grey-shaded cells. The vowel /əʊ/ should have been presented but was not.

Target	Responses																		Total	
	ʌ	ɜ:	ɑ:	æ	aʊ	aɪ	e	ɛə	eɪ	ɪ	i:	ɪə	ɔ	ɔ:	əʊ	ɔɪ	ʊ	u:		ʊə
ʌ	<b>0</b>						1						9							10
ɜ:		<b>4</b>	1											3						10
ɑ:			<b>0</b>											9						10
æ				<b>9</b>			1													10
aʊ					<b>5</b>				1						4					10
aɪ						<b>3</b>				5		2								10
e							<b>5</b>				1								1	10
ɛə		2							6											10
eɪ						1		1		<b>8</b>										10
ɪ							5				2	3								10
i:								1			4	5								10
ɪə									7				3							10
ɔ	3													7						10
ɔ:														<b>10</b>						10
əʊ																				0
ɔɪ																<b>10</b>				10
ʊ																	2	8		10
u:																	4	6		10
ʊə																			7	10
<b>Total</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>9</b>	<b>5</b>	<b>3</b>	<b>5</b>	<b>6</b>	<b>8</b>	<b>2</b>	<b>5</b>	<b>3</b>	<b>7</b>	<b>10</b>	<b>4</b>	<b>10</b>	<b>2</b>	<b>6</b>	<b>7</b>	<b>180</b>

### Discussion

The perception of the English vowels forms a serious problem for Sudanese Arabic listeners of this study. The listeners frequently confused the low central short vowel /ʌ/ for the peripheral low and back short vowel /ɔ/, whilst half open vowel /ɜ:/ was identified as /ɔ:/ because their L1(Arabic) inventory lacks central vowels (Brett 2004:103-133). As a matter of fact, the linguistic differences that exist between the listeners' L1 and L2 have a negative transfer

(mapping model, cf. Kuhl 2000: 99-115) on the listeners' perception process. That is, listeners are not familiar with the type of vowels needed in English because they are not distinguished in the Arabic phoneme system. Therefore they tend to adapt L2 vowel sounds to their L1 which causes perceptual problems.

A similar case reported by Tomokiyo, Black and Lenzo (2003: 1-4) describes difficulty to achieve inter-coder agreement between Arabic and English vowels, especially the presence of an /e/ or /o/ vowel is not easy for the Arabic listeners to identify with a great deal of consistency. They refer this to the influence of MSA, where formal methods (i.e. the writing system) indicate the existence of only /a/ , /i:/ and /u/. More importantly, duration often has a negative influence on the recognition of English vowels. This appears in several cases where the Sudanese listeners conflated /ʊ/ for /u:/ and /ɪ/ for /i:/, and confused /ɑ:/ for /ɔ:/. Such a type of error motivates the hypothesis that durations are important acoustic cues used in cross-linguistics of speech perception (Hillenbrand and Clark 2000: 3014-3022). According to Hillenbrand and Clark, due to duration shortening the vowel /æ/ tends to be heard as /ɛ/, and /ɑ/ as /ɔ/, whilst the lengthened /ɛ/ tends to shift to /ʌ/, and /ʌ/ as /ɑ/, or /ɔ/ a change process which leads to confusion. However, Hillenbrand and Clark observed slight alterations in the perception of /ʊ/, /u:/, and /ɪ/, /i:/ due to duration effect. A more specific case was reported by Munro (1993) that the English vowels interpreted by Arabic groups (including Sudanese) manifested the same ordering of vowel duration differences for front vowels, but different ordering for back ones. This is due to interference of L1 (Arabic), a quantity language where length is an intrinsic element that requires vowels to be realized as short/long (geminated). Thus, our subjects incorrectly interpret English tense-lax vowels in terms of Arabic long/short vowel categories. This data raises the prediction that English tense-lax vowels are close to Arabic long/short vowels in terms of quality and duration.

Moreover, it is possible to refer such perception errors to the inadequate knowledge of English vowels which motivates listeners to conflate, guess, or fall back on their L1 norms (Fokes and Bond 1995, Flege and Font 1980, Walker 2001: 1-6). It is also probable that because Sudanese listeners descend from a language background with a small number of vowels, they find the

perception of the English vowels difficult. According to Cruttenden (2001: 99-112) this is most predictable in those areas where vowels are close together in the vowel space, thus confusions are potential within these areas; [ɪ, i:], [ʊ, u:], [e, æ, ʌ], and [ɔ:, ɒ, ɑ:]. Incidentally, compared to the previously discussed levels, there are very few confusions on the level of diphthongs. The diphthong /aʊ/ is misidentified as /əʊ/, /ɪə/ as /eə/ and /aɪ/ as /eɪ/. Misidentification of such English vowels can be attributed to the fact that each two confused diphthongs share at least one sub-phoneme; a feature which serves to complicate the perception task for listeners. It seems as though the complete absence of such diphthongs in the listeners' L1 phonological system, may have helped them to achieve a better understanding.

Onset consonants results: Figure 2 shows the results of the perception test of ten Sudanese listener group on English consonants.

The results reveal that an overall identification of the onset is better than that of coda consonants (see table 3 ). On the onset level, listeners show near-perfect perception of stops /b/, /t/, /d/, and /k/, and the fricatives /f, v, s, ʃ/ as well as /m, n, h, y/. However, a few errors were made in the identification of voiceless labio-dental /p/ and voiced velar /g/. Listeners also substituted /g/ for /k/, which are produced at the same place of articulation (velar), and /dʒ/ for /d/. Other errors occurred in the recognition of the voiceless fricatives /θ/ and the voiced /z/. Here listeners confused the voiced /z/ for voiceless /s/, /θ/ for /s/, whilst /p/ was used for /tʃ/. An interesting finding is that listeners were observed to frequently perceive the retroflex /ɾ/ as /w/.

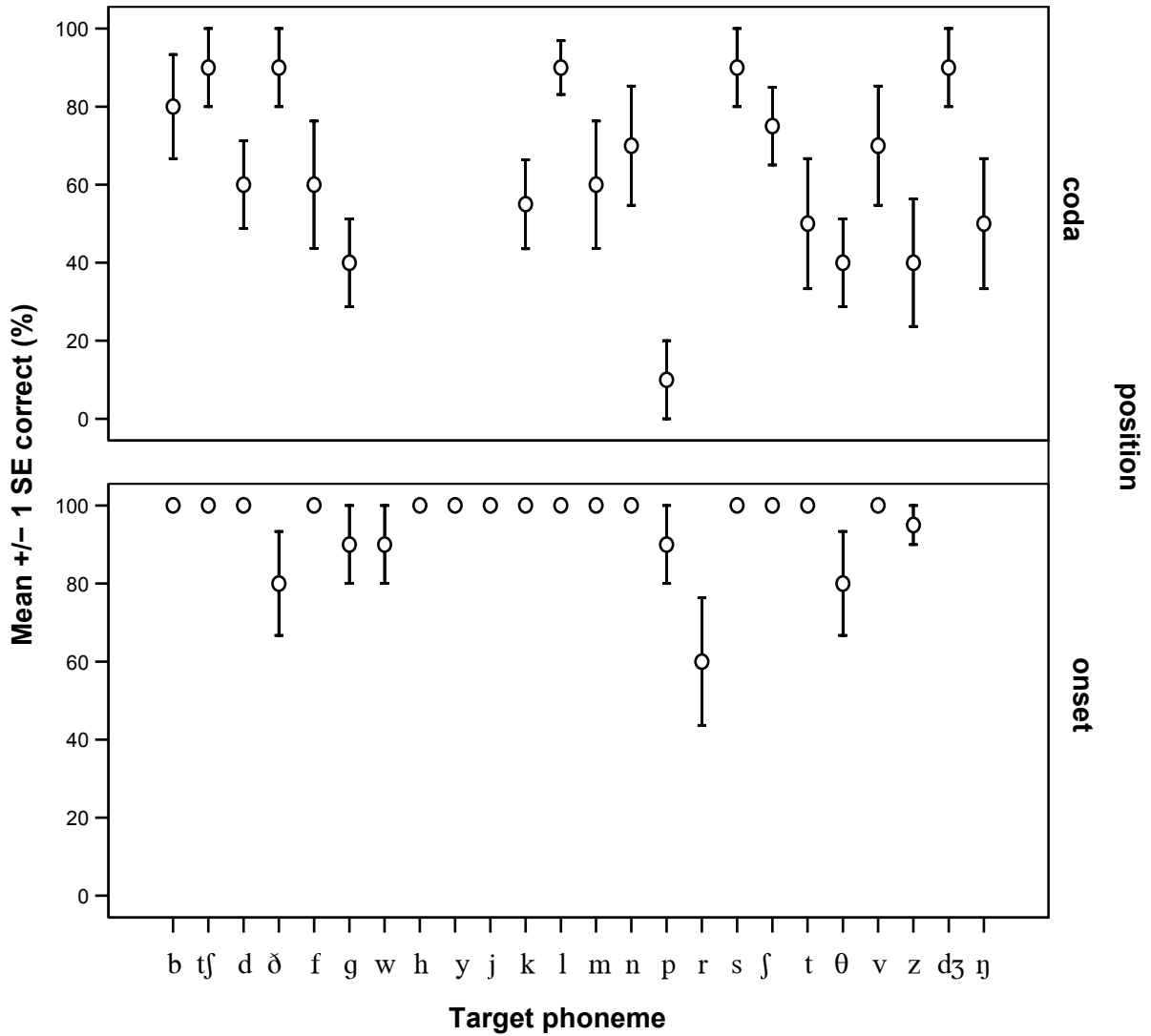


Figure 2. Correctly identified percentage of a perception test of English onset and coda consonants. The test was executed by ten Sudanese listeners. Error bars are +/- 1 Standard error.

Table 2 presents the Sudanese listeners' perception of English onset consonants in more detail. The diagonal line running across the table displays the correct scores of perception while the scores scattered around it represent the problem areas.

Table 2. Confusion matrix of 19 English stimulus onset consonants (in the rows) perceived by ten Sudanese-Arabic listeners (in the columns). Further see Table 1.

Target	Responses																				Total		
	b	tʃ	d	dʒ	f	g	h	j	k	l	m	n	p	r	s	ʃ	t	θ	v	w		y	z
b	10																						10
tʃ		9															1						10
d			10																				10
dʒ			2	8																			10
f					10																		10
g						9				1													10
h							10																10
j								10															10
k									10														10
l										10													10
m											10												10
n												10											10
p		1											9										10
r														6							4		10
s															10								10
ʃ																10							10
t																	10						10
θ													1		1			8					10
v																			10				10
w										1											9		10
y																						10	10
z															1								9
	10	9	8	8	10	9	10	10	10	9	10	10	9	6	10	10	10	8	10	9	10	9	220

Coda consonants results: Compared to onset consonants, results in Figure 2 show that more errors are made by the listeners in the perception of the coda consonants; the overall mean percentage of correctly identified consonants is poorer for codas than for onsets. A confusion of 90% was made in the recognition of the voiceless stop /p/ as /d/, /k/ and /n/. Listeners also made errors in the perception of /g/; i.e., they confused /g/ for /k/ and /g/ for /n/. Conversely, they

confused /k/ for /g/ and /k/ for /t/, whilst /t/ was misidentified as /d/ often as /k/. Nasal codas proved to be a problematic area of perception where the confusion rate ranged between 50% and 60%. For example, listeners frequently confused /ŋ/ and /m/ for /n/. On the other hand, labiodental /f/ was confused with /v/, and /v/ with /z/. Listeners show very few errors in identifying /b/, /s/, and /tʃ/, while they made no errors in the perception of /l/, /ʃ/, and /dʒ/. The confusion matrix of coda consonant perceptions is presented in table 3. In the table the plosives /p, t, d, k, g/ appear more problematic, whilst /l, ʃ, dʒ/ were perfectly perceived.

Table 3. Confusion matrix of 19 English stimulus coda consonants (in the rows) perceived by ten Sudanese-Arabic listeners (in the columns). Further see Table 1.

Target	Responses																		Total	
	b	tʃ	d	dʒ	f	g	k	l	m	n	ŋ	p	s	ʃ	t	θ	v	z		ð
b	8						1								1					10
tʃ		9												1						10
d			6			1									3					10
dʒ				10																10
f					6												4			10
g						5	3			2										10
k						1	6								2					10
l								10												10
m									6	4										10
n										7										10
ŋ			1							3	5									10
p			3				5			1		1								10
s													9		1					10
ʃ														10						10
t			2				1								5					10
θ																4	6			10
v																	7	3		10
z													6					4		10
ð								1											9	10
Total	8	9	6	10	6	5	6	10	6	7	5	1	9	10	5	4	7	4	9	200

## Discussion

One of our findings is that the Sudanese listeners confused English /r/ and /w/. This problem supports the claim that the learners' production of L1 sounds influences the way they perceive L2 counterpart. That is, it is very likely that the English /r/, which is not a trill but a frictionless continuant, is mistaken for the nearest vowel-like sound in Arabic, which would be /w/. There are strong indications that /w/ is perceptually close to English /r/. There is a sound change in progress in which young speakers of English now pronounce onset /r/ as /w/ (see Watt, Docherty and Foulkes 2003). In the majority of English accents /r/ is articulated as a voiced alveolar or post-alveolar approximant. The retroflex variant of /r/ is distinguished by a particularly low F3 that is close to F2, while energy above F3 is normally weak due to the existence of two anterior constrictions in the vocal tract, one made by the tip or blade of the tongue, and the other by the narrowed lip. The Arabic /r/, on the other hand, is normally a tap or an alveolar trill that requires alternative vibration of the tongue against the ridge. Allophonic variation is mainly concerned with the distinction between single and geminate /r/ in intervocalic position, whereby single /r/ is produced as a tap, and geminates as trills (as they are in Spanish). Because of these phonemic and acoustic differences, the substitution of /w/ for /r/ can occasionally occur (Khattab 2002). It is possible also to attribute such type of problem to the learners' lack of knowledge and to insufficient practice of the English [r] as a post-alveolar approximant.

On the other hand, the replacement of /g/ by /k/, /z/ by /s/, and /θ/ by /s/ shows a systematic pattern of errors. The first two errors are a shift of voiced to voiceless. These cases are produced at the same place of articulation; the sounds /g/ and /k/ are velar, while /z/ and /s/ are alveolar. It is most probable that the errors of perception /g, k/ and /z, s/ are the result of the effect of similarity of the place of articulation. Although it is possible to suggest that such problems can be interpreted as a violation of the norm of the voiced/voiceless feature; e.g. when [k] and [g] are confused, it is not just because they are both velar stops but because the voicing feature is not distinguished, or resists learning. However, Flege and Font (1981) attribute this type of error in English stops to the place of articulation rather than to voicing. Additionally, the confusion of /θ/

for /s/ is probably caused by interference of the perceptual strategies of the listener's L1 where the English (inter)dental /θ/ was mistaken for the nearest Arabic sound, which is the (alveolar) dental /s/. The substitution of /ð/ for /z/, and /θ/ for /s/ is often attributed to the L1 effect. That is, in the consonant inventory of Sudanese and other Arabic dialects, the interdental /θ, ð/ merged with the apico-dental (often labeled as alveolar or sibilant) /s, z/ (Dickins 2007: 23-27, Karouri 1996: 60-68, Janet 2002:13-20, Corriente1978: 50-55). Thus, Arabic words like /hæða/ 'this', is pronounced as [hæza], whilst /θæbit/ firm is pronounced as [sæbit], a problem which is reflected in the perception of L2 speech sounds. The affricate /tʃ/ was also misperceived as /p/ because the articulation of the two stops /t/ and /p/ involves a complete closure followed by a release. This makes listeners think of affricates as stops with a slow fricative release. It is very common among L2 interlocutors that when there is background noise or unfamiliarity with the speaker's accent, intelligibility is compromised (Ball and Rahilly 1999: 178-179, Subramaniam and Ramachandran 2006: 28-33).

In comparison to the onset, the perception of the coda consonants proved to be difficult for the Sudanese listeners. The listeners made more errors in the perception of voiceless stop /p/, which was substituted for /k/, /d/, and /n/. They also substituted /t/ for /d/ and /k/. This can be attributed to several factors. First, the sameness of the manner of articulation of such sounds; i.e., the sudden burst required in producing /p, k/, and /t, k/ makes such phonemes sound similar. When all acoustic correlates of L2 are not easy to pick up, listeners are forced to guess the identity of a stop; consequently they will choose the nearest place of articulation, or sound features that are relevant to the intelligibility of their native language which compromises recognition accuracy (Gimson 1989: 19-20). Second, the differences that exist between Arabic and English in both the phonetic detail specifying the voicing contrast, and the stop inventory, add to problems. In Arabic the voiceless stops are aspirated, while there is pre-voicing for syllable initial stops. English stops, on the contrary, exhibit a voicing contrast at all points of articulation; bilabial, alveolar, and velar. These differences function as sufficient cues for the distinction between the stops. Regardless of such differences, in perceiving the English stops particularly in cases like /p, d/ and

/t, d/ the Sudanese listeners use the acoustic correlates of Arabic stops instead which trigger the confusion. This type of error of English stops is described as a wrong approximation of the length of the vowel duration that should precede or follow such stops. To avoid these problems, Arabic speakers learning English need to do a modification in their L1 correlates of voiced and voiceless stops towards the English norm (Fokes et al. 1985: 81-84, Khattab 2000). They need to use a longer VOT values for initial voiceless and to lengthen the vowel preceding the syllable-final voiced stops/obstruent. Other perception errors are that the Sudanese listeners confused the voiceless coda consonants with their voiced counterparts as in /s and z/, and /f and v/ as a result of the similarity in the place of articulation, whilst the confusion of [n, ʃ and m] is due to nasality. Many types of errors of perception are the result of similarity of the place and manner of articulation, on both onset, and coda level. The absence of some phonemes like /v, ʃ and p/, etc., from the Arabic inventory adds up to the perception problems of listeners.

### **Consonant Onset and Coda Cluster Results**

Figure 3 shows means (and standard error) for a group of ten Sudanese listeners in the perception of English consonant clusters. As the figure shows, in contrast to vowels, consonant clusters yield fewer errors of perception. Furthermore, the performance of the listeners for onset clusters is better than for coda clusters; the overall correct scores being 75 and 71%, respectively.

Listeners misrecognized /dr/ as /gr/ which is more frequent than /dr/ as /kl/, and these are followed by the misidentification of /sl / as /sn/. They are also observed to interchangeably make errors in perceiving /spl/ as /spr/, /kl/ as /gr/, and /spr/ as /pr/ or /skw/. However, there are no errors shown in the perception of the initial clusters /gl/, /pl/, and /sw/. On the other hand, final clusters are more prone to misperception. That is, the rates of errors of perception shown in figure (3) indicate that the most perception errors manifest on the coda level; and these are the substituted /bd/ for /ld/, /st/ for /sk/, /nz/ for /mz/, and /nz/ for /dz/. Listeners also made errors in identifying /lm/, /ts/, /nt/ and /mp/, but fewer errors were observed in recognizing the item /gz/, whilst /ŋk/ was correctly recognized. More details are shown in Tables 4 and 5 below. They provide a clearer picture of the correct and confused consonant clusters. The correct scores of

perception appear on the diagonal line running across the table in bold face, while the cells scattering around represent the confusion areas.

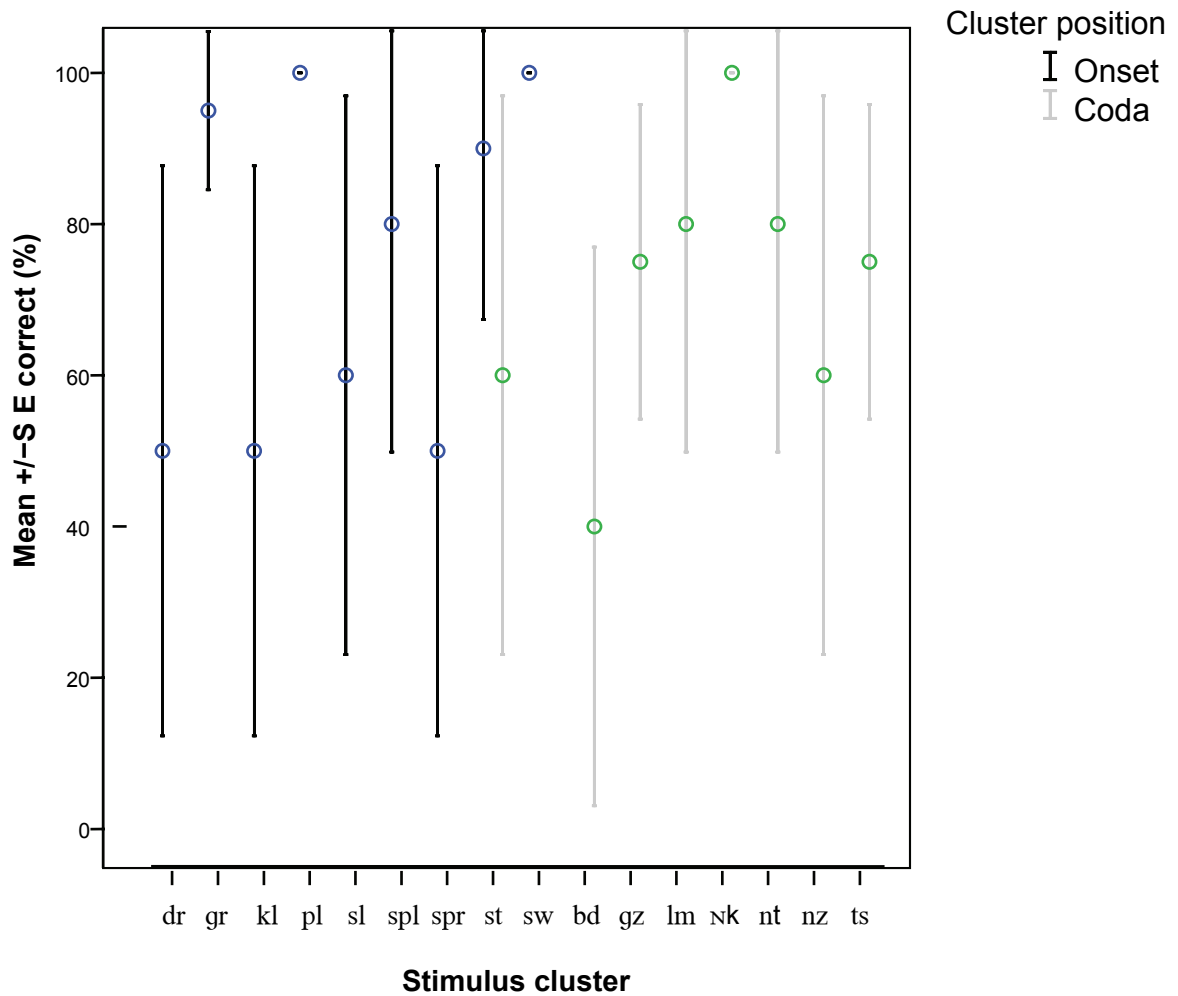


Figure 3. Mean percentage of English onset and coda clusters correctly identified by ten Sudanese listeners. Error bars are +/- 1 Standard error.

Table 4. Confusion matrix of 8 English stimulus onset consonant clusters (in the rows) perceived by ten Sudanese-Arabic listeners (in the columns). Further see Table 1.

Target Onset	Responses																Total	
	dr	gl	kl	pl	sl	spl	spr	st	sw	gr	kr	kw	pr	sk	skw	sm		sn
dr	5									4	1							10
gl		10																10
kl			5							2		3						10
pl				10														10
sl					6											1	3	10
spl						8	2											10
spr						3	5						1		1			10
st								9						1				10
sw									10									10
Total	5	10	5	10	6	8	5	9	10	6	1	3	1	1	1	1	3	90

Table 4. Confusion matrix of 8 English stimulus coda consonant clusters. Further see Table 4.

Target Coda	Responses																	Total		
	bd	gz	lm	ŋk	nt	nz	st	ts	dʒ	dʒ	lb	ld	lk	ls	lt	mp	mz		sk	zd
bd	4											6								10
gz		8									1								1	10
lm			8										2							10
ŋk				10																10
nt					8									1	1					10
nz						6				1							3			10
st							6											4		10
ts								7	2					1						10
Total	4	8	8	10	8	6	6	7	2	1	1	6	2	1	1	1	3	4	1	80

## Discussion

The plosive/liquid replacement of /dr/ by /gr/, and the fricative/plosive /st/ by /sk/ can be accounted for as an alveolar-to-velar shift within the same manner of articulation. The misperception of /kl/ as /gr/ (velar+liquid) can be referred to the factor of velarity in the first cluster members, and to the manners of articulation in the second. Generally speaking, these types of perception errors motivate the linguistic hypothesis that the perception of L2 sounds is often influenced by the perceptual and articulatory properties of L1 (Cruttenden 2001: 20-25, Canepari 2005: 38) where listeners often resort to the nearest corresponding sound. Moreover, such a type of perception error where voiced obstruent precedes the voiced liquid /r/ often takes place due to phonological alternations in similar consonant clusters-- mostly in homorganic C/liquid sequences. These phonological alternations usually occur when the speech signal is not detected well due to the lack of experience with voicing leads in phonetically voiced stops, or due to the absence of appropriate phonetic cues (Seo 2003: 20-59). Similar interpretations apply to the misperception of the voiceless sibilant/voiceless stops/liquid clusters /spl/ as /spr/ interchangeably, and the misperception of /spr/ as /pr/ and /skw/, where substitution errors of the third cluster member /l, r, w/ took place, respectively. However, this type of error points also to the influence of the similarity of the manner of articulation shared by such approximants. On the other hand, the confusion of the coda nasal/fricative clusters /nz/ as /mz/ is due to nasality, but the confusion of nasal/plosive clusters /nz/ for /dz/ is probably due to the influence of the place of articulation shared by such members. Additionally, listeners follow a repair strategy in perceiving /nt/ as /it/ (it is not a cluster member), and /bd/ as /ld/.<sup>3</sup> They adopt the nearest speech sound that aids them to understand a word/message; i.e. listeners transfer their L1 phonotactic constraints when listening to English. This strategy reflects the prominent role played by the Sonority Principle Sequence in accounting for phonotactic patterns across languages (Carr 1999: 14-39, Clements and Keyser 1988, Gierut 1999, Gierut and Champion 2001). Thus, the nasal/liquid, obstruent/liquid clusters of homorganic sequences and similar voiceless sibilant plus voiceless plosives etc. are more vulnerable to phonological change than those in heterorganic sequences.

### Results of sentence (SPIN) test

Background: The SPIN-test (Speech Perception in Noise test) targets word recognition at the sentence level. It aims to examine the learners' performance in speech perception by including the effect of semantic context. In the SPIN test listeners are exposed to a set of 25 specific meaningful sentences. Their task is to write down the last word embedded in each sentence. In this way, the final goal of such types of test is to provide a measure of the ability of a listener to understand speech in an everyday listening situation.

Figure 4 provides the means of Sudanese listeners' perception on the SPIN test.

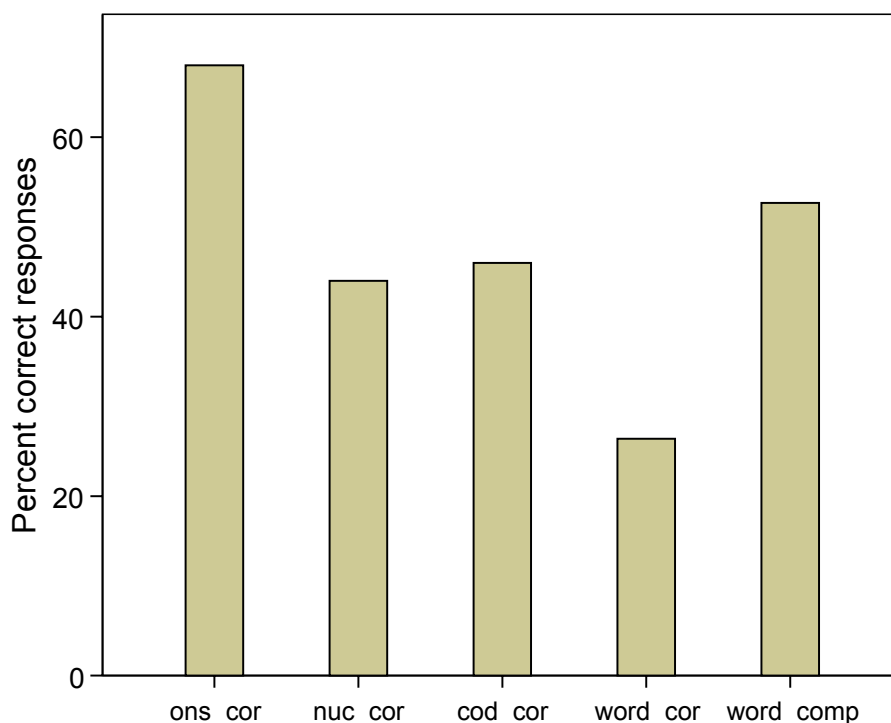


Figure 4. Percentage of (parts of) English words correctly recognized by Sudanese-Arabic listeners (further see text).

Correct perception of complete keywords ('word\_cor' in Figure 4) proved to be very difficult for listeners; scores are around 30% correct. However, listeners often managed to recognize some sounds in the words correctly. For instance, correct identification of sounds in the onset position

of syllables ('ons\_cor') is at 70%, whilst vowels ('nuc\_cor') and coda consonants ('cod\_cor') are around 45% correct. The mean of the component identification ('word\_comp') is about 50%. The observation that onsets were perceived more accurately than the vowels and codas ties in with the more detailed results of the MRT tests. Together, these results indicate that onsets consonants, whether single or clusters, were identified more successfully than vowels and codas.

## **Discussion**

The Sudanese listeners had a poor perception in simple, and predictable English sentences that reached 30% correct. However, they had a better performance on single and cluster consonants and were poor especially on vowel level. These observations provide empirical evidence that words and vowels are the most problematic aspects for our listeners. We predicted that vowel perception would be more of a challenge for Sudanese-Arabic listeners of English than single and cluster consonants. This is probably because their L1 has only five or six vowels which makes it difficult for them to attain the vowel system of any variety of English (Cruttenden 2001: 99-112). Moreover, observations bear out our prediction that the large number of consonant sounds existing in the listener's L1 facilitated the perception task (positive transfer); i.e. listeners are at least more familiar with consonants than vowels.

Table 6 presents the results of the correlations coefficient of vowels, single consonants, cluster consonants of English and SPIN sentences of ten Sudanese listeners. It shows the linear relation between the listeners perception scores at the four levels. In the table, the vowels, consonants and cluster components are shown in the upper part, whilst the SPIN sentences components are in the lower part of the table.

Table 6. Correlation matrix of scores on vowels, single consonants, cluster consonants and SPIN sentences. R-values indicate the linear relation between the listeners' perception scores at the four levels.

Items	Single Cons			Vow	Clusters			SPIN sentences			
	Onsets	Codas	Both		Ons	Cod	Both	Ons	Vow	Cod	Word
Codas	.591										
Both consonants	<b>.782</b>	<b>.965</b>									
Vowels	<b>-.682</b>	-.169	-.353								
Onset clusters	-.327	-.208	-.267	.312							
Coda clusters	-.391	-.057	-.172	.164	-.020						
Both clusters	-.505	-.152	-.282	.297	.470	<b>.873</b>					
ons_SPIN	-.135	.057	.000	.507	-.308	-.073	-.215				
vow_SPIN	-.209	.288	.154	.435	-.343	.227	.033	<b>.710</b>			
cod_SPIN	.288	.533	.505	.093	-.330	-.234	-.367	<b>.639</b>	.597		
word_SPIN	.194	.584	.514	.214	-.381	.070	-.124	.567	<b>.700</b>	<b>.899</b>	
comp_SPIN	.000	.327	.253	.386	-.370	-.064	-.237	<b>.908</b>	<b>.845</b>	<b>.866</b>	<b>.822</b>

Bolded  $|r| > .6$ : Correlation is significant at the 0.05 level (2-tailed).

Bolded  $|r| > .7$ : Correlation is significant at the 0.01 level (2-tailed).

As table 6 shows the correlation output of the four perception tests contributes to the issue under concern. The purpose of the correlation coefficient in this context is to provide information about matters such as how and to what extent the results of our subjects in the four sections relate to each other. The results show that there is a negative correlation of  $r = -.682$  ( $p < .05$ ) between vowels and onset consonants, indicating that poorer identification of vowels goes together with better results for onset consonants. On the other hand, vowels have high positive correlations with correct word identification ( $r = .700$ ,  $p < .01$ ). Positive correlations are also observed between word composition level and vowels  $r = .386$ , and consonants  $r = .253$ , which are not

significant. These data reveal that a weak relation of consonants identification with better word component results. Moreover, multiple regression predicting word scores (SPIN) from vowel, consonant, and cluster scores (MRT) is significant ( $R = .638$ ,  $p = .05$ ). These data suggest that an overall measure of performance on the vowels, consonants and clusters is a reasonable predictor of intelligibility at the sentence level.

To sum up, the perception of the listeners in the SPIN is very poor at the sentence level, but it provides feedback about which of the three types of English phonemes is most problematic for Sudanese listeners. In this connection, the results of the Sudanese listeners' correct word identification in the SPIN-test are comparable to those obtained for Mandarin Chinese listeners exposed to a similar SPIN test (Wang 2007: 70-71). Similarity of performance between the two groups can be attributed to the fact that both Chinese and Sudanese listeners speak English as a second/foreign language. The listeners also come from linguistic backgrounds that are entirely unrelated to English; Chinese is a Sino-Tibetan, whilst Arabic is a Semitic language. In contrast, Dutch listeners in Wang (2007) had high word correct percentage, due to more exposure to English than the non-Germanic groups. Furthermore, the Dutch L1 sounds are closer to the English targets either those of Arabic or Mandarin. Predictably, American listeners had the best performance on the SPIN test simply because they are native speakers of English (Wang 2007).

### **General conclusions**

Vowels proved to be a difficult area of perception for Sudanese listeners of English. This is most likely because they are unfamiliar with a large number of different types of vowel sounds present in the English language. Listeners found the perception of the English diphthongs, central, and back vowels the most problematic because such types of vowels are absent in their L1.




Durational aspects do not show serious effects on the identification of English vowels because there is some kind of correspondence between the listeners' L1 (Arabic) long/short vowel durations and those of English tense-lax vowels. However, the confusion within the tense-lax vowel pairs /ʊ, u:/ and less frequently /ɪ, i:/ indicates interference of the subjects' L1, and probably the lack of knowledge of English vowel sounds.

With regard to the interdependency existing between the perception and production of speech sounds, differences in the place and manners of articulation between English and Arabic phonetic systems require that the Sudanese listeners to enhance their L1 phoneme inventory to that of L2 so as to achieve a better performance of English speech.

The perception of the English single and cluster coda consonants is more difficult than that of the onset position. The listeners transfer their L1 phonotactic constraints when listening to English consonant clusters. This mostly occurs with coda consonants where the listeners fail to distinguish or implement certain phonetic features.

Conclusions drawn above provide cognitive insights that help us understand the nature and the causes of the speech perception problems, which are experienced by Sudanese listeners of English. Thus, they present useful guidelines that can contribute to the learning and teaching of such types of problems in ESL/EFL contexts. One important guideline is that successful pedagogical implications of speech perception should target the mastery of the basic principles of English phonology, phonetics, and acoustic cues. Many second/ foreign language learners lacking such knowledge, have difficulties treating English speech issues; e.g., recognizing English vowels in different contexts, or discriminating between quartets such as pit, pat, pot, put, etc. So, there is a need sometimes for pupil involvement in group work for task-based learning, whereby some pupils may have roles which require them to listen or speak quite a lot. Moreover, the listeners' L1 inventory has a real negative effect on the process of the speech intelligibility. This requires that it should be taken more seriously and more practically during the learning/ teaching tasks of English speech perception and production. The teachers, for example, need to create "English atmosphere" in the classroom where more exposure to native English speech is necessary to reduce the L1 effect.

## Notes

1. In the Arabic script, the harakat/diacritics are special unwritten marks which represent short vowels (a, i, u). The literal meaning of harakat is “movements”, e.g. in the context of moving air waves that we produce while pronouncing vowels as in the following examples:
  - i. Fathah; an oblique dash over a consonant like  represents "a" sounds
  - ii. Kasra sign; an oblique dash under a consonant like  represents "i" sounds
  - iii. Damma sign; a loop over ,resembles like comma, like  represents "u" sounds

In the Arabic language, diacritics are not part of Arabic alphabetic or ordinary spelling but understood from context (Hayat 2005: 29-33, Alan 1997:188-204, Chomsky and Halle 1968: 373-374). In the word structure of Arabic, they are sprinkled through the word rather than taking place as continuous segments, a characteristic that is clear in examples such as *darasa* ‘he studied’ and *hamala* ‘he carried’ where the vowels [a] are inflectional affixes marking tense, gender and number in a way that reveals the nature of Arabic non-concatenative morphological system “underlying” a deep phoneme regularities (Kenstowics 1994: 394-405).

2. Inadvertently, the vowel test did not include an item targeting the vowel /əʊ/ as in *boat*.
3. To achieve perceptible pronunciation or to facilitate perception and production of speech sounds, adult L2 learners are equipped with their L1 phonotactic constraints and have to deal with the mismatch that exists between L1 and L2. This process is referred to as a repair strategy (Kang and Hyunsook 2005: 407-419).

## References

- .Allen, J.S., & Miller, J.L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America*, 106 (4), 2031-2039.
- Amayreh, M.M., & Dyson, A.T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language & Hearing Research*, 41, 642-653.
- Ball, M.J., & Rahilly, J. (1999). *Phonetics: The science of speech*. Oxford University Press, New York.
- Boersma, P., & Weenink, D. (1996). Praat: a system for doing phonetics by computer. *Report of the Institute for Phonetic Sciences of the University of Amsterdam*, Vol. 132.
- Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL*, 16 (1), 103-113.
- Canpari, L. (2005). *A handbook of phonetics: Natural phonetics: Articulatory, auditory and function*. LINCOM, München.
- Carr, P. (1999). *An introduction: phonetics and phonology*. Blackwell, Oxford.
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. Harper & Row, New York, Evanston and London.
- Clements, G.N. & Keyser, S.J. (1988). From CV Phonology: A generative theory of the syllable. *Language*, 64 (1) 118-129.
- Collins, B., & Mees, I. (1981). *The sounds of English and Dutch*. Leiden University Press. The Hague, Boston, London.
- Corriente, F. (1978). D-L doublets in Classical Arabic as evidence of the process of de-lateralisation of DĀD and development of its standard reflex. *Journal of Semitic Studies*, 23 (1), 50-55.
- Cruttenden, A. (2001). *Gimson's Pronunciation of English*. Oxford University Press, New York.
- Dickins, J. (2007). *Sudanese Arabic: Phonemtics and syllable structure: Integrating consonants and vowels*. Otto Harrassowitz Verlag, Wiesbaden.
- do Val Barros, A.M. (2003). Pronunciation difficulties in the consonant system by Arabic speakers when learning English after puberty. Unpublished M.A. thesis, University of West Virginia.

- Flege, J.E., & Font, R. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30 (1), 117-134.
- Flege, J.E., & Font, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24 (2), 124-145.
- Fokes, J. Bond, Z.S., & Steinberg, M. (1985). Acquisition of the English voicing contrast by Arab children. *Language and Speech*, 28, 81-91.
- Frisch, S. (1996). Similarity and frequency in phonology. Ph.D. dissertation, Northwestern University, Evanston, IL.
- Giegerich, H.J. (1992). *An introduction to English phonology*. Cambridge University Press, Cambridge.
- Gierut, J. (1999). Syllable onsets: clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research*, 4, 708-726.
- Gierut, J., & Champion, A.H. (2001). Syllable onset II: Three-element clusters in phonology treatment. *Journal of Speech, Language, and Hearing Research*, 44 (4), 886-904.
- Gimson, A.G. (1989). *An introduction to pronunciation of English*. Cambridge University Press.
- Hayat, A. (2005). Transcribing Arabic phonemes. A preliminary attempt. *I-MAG*, 3, 29-34.
- Hillenbrand, J.M., & Clark, M.J. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108 (6), 3014–3022.
- Huthaily, Kh. (2003). Contrastive phonological analysis of Arabic and English. Unpublished M.A. thesis, The University of Montana.
- Hyman, L.M. (1975). *Phonology: Theory and analysis*. Holt, Rinehart & Winston, New York.
- Janet, E.W. (2002). *The phonology and morphology of Arabic*. Oxford University Press.
- Kang, H., & Yoon, K. (2005). Tense and lax distinction of English [s] in intervocalic position by Korean speakers: consonant/vowel ratio as a possible universal cue for consonant distinctions. *Studies in Phonetics, Phonology, and Morphology*, 11 (3), 407- 419.
- Kaye, A.S. (1997). Arabic and its relationship to the other Semitic languages. In A.S. Kaye (ed.) *Phonologies of Asia and Africa (including the Caucasus)*, Vol 1. Eisenbrauns, Winona Lake, IN, 188-204.
- Karouri, A.M. (1996). *Phonetics of classical Arabic: A selectional study of the problematic sounds*. Khartoum University Press, Khartoum, Sudan.

- Kawasaki, H. (1982). An acoustic basis for universal constraints on sound sequences. Ph.D. dissertation. University of California, Berkeley.
- Kawasaki, H. (1993). The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*. Longman, London.
- Khattab, Gh. (2002). /r/ Production in English and Arabic and monolingual speakers. *Leeds Working Papers in Linguistics and Phonetics*, 9, 91-129.
- Kenstowics, M.J. (1994). *Phonology of generative grammar*. Blackwell, Cambridge MA.
- Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97 (22), 11850-11857.
- Lafon, J.C. (1966). *The Phonetic test and the measurement of hearing*. Centrex, Eindhoven.
- Laufer, A. (1988). The emphatic and pharyngeal sounds in Hebrew and Arabic. *Language and Speech*, 31, 181-199.
- Logan, J.S., Beth, G., & Pisoni, D.B. (1989). Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America*, 86, 566-581.
- Luchini, P. (2005). Task-based pronunciation teaching: A state-of-the-art perspective. *Asian EFL Journal*, 7 (4), 191-202.
- McLeod, S., van Doon, J., & Reed, V.A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, 10, 99-110.
- Munro, J. M., Derwing, T.D., & Morton, L.S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Munro, J.M. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurement and accentedness ratings. *Language and Speech*, 36, 39-61.
- Nwesri, A.F.A., Tahaghoghi, S.M.M., & Scholer, F. (2006). Capturing out-of-vocabulary words in Arabic text. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia*, 258-266.
- Patil, Z.N. (2006). On the nature and role of English in Asia. *Linguistics Journal*, 2 (2), 88-131.
- Pascoe, M. (2005). What is intelligibility? How do SLP's evaluate and address children's intelligibility intervention? *The Apraxia-Kids Monthly*, 6, 5.
- ([www.apraxia-kids.org/site/c.chKMIOPiIsE/b.980831/apps/s/content.asp?ct=911039](http://www.apraxia-kids.org/site/c.chKMIOPiIsE/b.980831/apps/s/content.asp?ct=911039))

- Raimy, E. (1997). Syllable repair in Sudanese Arabic. *Toronto Working Papers in Linguistics*, 16, 117-131.
- Rasmussen, Z.B. (2007). The inter-language speech intelligibility benefit: Arabic-accented English. BA Honors Thesis in Linguistics. The Speech Acquisition Lab: University of Utah. 6-8. ([www.linguistics.utah.edu/speechlab/prespub.html](http://www.linguistics.utah.edu/speechlab/prespub.html))
- Rhebergen, K.S. & Versfeld, N.J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *Journal of the Acoustical Society of America*, 117, 2181-2192.
- Seo, M. (2003). A Segment contact account of the patterning of sonorants in consonant clusters. Ph.D. Dissertation, Ohio State University.
- Subramaniam, N., & Ramachandraiah, A.(2006). Speech intelligibility issues in classroom acoustics: A review . *IE(I) Journal-AR*, 87, 28-33.
- Suhana, Sh. (2001). A cross-linguistic study of phonological development. *Journal of Undergraduate Research*, University of Florida, 2 (11).
- Tomokiyo, L.M., A.W. Black, & K.A. Lenzo (2003). Arabic in my hand: Small-footprint synthesis of Egyptian Arabic. *Proceedings of Eurospeech 2003*, 2049-2052.
- Trudgill, P., & J. Hanan (2002). *Guide to the variations of standard English*. Oxford University Press, New York.
- Walker, R. (2001). Pronunciation for international intelligibility. Karen's linguistics issues: Free resources for teacher and students of English. *English Teaching Professional Magazine*, 21. ([www3.telus.net/linguisticsissues/internationalintelligibility.html](http://www3.telus.net/linguisticsissues/internationalintelligibility.html)).
- Wang, H. (2007). *English as a Lingua Franca. Mutual intelligibility of Chinese, Dutch and American speakers of English*. LOT Dissertation series nr. 143, LOT, Utrecht.
- Wang, H., & Heuven, V.J. van (2007). Quantifying the interlanguage speech intelligibility benefit. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1729-1732.
- Watt, D.J.L., Docherty, G.J., & Foulkes P. (2003). First accent acquisition: a study of phonetic variation in child-directed speech. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1959-1962.

## Appendices

**Table A1.** Stimuli used in Modified Rhyme Test: Vowels

	Target		Distracters		
1.	i:	peat	pit	pet	put
2.	ɪ	pit	peat	pat	pet
3.	ei	late	let	lit	light
4.	e	pet	put	pit	pat
5.	æ	pat	put	pet	pot
6.	ɑ:	bard	board	bird	beard
7.	ʌ	nut	net	not	nit
8.	ɔ	pot	pat	putt	put
9.	əʊ	bow	boy	buy	bay
10.	ʊ	full	fill	fool	fell

	Target		Distracters		
11.	u:	fool	full	fill	fell
12.	ai	mile	male	mill	meal
13.	ɑu	out	ate	oat	at
14.	ɔi	boy	buy	bay	bow
15.	ɪə	peer	pair	poor	pore
16.	ɛə	air	err	or	ear
17.	ɜ:	bird	bard	board	beard
18.	ɔ:	board	beard	bard	bird
19.	ʊə	poor	peer	pair	pore

**Table A2.** Stimuli used in Modified Rhyme Test: Single onset consonants

	Target		Distracters		
1.	p	pin	tin	fin	chin
2.	t	tame	name	game	dame
3.	k	cold	hold	told	gold
4.	b	bang	fang	rang	gang
5.	d	den	ten	men	pen
6.	g	got	pot	cot	jot
7.	f	fid	hid	lid	bid
8.	θ	thaw	law	paw	saw
9.	s	sip	rip	dip	tip
10.	ʃ	shut	but	nut	gut
11.	v	vest	test	best	nest

	Target		Distracters		
12.	z	zeal	peel	feel	seal
13.	ð	then	pen	den	ten
14.	tʃ	chit	bit	fit	sit
15.	dʒ	job	rob	bob	cob
16.	m	must	bust	dust	gust
17.	n	not	tot	pot	lot
18.	l	led	bed	red	wed
19.	r	rent	went	bent	dent
20.	w	wick	pick	tick	lick
21.	j	yen	fen	pen	hen
22.	h	hit	lit	bit	wit

**Table A3.** Stimuli used in Modified Rhyme Test: Single coda consonants

	Target		Distracters		
1.	p	<b>cop</b>	con	cock	cog
2.	t	<b>pat</b>	pad	pan	pack
3.	k	<b>lack</b>	lap	lab	lad
4.	b	<b>lab</b>	lap	lad	lack
5.	d	<b>mad</b>	map	mat	man
6.	g	<b>cog</b>	cod	cop	cock
7.	f	<b>safe</b>	save	sane	sage
8.	θ	<b>heath</b>	heat	heave	he's
9.	s	<b>lace</b>	lane	late	lame
10.	ʃ	<b>mash</b>	match	map	man

	Target		Distracters		
11.	v	<b>rave</b>	raise	rate	ray
12.	ð	<b>with</b>	wid	wick	wiz
13.	z	<b>raise</b>	race	rate	ray
14.	tʃ	<b>match</b>	mass	mash	mad
15.	dʒ	<b>page</b>	pale	pane	pave
16.	m	<b>dam</b>	dad	dan	dab
17.	n	<b>can</b>	cat	cab	cam
18.	ŋ	<b>bang</b>	ban	bad	back
19.	l	<b>pale</b>	page	pane	pave

**Table A4.** Stimuli used in Modified Rhyme Test: Onset consonant clusters

	Target		Distracters		
1.	pl	<b>ply</b>	pry	try	dry
2.	kl	<b>clean</b>	queen	green	glean
3.	dr	<b>drain</b>	brain	grain	crane
4.	gl	<b>glaze</b>	blaze	craze	graze
5.	st	<b>sty</b>	spy	sky	sly
6.	sl	<b>slack</b>	smack	snack	stack
7.	sw	<b>swine</b>	shrine	twine	spine
8.	spl	<b>splint</b>	sprint	squint	print

**Table A5.** Stimuli used in Modified Rhyme Test: Coda consonant clusters

	Target		Distracters		
1.	st	<b>mast</b>	mask	marched	marked
2.	nt	<b>lint</b>	lilt	limp	link
3.	ŋk	<b>wink</b>	winch	wins	wind
4.	bd	<b>fibbed</b>	filth	film	filled
5.	ts	<b>putts</b>	pudge	pulse	punt
6.	gz	<b>bugs</b>	buzzed	bulb	bussed
7.	nz	<b>buns</b>	butts	buds	bums
8.	lm	<b>elm</b>	else	elk	elf

**Table A6.** SPIN (Speech in Noise) sentence intelligibility test. Only contextually highly predictable keywords were used. Keywords are always sentence final.

1.	Throw out all the useless junk.
2.	She cooked him a hearty meal.
3.	Her entry should win the first prize.
4.	The stale bread was covered with mold.
5.	The fireman heard her frightened scream.
6.	Your knees and your elbows are joints.
7.	I ate a piece of chocolate fudge.
8.	Instead of a fence plant a hedge.
9.	The story had a clever plot.
10.	The landlord raised the rent.
11.	Her hair was tied with a blue bow.
12.	He's employed by a large firm.
13.	To open the jar twist the lid.
14.	The swimmer's leg got a bad cramp.
15.	Our seats were in the second row.
16.	The thread was wound on the spool.
17.	They tracked the lion to his den.
18.	Spread some butter on your bread.
19.	A spoiled child is a brat.
20.	Keep your broken arm in a sling.
21.	The mouse was caught in the trap.
22.	I have got a cold and a sore throat.
23.	Ruth poured herself a cup of tea.
24.	The house was robbed by a thief.
25.	Wash the floor with a mop.