

AN EXPLORATORY STUDY OF FIRST EXPOSURE TO URDU – DISCOVERING WHAT MATTERS

EVA MALESSA & REBECCA LIM
(Newcastle University) ¹

Abstract

Little research has been conducted on how adults identify sound strings as words and map meaning to these strings when they encounter naturalistic input for the first time. FIRST EXPOSURE studies examine how adults process novel stimuli in an unfamiliar language. This paper reports on a replication of Gullberg et al.'s (2012) work on Dutch speakers' first exposure to Mandarin Chinese. In our small-scale study, 38 English-speaking participants were exposed to Urdu in a four-minute weather report and then tested on word recognition and sound-to-meaning mapping. Controlled variables included item frequency, loanwords and visual highlighting. The current study found that the loanwords *traffic* and *degree* facilitated word recognition. Results on item frequency and visual highlighting were not deemed sufficiently conclusive to suggest facilitation effects. D-prime (d') calculations for the tasks revealed a very low discriminability of target and filler items. The current study is thus unable to support Gullberg et al.'s (2012) findings that adults are able to extract language-specific information about a novel language after minimal exposure. Future studies are needed to establish whether sensitivity to a novel language is language-specific or not, and to further study the effect of first exposure with regard to different novel languages.

1 Introduction to FIRST EXPOSURE

Making sense of the speech stream is a challenge for language learners confronted with a new and unknown language for the first time. The first crucial steps include SEGMENTATION (identification of relevant sound strings as 'words') and MEANING-MAPPING (assigning relevant meaning onto those 'words'). These steps constitute the complex 'learner's problem of analysis' (Klein 1986: 59). Shoemaker and Rast (2013: 167) point out that while the limitations of second language (L2) speech processing at intermediate and advanced proficiency level have been extensively studied, very little research has focused on 'how learners initially break into the sound stream'.

Remarkably little research has, in fact, been conducted in second language acquisition (SLA) on the topic of first exposure, investigating how adult L2 learners master segmentation and meaning-mapping challenges when exposed to naturalistic input² in form of complex and coherent speech for the first time. This scarcity of relevant research has been noted by the SLA research community (Carroll & Windsor 2015: 58, Han & Liu 2013: 145, Carroll 2004: 236). This paper reports on a small-scale replication of Gullberg et al.'s study (2012), investigating how adult learners process naturalistic yet controlled input after minimal exposure to a novel language. The following section presents an overview of relevant studies. Section 3 illustrates this study's methodology. Results are presented in section 4. Finally, section 5 concludes with WHAT really matters in first exposure.

¹ We would like to acknowledge Professor Martha Young-Scholten, who was a great source of inspiration, for her patient guidance throughout the process of this study.

² In this paper, the term 'input' is used interchangeably with the term 'stimuli', following standard practice in SLA literature. Carroll (2001: 8) points out that 'most writers, however, do not bother to define what they mean by the term'. Flege (2009: 175) makes a similar point, further defining L2 speech input as 'all L2 vocal utterances the learner has heard and comprehended, including his own'. In this study, the terms 'input' and 'stimuli' are employed 'in a general sense to refer to the language a learner hears' (Piske & Young-Scholten 2009: 263). Following Rast (2010), the terms 'learning' and 'acquisition' are also used interchangeably.

2 Background literature: From hearing noise to decoding

The initial stages of SLA have recently started to attract researchers' attention, in particular with regard to first language (L1) effect and L2 input processing and constraints (Carroll & Windsor 2015: 58). Emerging findings indicate that learners go beyond their L1 perceptual filters and point to sensitivity to phonotactic and other information providing cues for speech stream segmentation (Carroll 2013; Shoemaker & Rast 2013). Carroll (2004: 229) stresses that 'hearing words is no primitive operation' as it results from 'some rather complex phonetic and phonological parsing and word-recognition operations'. Her focus is thus on how the learner makes the transition from hearing incomprehensible noise to hearing some sequence of syllables (ibid: 236).

Most first exposure studies (Carroll 2014, 2012, Rast 2010, Ristin-Kaufmann & Gullberg 2014, Shoemaker & Rast 2013) have focused on the adult L2 learners' ability to segment prosodic words from continuous speech. They have shown that learners are able to segment L2 sound forms after very little input, regardless of the similarity between L1 and L2 (Carroll & Windsor 2015: 59). Whether or not segmentation is easy and/or rapid remains a debated issue. Carroll and Meisel (2015: 11) point to earlier research showing that 'segmentation even at the initial stage is rapid', while Han and Liu (2013) claim otherwise.

The role of INPUT remains a highly debated topic in SLA, in particular with regard to adult learners' capabilities to process natural speech (see Carroll 2014). Gullberg et al. (2012) emphasize that while 'input' as a central issue in SLA research has received a lot of attention, surprisingly little is known about HOW input matters. However, input undoubtedly has an effect on SLA acquisition (Piske & Young-Scholten 2009: 1). The studies by Gullberg et al. (2010, 2012) indicate that 'at the earliest stages of L2 acquisition [...] the adult learning mechanism can deal efficiently with very little input and very complex input' (Gullberg et al. 2010: 18). These studies differ from existing studies as they employ 'natural input', visual cues, the effect of item frequency, 'implicit/incidental learning outside of a classroom context' and extremely limited exposure (Gullberg et al. 2012 : 244-245).

Ristin-Kaufmann and Gullberg (2014: 18) point out that while the contribution of artificial and statistical language studies to the understanding of L2 learning has been enormous, they are not comparable to 'naturalistic' language exposure studies. In the study conducted by Gullberg et al. (2012), native Dutch participants were exposed to seven minutes of controlled naturalistic input in Chinese Mandarin without explicit instruction and then tested on their segmentation and meaning-mapping abilities. Their results suggested that adults draw on frequency, gestural cues and phonotactic sensitivity in decoding a novel language. Gullberg et al. (2012: 259) proposed that 'the adult learning mechanism appears to be a great deal more powerful than typically assumed in the L2 acquisition literature'.

2.1 Background of the present study, a replication of Gullberg et al.'s (2012) study

The following research questions, based on Gullberg et al. (2012), were adopted for this study:

1. Can adults EXTRACT information about word forms and meanings after minimal exposure to naturalistic yet controlled input in a previously unknown language?
2. What are the EFFECTS of item frequency, visual highlighting (icons and gestures) and lexical transparency (loanwords)?

Variables of this study included frequency of item appearance and visual highlighting. Lexical transparency was included as a variable to investigate the lexical effects on segmentation as suggested in existing literature (Carroll & Windsor 2015: 59). Shoemaker and Rast (2013: 169) highlight that 'transparent words' show a higher detectability as they 'are mapped onto existing mental representations'.

It was hypothesized that participants would be able to segment sound strings after a very short exposure in the pre-testing treatment. Learners were also predicted to be able to extract word-related information from the complex and coherent speech they were exposed to in the weather report. Finally, frequency and visual highlighting were expected to facilitate detectability of target words and the presence of loanwords was anticipated to enhance participants' ability to segment and map meaning to form.

Our study investigated whether or not Gullberg et al.'s (2012) findings for Dutch participants were possibly language-specific. This formed part of the motivation for this study, in which English-speaking participants were exposed to Urdu. Urdu, an Indo-Aryan language, was selected due to its typological distance to English and the fact that participants had had no contact to Urdu prior to treatment. Urdu is a syllable-timed language, differing from stress-timed English. In addition, its phonemic inventory consists of fewer vowels. There are, however, more consonants in Urdu compared to English. Urdu and English also differ with regard to orthographical and typographical characteristics. Urdu has, additionally, a complex morphology. In terms of phonotactics, complex onset and coda clusters are, nevertheless, allowed in both languages. (Schmidt 1999).

The following section presents the methods and materials employed in this study.

3 Methods and materials

3.1 Participants

Participants in this study were linguistics students from Newcastle University. The study involved one experimental group consisting of 38 participants (7 males and 31 females). A background questionnaire³ revealed that the majority of thirty English L1 speakers was complemented by eight English L2 speakers. All but six English L1 participants reported that they were bilingual to some degree. Fourteen English L1 speakers reported knowledge of one L2 and ten indicated having knowledge of two or more L2s. The non-native English speakers were multilingual to various degrees (see Appendix 2 for details). All participants reported no prior knowledge of Urdu.

3.2 Treatment

The current study incorporated a pre-testing treatment, in which the participants were exposed to fully controlled, continuous novel stimuli in Urdu (see Appendix 3). During the treatment, participants received naturalistic input in the form of a Pakistan audio-visual weather report. The four-minute video, featuring a female native Urdu speaker, was made available online for participants' access. The video contained six visual weather charts. The script of the continuous and coherent weather text consisted of 120 clauses and 24 target words, which the participants were later tested on (see Appendix 5). The target word **mosam** 'weather' in the first clause of the weather report is highlighted in Example 1 below.

- (1) peer k Pakistan main **mosam** k hal k Main hu Qudsia
 din sath,
 monday Pakistan in weather report with I am Qudsia
 'This is Qudsia with your Monday weather report in Pakistan.'

Item frequency and the speaker's gestures were controlled. Target words were, however, not controlled for sentence position. The participants were instructed to watch the four-minute video at least twice at their own convenience within one week. In addition, participants were

³ The questionnaire required participants to self-indicate their knowledge and proficiency level of other languages. It is therefore possible that participants either over- or underestimated their language knowledge/proficiency level.

told to pay attention to loanwords in the treatment video. All participants reported on their total number of viewings in a background questionnaire during the testing (see Appendix 2).

3.3 Testing

The testing consisted of two tasks, completed in one session. Task 1 was a word recognition task (see Appendix 4). Task 1 consisted of 42 items, including 20 Urdu target words, 12 Arabic filler words and 10 Urdu filler words (see Appendix 5). The word order was randomized to ensure they did not appear in the same order as the treatment video. Participants were instructed to listen to audio clips recorded prior to testing by the same native Urdu speaker who presented the weather report. After each clip participants had to indicate whether they had heard the word in the weather report during the treatment by ticking the corresponding *Yes* or *No* box, see Example 2. There was a time-limit of two seconds between each item to ensure instantaneous response. The audio clips were played only once.

(2) Item 7 traffic

Yes No

Task 2 was a sound-to-meaning mapping task with six pre-recorded audio files containing four Urdu target words, as well as one Urdu and one Arabic filler word (see Appendix 4 and 5). For each item, a visual highlight (in form of an icon) used in the treatment video was shown on the screen, followed by an audio clip containing a word. The icon order was randomized. Participants were instructed to indicate whether the utterance in the audio clip correctly represented the visual highlight shown on the screen by indicating *Yes* or *No* in the corresponding boxes, see Example 3 below. Participants were given only four seconds between each set of stimuli.

(3) Item 1 barf

Yes No



The following section presents this study's results.

4 Results

This study's data was analysed by obtaining the percentage of correct identification of TARGET words (used in the treatment) and FILLER words (not used in the treatment video). Task 1 and 2 contained 24 target and 24 filler items in total. Each item, correctly identified as either target item (heard in the treatment) or filler item (not heard in the treatment) was awarded a score of 1. An incorrect response was awarded a score of 0. The data of three participants (*36, *37, *38) were excluded as these participants were directly involved in the set-up of the treatment and testing, rendering their data inappropriate for the study (see Appendix 2). This reduced the number of subjects in the data analysis to 35.

4.1 Overall performance

In Task 1, participants were able to correctly identify 55.24% of all 42 items as either target words or filler words. On average, they were able to correctly map the meaning of the word to icons for 60.95% of all six items in Task 2.

Table 1: Percentage of overall performance

Overall performance (Target + Filler)	
Task 1	55.24%
Task 2	60.95%
Task 1 + 2	53.99%

The overall performance in both tasks was just above chance (53.99%), which was lower than anticipated. One possible explanation for the low performance might have been the lack of control in ensuring that all participants viewed the treatment video at least twice. This lack of control might have resulted in participants not watching the video at all, or only watching it once. Since the number of viewings was not a controlled variable, this study relied on the collaboration of the participants and their honesty in reporting their number of viewing times.

4.2 Comparison of the performance on target and filler items

In Task 1, the two loanwords *traffic* and *degree* were included as target items. Due to the participants' bias resulting from the viewing instructions, it came as no surprise that participants were highly successful in recognizing both L7 *traffic* and L22 *degree*, with 85.71% and 88.57% success rates respectively (see Appendix 7). Performance for these target words was therefore distinctively higher compared to the performance of target words (see Table 2).

Table 2: Percentage of performance of target words

Overall performance (Target only)	
Task 1	43.13%
Task 2	38.57%
Task 1 + 2	42.40%

The overall performance on target words (42.40 %) was very low, compared to the overall performance on filler words (67.14%), as shown in Table 3 below. In Task 1, both Arabic and Urdu filler words were identified as new input in 67.92% of cases. A more detailed breakdown revealed greater detectability of Arabic than Urdu filler words. On average, 68.79% of Arabic fillers and 65.19% of Urdu filler items were successfully recognized.

Table 3: Percentage of performance on filler words

Overall performance (Filler only)	
Task 1 (Arabic)	69.52%
Task 1 (Urdu)	66.00%
Task 1 (Overall)	67.92%
Task 2 (Arabic)	60.00%
Task 2 (Urdu)	57.14%
Task 2 (Overall)	58.57%
Task 1 + 2 (Arabic)	68.79%
Task 1 + 2 (Urdu)	65.19%
Task 1 + 2 (Overall)	67.14%

Based on the results of performance, i.e. the correct identification as either previously heard items or new items, participants seemed to perform better at correctly identifying filler words in both tasks as non-target words (i.e. new items) in contrast to identifying previously heard items as target items. In order to obtain a more distinct indication of performance, response

bias (β) and D-prime (d') were measured. The following section presents the results for the β and d' calculations and discusses some problematic issues of the current study.

4.3 Response bias and sensitivity

In order to separate detection, i.e. correct identification, from RESPONSE BIAS, ‘the general tendency to respond *yes* or *no*’ (Stanislaw & Todorov 1999: 139), values for the factors response bias (β) and sensitivity (d'), in the obtained performance results in Task 1 and 2, were calculated, based on the Signal Detection Theory (SDT)⁴.

The obtained value for sensitivity (d')⁵ in Task 1 (see Table 4 below) indicated low discriminability, signifying participants’ minimal sensitivity and hence, inability to distinguish target items from filler items. For Task 1, the quantification of the response bias⁶ with β indicated a bias towards responding with *no* (< 1), i.e. they have not heard the word before. This bias towards responding *no* possibly accounts for the unexpected higher detectability for filler items (67.14%) in comparison to target items (42.40%), as mentioned in section 4.2.

Table 4: Calculation of response bias and sensitivity for all participants

	Task 1	Task 2
d'	0.245	- 0.056
β	1.087749	0.984402

For Task 2, the quantification of the response bias with β showed a bias towards responding with *yes* (> 1). As two thirds of the tested items in Task 2 were target items, this bias may have resulted in the higher overall performance in Task 2 (60.95%) compared to the overall performance in Task 1 (55.24%). In addition, the negative value of d' (-0.056) suggests response confusion, i.e. ‘responding *yes* when intending to respond *no*, and vice versa’ (Stanislaw & Todorov 1999: 139-140).

Both the bilingual, and in particular the multilingual native-English speakers, were affected by this response confusion in Task 2, while the monolingual speakers were only slightly affected and the non-native speakers remained unaffected (see Appendix 11). Based on the calculations for d' , no similar observations were made for Task 1 (see Appendix 10).

Interestingly, the response bias values for the monolingual English-native speakers and bi-/multilingual native speakers differed slightly. While the monolingual speakers tended to respond *no*, according to their response bias value (1.07), the bi-/multilingual speakers favoured *yes* (0.98), see Appendix 11 for details.

4.4 Item frequency and visual highlighting

For the purpose of this small scale-study, an item was classified frequent⁷ if it occurred five or more times in the weather report. Only 18.75% of the target words were classified as frequent in this study, with the majority of target items appearing infrequently. The most frequent word T19 *ilaqay* ‘areas’ had an unexpectedly low detectability (22.86%) compared to an average of 40.48 % for target items in Task 1, as shown in Table 5 below.

⁴ See Table 1 and Table 2 in Stanislaw & Todorov (1999: 145).

⁵ A value of 0 for d' indicates an inability to distinguish signals from noise. Larger d' values (> 0) indicate a correspondingly greater ability to distinguish signals from noise.

⁶ A β value of 1 indicates no bias. A value less than 1 signifies a bias towards *yes*, and a value more than 1 signifies a bias towards *no*.

⁷ In Gullberg et al.’s (2012: 245) study frequent words occurred eight times or more.

Table 5: Percentage performance on frequent and infrequent target words

Overall Performance (Target Items)			
Task	Frequent Items	Infrequent Items	Overall
1	40.48%	38.18%	55.24%
2	45.71%	31.43%	60.95%

Q7 *traffic* and Q22 *degree*, both loanwords as well as frequent target items in Task 1, constituted the most successfully detected items in the study, with over 85% recognition success (refer to Appendix 6 for details). This might have affected the outcome of the overall performance for frequent items. The recognition success of frequent items (40.48%) in Task 1 is thus seen as comparable to that of infrequent items (38.18%). Contrary to prior expectations, this study did not find any significant correlation between item frequency and the successful recognition of target items, and differs, therefore, from Gullberg et al.'s (2012: 254) finding that 'if a word appeared frequently in the input, it was more likely to be recognized than if it had appeared infrequently'.

In Task 1, target items denoting directions, T5 *shumal* 'north', T8 *janub* 'south' and T26 *mashraq* 'east', were accompanied by a gesture in the treatment video. The average performance for these three target items was extremely low, under 40.00%. It is important to consider that the lack of rehearsal by the speaker prior to recording the treatment video might have resulted in spatial and temporal inaccuracy of the gestures. As a result, the gestures in the treatment video might not have been sufficiently obvious to participants, providing no 'reliable deictic link between gesture, speech and weather charts' (Gullberg et al. 2012: 246).

In Task 2, only one outlier (5F) correctly identified all six items in the task (see Appendix 9). The meanings of the aural items were only 45.23 % of the time successfully mapped by all participants to the corresponding icons. It seems very likely that the role of visual highlighting in form of icons might have been overestimated in this study due to the small number of items in Task 2 (6 items in total: 4 target items and 2 filler items) compared to Task 1 (42 items in total: 10 target items and 12 filler items).

Our initial results seem to suggest no effect of visual highlighting on word recognition in this study. This would be in line with the observations by Gullberg et al. (2012: 254). Taking into account the overall results and problematic issues such as the frequency of target words, number and distribution of target and filler words as well as participants' sensitivity (d') and response bias (β), this study's findings could, however, not convincingly show whether item frequency and visual highlighting facilitate detectability.

5 Discussion – What really matters

D-prime measurements revealed the participants' miniscule ability to distinguish target from filler items. This is also reflected in the overall performance results, which were minimally above chance (see Table 1). Further this study found that participants were compromised by a response confusion in Task 2. Based on these shortcomings, this study's results were interpreted as non-significant regarding the original hypothesis that adults are able to extract language-specific information after minimal exposure to naturalistic yet controlled input in a novel language. Future studies are needed to shed more light on the (in)effectiveness of first exposure.

This study found, nevertheless, that, according to prior expectations, LEXICAL TRANSPARENCY displayed by the two loanwords *traffic* and *degree* MATTERS, as it enhances segmentation and meaning-mapping in a novel language. Previously, Shoemaker and Rast (2013) have demonstrated that loanwords facilitate segmentation and meaning-mapping in speech streams and have also provided evidence of such lexical effects on segmentation (see Carroll & Windsor 2015: 59).

In order to boost scientific rigor of the current study, monitoring the treatment input in terms of length, frequency and setting as well as employing a control group is seen to be essential. Additionally, controlling participants' language backgrounds to ensure methodological rigor in future studies is considered crucial. According to Rast (2010: 80) not only the learners' native languages, but also other languages play an important role in the acquisition of an additional L2. Carroll and Windsor (2015: 55) claim that 'familiarity with other L2s can offer learners phonetic variants as possible targets for segments recognized as functionally equivalent'. Similarly, Carroll (2004: 235) proposes that bilinguals use a segmentation strategy that is 'based on exposure to the prosodic properties of their dominant language, and transfer it to their weaker one'. It seems clear then that knowledge of both L1 and L2(s) matters in first exposure studies.

In addition, aptitude, especially phonetic coding ability (Skehan 1989), appears to be an interesting variable that should be considered in future first exposure studies. Shoemaker and Rast (2013: 180) suggest that generalizations from studies involving experienced language learners to 'less experienced learners should be made with caution'. It is obvious that this also concerns generalizations from less experienced/monolingual learners to multilingual learners.

Generalizing from one particular type of input, in this case a specific text type of a 'pseudo-weather report' (Carroll 2014: vii-viii) is certainly likewise problematic. The lack of variability in text type and structure needs to be taken into account (Carroll 2013: 137). Variables including the number of speakers used in the input, speech rates and environment (noise), which are typical of normal speech, need to be considered as well (Carroll 2014: ii). In a follow-up of the present study, input containing two or more speakers engaged in a two-way conversation could be employed to increase the authenticity of the input and improve variability. Rast (2010: 80) emphasizes further 'the relative importance of interacting variables, such as frequency and transparency, on different types of language activity'. Carroll (2012: 40), on the other hand, stresses that 'repeated exposure to the same input is not necessary to memorize a sound token'. Is less then really more in the initial stages of learning? (cf. Denhovska et al. 2016). There remains a lot to be discovered about the implicit learning mechanisms of adults, and first-exposure studies are just the tip of the iceberg revealing what matters in SLA.

Appendix 1: Questionnaire

QUESTIONNAIRE

School of English Literature, Language and Linguistics,
Percy Building,
Newcastle upon Tyne,
NE1 7RY, UK

First name + initial of family name: _____

First language(s): _____

Other languages:

Beginner level _____

Intermediate level _____

Advanced level _____

Have you ever had any contact with Urdu before watching the video (e.g. Urdu-speaking friends, shopping at a Pakistani grocery store)?

How many times did you watch the video?

once twice three times or more

When was the last time you watched the video? _____

Appendix 2: Participants' background information based on the questionnaire

Participants are referred to according to a number followed by the letter M (male) or F (female), e.g. 1M.

No.	Gender	First Lg	BEG	INT	ADV	Urdu Contact	Times watched video	Last watched	Number of Languages	
1	M	EL	NIL			N	2	6-Dec	1	NATIVE'
2	F	EL	NIL			N	1	30-Nov	1	
3	F	EL	NIL			N	1	30-Nov	1	
4	F	EL	NIL			N	1	30-Nov	1	
5	F	EL	NIL			N	1	29-Nov	1	
6	F	EL	NIL			N	2	6-Dec	1	
10	M	EL	FRENCH			N	1	30-Nov	2	NATIVE' + L2
24	F	EL	FRENCH			N	1	30-Nov	2	
25	F	EL	FRENCH			N	1	30-Nov	2	
26	F	EL	FRENCH			N	1	30-Nov	2	
27	F	EL	FRENCH			N	1	30-Nov	2	
28	F	EL	FRENCH			N	1	30-Nov	2	
19	F	EL	FRENCH			N	1	30-Nov	2	
20	F	EL	FRENCH			N	1	3-Dec	2	
11	F	EL	FRENCH			N	1	1-Dec	2	
12	F	EL	FRENCH			N	2	6-Dec	2	
13	F	EL	FRENCH			N	1	30-Nov	2	
14	F	EL	SPANISH			N	1	30-Nov	2	
15	F	EL	FRENCH			N	2	3-Dec	2	
29	F	EL	SPANISH			Y Friends	2	30-Nov	2	
7	M	EL	SPANISH GERMAN			N	2	7-Dec	3	NATIVE' + L2+
8	M	EL	ITALIAN FRENCH			N	2	2-Dec	3	
9	M	EL	FRENCH SPANISH			N	1	6-Dec	3	
21	F	EL	FRENCH SPANISH			N	2	30-Nov	3	
22	F	EL	FRENCH	GERMAN		N	1	30-Nov	3	
23	F	EL	GERMAN		FRENCH	Y Travel	1	30-Nov	3	
30	F	EL	GERMAN FRENCH			Y Classmate	2	30-Nov	3	
16	F	EL	FRENCH GERMAN			N	1	30-Nov	3	
17	F	EL	DUTCH DANISH ITALIAN	FRENCH	SPANISH	N	3	30-Nov	6	
18	F	EL	ITALIAN GREEK BSL		FRENCH	N	2	6-Dec	5	
31	M	CANTONESE		MANDARIN	EL	N	2	30-Nov	3	NON-NATIVE'
32	M	CANTONESE	JAPANESE	KOREAN	EL MANDARIN	N	1	30-Nov	5	
33	F	POLISH	ITALIAN	PORTUGUESE	GERMAN EL	N	3	3-Dec	5	
34	F	FINNISH SWEDISH	MANDARIN	FRENCH	GERMAN EL	N	3	6-Dec	6	
35	F	MALAY		EL		N	1	30-Nov	2	
*36	F	MANDARIN		EL	CANTONESE	N	2	30-Nov	3	
*37	F	RUSSIAN	MANDARIN		EL	Y Knows Arabic	2	30-Nov	3	
*38	F	MANDARIN	HOKKIEN		EL	N	1	30-Nov	3	

Appendix 3: The Pakistan weather report

Target words are marked in **bold**. Target words also highlighted by a gesture are marked in **green**.

Text associated with weather slide 1

1. **Subho** bakhair
morning good
'Good morning.'
2. peer k din Pakistan main **mosam** k hal k sath, Main hu Qudsia
monday__ Pakistan__in__weather report with I am Qudsia
'this is Qudsia with your Monday weather report in Pakistan.'
3. Guzishta roz k **mosam** k muqably main
yesterday 's weather in comparison with
'In comparison to yesterday's weather.'
4. Aj ka **derja herarat** berqarar nhe hai
today 's temperature stable not is
'today's temperature is not stable.'
5. **Subo** bohet kum
morning very low
'very low in the early morning.'
6. Jab k bhar ker challis **degree** centigrade tak pohnc gaya
whweras rise forty degree centigrade upto reached
'and could climb to as high as 40 degrees Celsius'
7. Bad **dopehr**
late Afternoon
'by the late afternoon.'
8. Lakin mulk k kuch **ilaqay**
but country of some areas
'But some areas of the country'
9. Abi bi **dhop** main kuch lamhat se lutf andoz ho saktay hain
still sunshine in moments enjoy can
'can still enjoy a moment of sunshine.'

Text associated with weather slide 2 (snowy)

10. Aayiay **shumal** ka ahwal jantay hain
let us north situation know
'Let's look at the north.'
11. **Shumal** main
north in
'In the north.'
12. Abi **barf** giray ge
still snow fall will
'it will be snowy.'
13. **Darja-hararat** kam ho ga
temperature low be will
'The temperature will be low.'

14. Ye munfi pandra **degree** centigrade tak jae ga
 this minus fifteen degree centigrade to go will
 'It will be minus 15 degree Celsius.'

15. Aur mazid munfi taees **degree** centigrade tak girnay ka andesha hai
 and more minus twenty-three degree centigrade to go down chance is
 'and it might go down to minus 23 degree Celsius'

16. **Subo** k ibtadaee auqat main
 morning of early time in
 'during the early hours of the morning.'

17. Apko bohet **thand** mehsoos ho ge
 to you very cold feel be will
 'You will feel very cold.'

18. **Sham** k waqt
 evening of time
 'In the evening.'

19. Mazid **barf** bari ho sakti hai
 more snowfall be can is/will
 'there will be more snowy.'

20. Ku k **shumali** Peshawar num hawao ki lupait main hai
 because northern Peshawar moist air of Surrounded by is
 'because the moist air will gather north of Peshawar.'

21. Khas toer se wo **ilaqay**
 especially those areas
 'especially in the areas'

22. paharu k karib **derja herarat** nukta e injimat per hai
 mountains of near temperature below zero on are
 'near the mountains with below-zero temperature.'

23. **Sarku** pe pani jam jany se phislan ho ge
 roads water frozen from slip be will
 'The roads will be very frosty and slippery.'

24. Bahir jatay huway **ahtiat** kijeay
 out going while careful be
 'Be careful when you are out.'

Text associated with weather slide 3 (rainy)

25. Ab **janub** ka ahwal janiay
 now south of situation know
 'Let's look further down at the south.'

26. **Janub** main
 south in
 'In the south.'

27. **Barish** k asaar hain
 rain of chances are

‘It will be rainy.’

28. Aur kuch had tak **garaj chamak** k sth yalabari k imkanat b hain
and some of Thunder with hail fall of chances there are
‘and also has a bit of hail and thunder’

29. **Darja - hararat** kam ho ga
temperature low be will
‘Temperature will be lower.’

30. **Sham** main
evening in
‘In the evening,’

31. Ye taqriban das **degree** centigrade tak pohncay ga
it approximately ten degree centigrade to reach will
‘it could be 10 degree Celsius.’

32. Ap **thand** mehsus kerain gay
you cold feel will
‘You will feel cold.’

33. Apna **chataa** yaad se rakhain
your umbrella Remember take
‘Remember to take your umbrella’

34. Aur apni **barsaatian** b
and your raincoat
‘and your raincoat.’

35. Bahir jatay huway **ahtiat** kijeay
out going while be careful
‘Be careful when you are out,’

36. Jeb ap Gari **chala** rhy hun
when you (vehicle) driving are
‘when you are driving.’

37. Khas toer se Karachi k beshter **ilaqu** main
especially Karachi of most areas in
‘Especially in most areas of Karachi,’

38. Ziada **barish** ho ge
very rainy be will
‘it will be very rainy’

39. Aur **taiz andhi**
and very windy
‘and very windy.’

40. **Traffic** mutasir ho ga
traffic affected be will
‘The traffic will be affected,’

41. Ku k beshter **ilaqu** main mustaqil barish ho gi
because most areas over continuously rain be will
‘because it will rain continuously over most areas.’

Text associated with weather slide 4 (sunny, clear)

42. Ain **mashraq** ki terf dekhien
let's east at look
'Let's look at the east.'

43. **Mashraq** main
east in
'In the east,'

44. Pura din barpoor **dohp** mutwaky hai
all day a lot of sunshine expect is
'we can expect a lot of sunshine all day.'

45. Zyada ter **ilaqon** main matla saaf rhy ga
most areas in the sky clear be will
'Most areas, the sky will be very clear.'

46. **Derja herarat** kafi blund rhy ga
temperature quite high be will
'The temperature will be quite high,'

47. jo ibtadae **subha** k waqt apnay urooj pe ho ga
which early morning of time its highest on be will
'and much higher in the early morning.'

48. Jabk Lahore k zyada **ilaqon** main
whereas Lahore of most areas in
'But most areas of Lahore,'

49. **Dohp** dair tk rhy gi
sunshine last longer will
'the sunshine will last longer.'

50. **derja herarat** main tezi k sath bohet garmi rhy gi
temperature in high of with very hotter be will
'It will be very hotter with high temperature'

51. **Dopehr** gey tak
afternoon late by
'by the late afternoon,'

52. **Mashraq** se aany wali garm lehron ki weja se
east from coming heat wave due to
'due to the heatwave coming from the east.'

53. **Derja herarat** bery ga
Temperatures climb up will
'Temperatures will climb up to'

54. 30 se 40 **degree celsius** tak
30 to 40 degree Celsius upto
'as high as 30 to 40 degree Celsius.'

55. Bahir jatay huway **ahtiat** kijeay
out going while be careful
'Please be careful when you are out,'

56. Aik dasti **chata** rakna soodmand hai

a handy umbrella possession useful is
'a handy umbrella could be useful.'

57. Jeb ap bahir **gari chala** rhy hun
when you out go ing are
'When you are driving out.'

58. Khas ker Pur hujom **sarkon** per
especially crowded roads on
especially on the crowded roads.'

59. Pur sakon rehye aur **ehtiyat** beratye
calm be and careful be
'remember to be calm and careful.'

60. **Garm** surat-e-hal ki wja se
hotter situation, due to
'due to the hotter situation.'

61. Aur blund **derja herarat**
and higher temperature,
'and the higher temperature.'

62. Is se **traffic** masail ho saky hain
it cause traffic problems be can
'it can cause the traffic problems.'

Text associated with weather slide 5 (partly sunny)

63. Ain eb darul hukomat Islamabad me **mosam** per nazr dalty hain
let's now capital Islamabad in weather at look
'Now let's look at the weather in the capital Islamabad'

64. Log eb b **dohp** k lamhat se lutaf andoz ho saky hain
people still sunshine moment enjoy can be
'People can still enjoy a moment of sunshine'

65. Ku k juzwi toer se **dohp** rehay ga
because partly sunny be will
'because it will be partly sunny'

66. Acha **mosam** shehr pe cha jay ga
fine weather city over spread will
'Fine weather will spread over the city.'

67. **Mosam** bohat garm rhy ga
weather very warm be will
'The weather will be very warm.'

68. Gerchy chamakti **dohp** thori dair he rhy gi
though bright sunshine shortly stay will
The bright sunshine will stay shortly, though.

69. Jumarat ko asmaan **badlon** se daka rhy ga
Thursday on sky clouds by covered be will
'The sky will be covered by clouds on Thursday.'

70. **Baadal** door shamaal me jama hon gy
clouds far south in gather will
The clouds will gather in the far south.

71. Aur darul hukumat aur karibi **ilaqun** ki terf berien gy
and capital(city) and near areas upto move will
'and move up to the capital city as well as cities in nearby areas.'

72. Lakin **derja herarat** zyada tabdil nhe ho ga
but temperature too much change not will
'But the temperature will not change too much.'

73. **Derja herarat** blund ya am ho sakta hai
temperature high or low could be
'The temperature could go up or go down'

74. 24 se 28 **degree Celsius** key kareeb
24-28 degree Celsius around
'to around 24-28 degree celsius.'

75. Mujmuae toer pe mutla juzwi **aber alood** rehay ga
collectively sky partly cloudy remain will
'In general, it will be partly cloudy.'

76. **Shumal** se aany wali hewaon ki wja se
north from coming breeze due to
'due to a breeze coming from the north'

77. Aur us k bad shandar **dohp** ho gi
and afterwards brilliant sunshine be will
'and afterwards there will be brilliant sunshine.'

Text associated with weather slide 6 (windy and rainy)

78. Ain **maghrib** ki terf dekhien
let's west at(towards) look
'Let's look at the west.'

79. **Subha** main
morning in
'In the morning.'

80. **Badil** batadreej bardh rehay gy
clouds gradually grow ing will
'clouds will gradually accumulate.'

81. Aur ye inn **ilaqun** k darmiyan rhy gy
and these areas to center be will
'and it will center mainly around these areas.'

82. Boat **andhi** chaly gi
very windy be will
'it will be very windy.'

83. Ghaliban **barishho** sakti hai ksi bi wakt
probably rainbe can any time
'It will probably rain at any time.'

84. **Derja herarat** zyada blund ya kam nhe ho ga
temperature too high or low not be would
'The temperature wouldn't be too high or too low.'

85. Log thori **thand** mehsos krien gy

people bit cold feel will
 'People will feel a bit cold.'

86. Sara din (matla) zyada ter **aber alood** he rhy ga
 throughout day mostly cloudy be will
 'It will be mostly cloudy throughout the day.'

87. Khas ker **sham** main
 especially evening in
 'Especially, in the evening,'

88. Wakfy wakfy se **garaj chamak** ho gi
 time to time from thunder be will
 'thunder will appear from time to time,'

89. Aur isi doraan **andhi** be chaly gi
 and at the same time windy be will
 'and at the same time it will be windy,'

90. ahistagi se barti **Sard** hewa k sath
 slowly moving cold air with
 'with the cold air moving slowly'

91. Bahir jatay huway **ahtiat** kijeay
 out going while be careful
 'Please be careful when you are out,'

92. Aik dasti **chata** rakna soodmand hai
 a handy umbrella possession useful is
 'a handy umbrella could be useful,'

93. Aur apki **barsati** b
 and your raincoat also
 'as well as your raincoat.'

94. Jeb ap bahir gari **chala** rhy hun
 when you out go ing are
 'When you are driving out,'

95. Khas ker Pur hujom **sarkon** per
 especially crowded roads on
 'especially on the crowded roads,'

96. yad rakhye pur sakon rehna hai aur **ehtyat** beratni hai
 remember to calm be and careful be
 'remember to be calm and be careful,'

97. Khrab **mosam** ki wja se
 bad **weather** due to
 'due to the bad weather,'

98. Aur waqt ba waqt chalti **andhi**
 and time to time blowing wind
 'and the wind blowing from time to time,'

99. Is se **traffic** masail ho sakty hain
 it cause traffic problems be can
 'it can cause the traffic problems.'

Text associated with weather slide 7 (weather warnings)

100. Akhar mai aik nazar sahili **ilaqun** pe
 lastly in a look coastal areas upon
 ‘Lastly, let’s have a look at the coastal areas.’

101. Door **janub** main
 far south in
 ‘Here in the far south,’

102. **Derja hararat** maloom nai kia ja sakta
 temperature known not do go can
 ‘the temperature during the day couldn’t be expectable.’

103. Loag **mosami** italat se juray rehain
 people weather report to connected be
 ‘People need to stay tuned for the weather report,’

104. Aur **traffic** k halaat baray
 and traffic of situation about
 ‘and the traffic report,’

105. Khasusan **subo** sweray bheer k waqt
 especially morning early rush of time
 ‘especially in rush hours like early in the morning’

106. Ya pahir **shamko**
 Or then evening
 ‘or in the evening,’

107. Jab har aik gar se kam k liay **sarak** pe hota hai
 when everyone home from work for roads on be would
 ‘when everyone are on the roads from home to work.’

108. Samandaer se uthnay walay ganay **badil**
 ocean from rising massive clouds
 ‘due to massive clouds from the ocean,’

109. Moosla dhahar **barish** ka baies bntay dikhaeee de rehay hain
 cats and dogs rain of cause becoming are
 ‘it will rain like cats and dogs.’

110. Lihaza tofani **barish** se shadeed sailaab mutwaky hai
 so heavy rain from severe flood expected is
 ‘So, severe flooding from heavy rain is expected,’

111. Braye mehrbani **ehtyat** kijye
 please careful is
 ‘please be careful,’

112. Ghair zaruri **safar**/drive se ijtinab kijeay
 non-urgent drive from avoid do
 ‘avoid non-urgent driving out of the house’

113. Jin **ilaqun** main sailab ki paishangoi ki gai hai
 to areas in flood of warning issued is
 ‘to areas which are issued a flooding warning.’

114. Kuch **sarkain** band rehni gi
 few roads block be will
 ‘There will be a few road closures,’

115. Khasusan bheer k auqat main **traffic** musael paida ho saktay
 especially rush of time in traffic problems create be can
 ‘especially in a rush hour that can causes the traffic problems.’

116. Humaray andazay main ganay **badil** mutwaqo hain
 our prediction in massive clouds expected are
 ‘We predict that massive clouds will appear.’

117. Aur tez **andhi** ay gi
 and strong wind come will
 ‘and there will be a strong wind.’

118. Natijatan **garaj chamak** k imkanat hain
 as a result thunder light of chances are
 ‘This can cause thunder and lightning.’

119. Sahili **ilaqu** k janubi shehr kai roaz tak mutasar rehain gy
 coastal areas of south cities several days to affected be will
 ‘Southern city in coastal areas will be affected for several days.’

120. Aj k **mosam** mai bus itna he... mazid jannay k liay humaray sath rehiay
 today's weather in only this much more information for us with let's be
 ‘That's all for today's weather. Stay tuned.’

Appendix 4: Participants' answer sheet

Part 1:

Listen to the following 42 audio files.

Decide whether you had heard these words in the weather report: Yes? No?

Don't think, just **tick!** (You have only 2s for each item).

1.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
2.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
3.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
4.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
5.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
6.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
7.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
8.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
9.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
10.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
11.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
12.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
13.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
14.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
15.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
16.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
17.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
18.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
19.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
20.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
21.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
22.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
23.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
24.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
25.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
26.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
27.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
28.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
29.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
30.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
31.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
32.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
33.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
34.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
35.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
36.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
37.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
38.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
39.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
40.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
41.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
42.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No

Part 2:

Look at the icon shown on the screen AND **listen** to the audio file.

Decide whether the audio file represents the icon: Yes? No?

Don't think, just **tick!** (You have only 4s for each item).

1.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
2.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
3.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No
4.	<input type="checkbox"/>	Yes	<input type="checkbox"/>	No

5. Yes
6. Yes

No
 No

Appendix 5: List of tested items (filler and target items)

Part 1

1	Juma	FW
2	Nujoom	FW
3	Barid	FW
4	Ehtiat	TW
5	Shumal	TW
6	Kursi	FW
7	Traffic	TW
8	Janub	TW
9	Chataa	TW
10	Hafta	FW
11	Thalj	FW
12	Kharif	FW
13	Mosam	TW
14	Itwaar	FW
15	Lahore	TW
16	Bahaar	FW
17	Mosam garma	FW
18	Arabah	FW
19	Ilaqay	TW
20	Budh	FW
21	Khazaan	FW
22	Degree	TW
23	Phool	FW
24	Fannan	FW
25	Kutta	FW
26	Mashraq	TW
27	Sayara	FW
28	Dopeher	TW
29	Karachi	TW
30	Thand	TW
31	Rathath	FW
32	Badal	TW
33	Andhi	TW
34	Chala	TW
35	Sarku	TW
36	Mard	FW
37	Derja herarat	TW
38	Parinda	FW
39	Lail	FW
40	Sham	TW
41	Barsaatian	TW
42	Daftar	FW

Part 2

1	Barf	TW
2	Fajr	FW
3	Dohp	TW
4	Garja chamak	TW
5	Barish	TW
6	Janwar	FW

**TW – target
word**
FW – filler word

Appendix 6: Frequency of target words

Note: The accuracy in classification of frequent and infrequent items might vary to some degree, due to multiple acceptable orthographic variations in Urdu, e.g. *ilaquay* and its variations *ilaqu*, *ilaqon*, *ilaqun*.

Target words	Frequent	Infrequent	Target words	Frequent	Infrequent
Ehtyat, ahtiat	6x		Thand		3x
Shumal(i)		4x	Badal (badil/baadal)		2x
Traffic (trafic)	5x		Andhi	5x	
Janub(i)		4x	Chala		3x
Chataa		1x	Sarku		1x
Mosam(i)	9x		Derja herarat	9x	
Lahore		1x	Sham		3x
Ilaqay/ilaqu/ ilaqon/ ilaqun	11x		Barasaatian (barsati)		2x
Degree	5x		Barf		2x
Mashraq		3x	Dhop (doph)	7x	
Dopehr		1x	Garaj chamak		1x
Karachi		1x	Barish	6x	

Appendix 7: Raw scores for target items (Task 1)

TARGET RESPONSES (TASK 1 - WORD RECOGNITION)																									
No.	Gender	T4	T5	L7	T8	T9	T13	P15	T19	L22	T26	T28	P29	T30	T32	T33	T34	T35	T37	T40	T41	TOTAL CORRECT	% CORRECT		
1	M	1	1	1	0	1	1	1	1	1	0	0	1	0	1	0	1	0	0	1	0	12	60.00%	6	NATIVE MONO
2	F	0	1	1	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	14	70.00%		
3	F	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	15.00%		
4	F	0	1	1	0	0	0	1	1	0	1	0	0	0	1	0	1	1	0	0	0	8	40.00%		
5	F	0	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	14	70.00%		
6	F	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5.00%		
7	M	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	5	25.00%	14	NATIVE BI
8	F	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	10.00%		
9	F	1	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	7	35.00%		
10	F	0	0	0	1	1	0	1	0	1	0	0	0	0	1	0	1	1	1	1	0	9	45.00%		
11	F	1	0	1	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0	1	1	13	65.00%		
12	F	0	1	1	1	1	0	1	0	1	1	1	1	0	1	1	0	0	0	0	1	12	60.00%		
13	F	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	10.00%		
14	F	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	10.00%		
15	F	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	1	8	40.00%		
16	F	1	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1	0	1	1	0	12	60.00%		
17	F	1	0	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0	0	1	0	8	40.00%		
18	F	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	8	40.00%		
19	F	0	0	1	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0	0	0	7	35.00%		
20	F	0	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	0	0	1	0	12	60.00%		
21	M	0	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	14	70.00%	10	NATIVE MULTI
22	F	1	1	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	10	50.00%		
23	F	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	4	20.00%		
24	F	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	15.00%		
25	M	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	0	15	75.00%		
26	M	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1	1	1	0	1	0	14	70.00%		
27	F	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	10.00%		
28	F	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	0	1	0	14	70.00%		
29	F	0	0	1	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	6	30.00%		
30	F	0	1	1	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	1	10	50.00%		
31	M	0	0	0	1	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	6	30.00%	5	NON-NATIVE
32	F	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	3	15.00%		
33	M	0	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0	1	11	55.00%			
34	F	0	1	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	0	0	0	8	40.00%		
35	F	1	1	1	0	1	0	0	0	1	0	0	0	0	1	0	1	1	1	1	0	10	50.00%		
		11	18	30	12	16	12	17	8	31	8	10	18	11	13	14	15	13	10	15	7	289	41.29%	35	
		31.43%	51.43%	85.71%	34.29%	45.71%	34.29%	48.57%	22.86%	88.57%	22.86%	28.57%	51.43%	31.43%	37.14%	40.00%	42.86%	37.14%	28.57%	42.86%	20.00%				

Legend:
 T4 – target item
 L7 – loanword
 P15 – place name

Appendix 8: Raw scores for filler items (Task 1)

FILLER RESPONSES (TASK 1 - WORD RECOGNITION)																												
No.	Gender	U1	A2	A3	A6	U10	A11	A12	U14	A16	U17	A18	U20	U21	U23	A24	U25	A27	U31	U36	Q38	A39	A42	TOTAL CORRECT	% CORRECT			
1	M	1	1	0	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	10	45.45%	6	NATIVE MONO
2	F	1	0	0	0	0	0	1	1	1	1	0	1	0	1	0	0	1	1	1	1	1	1	13	59.09%			
3	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	100.00%			
4	F	1	1	0	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	15	68.18%			
5	F	1	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	1	0	10	45.45%			
6	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	100.00%			
7	M	1	1	0	1	1	1	1	0	0	1	1	0	0	0	1	1	1	0	1	1	1	1	15	68.18%	14	NATIVE BI	
8	F	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	21	95.45%			
9	F	1	1	0	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	17	77.27%			
10	F	0	1	0	0	0	1	0	1	0	1	1	1	0	0	1	0	0	1	0	1	1	0	10	45.45%			
11	F	0	0	1	1	1	0	0	1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	13	59.09%			
12	F	0	1	0	1	0	1	1	0	0	1	0	0	1	1	1	0	0	1	1	0	0	1	12	54.55%			
13	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	100.00%			
14	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	100.00%			
15	F	1	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	8	36.36%			
16	F	1	1	1	1	1	1	1	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	14	63.64%			
17	F	0	1	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1	1	0	1	1	0	14	63.64%			
18	F	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0	1	0	1	1	16	72.73%			
19	F	1	1	0	0	1	1	0	0	1	0	0	1	0	1	1	1	0	1	0	1	0	1	12	54.55%			
20	F	1	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0	1	1	0	0	1	1	11	50.00%			
21	M	0	0	1	1	0	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	0	14	63.64%	10	NATIVE MULTI	
22	F	1	1	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	1	0	1	0	1	12	54.55%			
23	F	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	18	81.82%			
24	F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	20	90.91%			
25	M	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	8	36.36%			
26	M	1	1	0	0	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	9	40.91%			
27	F	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	21	95.45%			
28	F	0	0	1	1	1	1	1	1	0	0	1	0	0	0	0	1	1	0	0	1	1	0	11	50.00%			
29	F	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	19	86.36%			
30	F	1	1	0	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	16	72.73%			
31	M	0	1	1	1	0	1	0	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	16	72.73%	5	NON-NATIVE	
32	F	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	21	95.45%			
33	M	1	0	0	1	0	1	0	0	1	1	0	0	1	0	0	1	0	1	0	1	1	1	11	50.00%			
34	F	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1	0	0	0	0	1	1	0	12	54.55%			
35	F	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	0	0	1	16	72.73%			
		27	25	19	28	27	28	23	19	19	26	24	21	24	15	25	23	25	25	25	22	27	27	24	523	67.92%	35	URDU
		77.14%	71.43%	54.29%	80.00%	77.14%	80.00%	65.71%	54.29%	54.29%	74.29%	68.57%	60.00%	68.57%	42.86%	71.43%	65.71%	71.43%	71.43%	62.86%	77.14%	77.14%	68.57%				ARABIC	

Legend:

U1 – Urdu filler item

A2 – Arabic filler item

Appendix 9: Raw scores on target and filler items (Task 2)

No.	Gender	Q1	Q2	Q3	Q4	Q5	Q6	TOTAL CORRECT	% CORRECT	
1	M	0	0	0	1	1	0	2	33.33%	NATIVE MONO
2	F	0	1	0	0	0	1	2	33.33%	
3	F	0	1	0	0	0	0	1	16.67%	
4	F	1	1	0	1	1	1	5	83.33%	
5	F	1	1	1	1	1	1	6	100.00%	
6	F	1	0	0	0	0	1	2	33.33%	
7	M	0	0	1	0	0	0	1	16.67%	NATIVE BI
8	F	0	0	1	0	0	0	1	16.67%	
9	F	0	0	1	0	0	0	1	16.67%	
10	F	0	1	0	1	1	1	4	66.67%	
11	F	0	1	1	1	1	1	5	83.33%	
12	F	0	0	1	0	1	0	2	33.33%	
13	F	1	1	1	1	0	0	4	66.67%	
14	F	0	1	0	1	1	1	4	66.67%	
15	F	0	1	0	0	0	1	2	33.33%	
16	F	0	1	0	0	0	1	2	33.33%	
17	F	0	1	1	0	1	0	3	50.00%	
18	F	1	1	1	0	0	1	4	66.67%	
19	F	0	1	1	1	0	1	4	66.67%	
20	F	0	0	0	0	1	0	1	16.67%	
21	M	0	0	1	1	1	0	3	50.00%	NATIVE MULTI
22	M	0	0	0	1	1	0	2	33.33%	
23	M	1	0	1	0	0	0	2	33.33%	
24	F	0	0	1	0	1	1	3	50.00%	
25	F	0	1	0	1	1	0	3	50.00%	
26	F	0	1	0	0	0	1	2	33.33%	
27	F	1	1	0	1	0	1	4	66.67%	
28	F	0	0	0	0	0	1	1	16.67%	
29	F	0	1	0	0	0	0	1	16.67%	
30	F	0	0	1	0	1	0	2	33.33%	
31	M	1	1	0	1	1	1	5	83.33%	NON-NATIVE
32	M	0	1	0	0	0	1	2	33.33%	
33	F	0	0	1	0	0	1	2	33.33%	
34	F	0	1	0	1	1	1	4	66.67%	
35	F	0	1	1	0	0	1	3	50.00%	
		8	21	16	14	16	20	95	45.24%	Arabic filler
		22.86%	60.00%	45.71%	40.00%	45.71%	57.14%			Urdu filler

Legend:

Q1, 3-5 – Target items

Q2 – Arabic filler item

Q6 – Urdu filler item

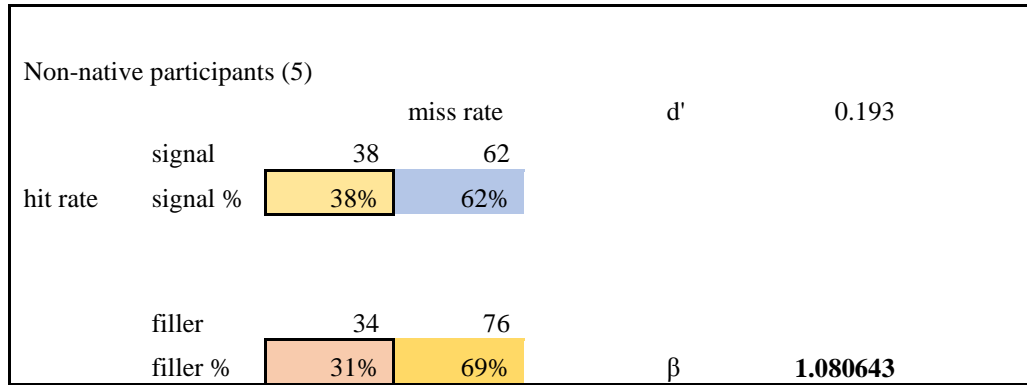
Appendix 10: Values for sensitivity (d') and response bias (β) for Task 1

All participants(35)				
			miss rate	d' 0.245
	signal	289	411	
hit rate	signal %	41%	59%	
	filler	247	523	
	filler %	32%	68%	β 1.087749

Native Monolingual participants (6)				
			miss rate	d' 0.348
	signal	52	68	
hit rate	signal %	43%	57%	
	filler	40	92	
	filler %	30%	70%	β 1.126236

Native Bilingual participants (14)				
			miss rate	d' 0.146
	signal	107	173	
hit rate	signal %	38%	62%	
	filler	101	207	
	filler %	33%	67%	β 1.055852

Native Multi-lingual participants (10)				
			miss rate	d' 0.347
	signal	92	108	
hit rate	signal %	46%	54%	
	filler	72	148	
	filler %	33%	67%	β 1.099731



Appendix 11: Values for sensitivity (d') and response bias (β) for Task 2

All participants (35)					
			miss rate	d'	-0.056
	signal	53	87		
hit rate	signal %	38%	62%		
	filler	28	42		
	filler %	40%	60%	β	0.984402

Native Monolingual participants (6)					
			miss rate	d'	0.220
	signal	10	14		
hit rate	signal %	42%	58%		
	filler	4	8		
	filler %	33%	67%	β	1.073177

Native Bilingual participants (14)					
			miss rate	d'	-0.047
	signal	21	35		
hit rate	signal %	38%	63%		
	filler	11	17		
	filler %	39%	61%	β	0.986289

Native Multi-lingual participants (10)					
			miss rate	d'	-0.572
	signal	15	25		
hit rate	signal %	38%	63%		
	filler	12	8		
	filler %	60%	40%	β	0.9815

Non-native participants (5)					
			miss rate	d'	0.896
hit rate	signal	7	13		
	signal %	35%	65%		
	filler	1	9		
	filler %	10%	90%	β	2.110556

References

- Carroll, S. E. & Meisel, J. M. (2015). Input, learner populations, and the human language-making capacity. In C. Hamann & E. Ruigendijk (eds.), *Language Acquisition and Development. Proceedings of GALA 2013*. Newcastle upon Tyne: Cambridge Scholars Publishing, 1-16.
- Carroll, S. E. & Windsor, J. W. (2015). Segmental targets versus lexical interference. Production of second-language targets on first exposure and the result of minimal training. In Peukert, H. (ed.), *Transfer effects in multilingual language development*. Amsterdam: John Benjamins Publishing Company, 53-86.
- Carroll, S. E. (2014). Processing 'words' in early stage SLA: A comparison of first exposure and early stage learners. In Z-H. Han & Rast, R. (eds.), *Input processing in second language acquisition*. Cambridge: Cambridge University Press, 107-138.
- Carroll, S. E. (2013). Introduction to the special issue: Aspects of word learning on first exposure to a second language. *Second Language Research* 29(2): 131-144.
- Carroll, S. E. (2012). First exposure learners make use of top-down lexical knowledge when learning words. In K. Braunmüller, Gabriel, C. & Hänel-Faulhaber, B. (eds.), *Multilingual individuals and multilingual societies*. Amsterdam: John Benjamins, 23-45.
- Carroll, S. E. (2004). Segmentation: Learning how to 'hear' words in the L2 speech stream. *Transactions of the Philological Society* 102: 227-254.
- Carroll, S. E. (2001). *Input and evidence: the raw material of second language acquisition*. Amsterdam: John Benjamins.
- Denhovska, N., Serratrice, L. & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: the role of frequency and working memory. *Language Learning* 66 (1), 159-190.
- Flege, J. E. (2009). 'Give input a chance!'. In T. Piske & Young-Scholten, M. (eds.), *Input matters in SLA*. Bristol: Multilingual Matters, 175-190.
- Gullberg, M., Roberts, L. & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics* 50: 239-276.
- Gullberg, M., Roberts, L., Dimroth, C. Veroude, K. & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning* 60: 5-24.
- Han, Z-H. & Liu, Z. (2013). 'Input processing of Chinese by *ab initio* learners'. *Second Language Research* 29: 145-164.
- Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.
- Piske, T. & Young-Scholten, M. (eds.) (2009). *Input matters in SLA*. Bristol: Multilingual Matters.

- Rast, R. (2010). The role of linguistic input in the first hours of adult language learning. *Language Learning* 60: 64-84.
- Ristin-Kaufmann, N. & Gullberg, M. (2014). The effects of first exposure to an unknown language at different ages. *Bulletin Suisse de Linguistique Appliquée*, 17-29.
- Schmidt, R. L. (1999). *Urdu: An Essential Grammar*. London: Routledge.
- Shoemaker, E. & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research* 29: 165-183.
- Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.
- Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers* 31(1): 137-149.

EVA MALESSA

School of English Literature, Language and Linguistics

Newcastle University

Newcastle upon Tyne

Tyne and Wear, NE1 7RU

United Kingdom

E.I.Malessa2@newcastle.ac.uk

REBECCA LIM

School of English Literature, Language and Linguistics

Newcastle University

Newcastle upon Tyne

Tyne and Wear, NE1 7RU

United Kingdom

r.lim2@newcastle.ac.uk