Boards for Automated Referential Communication Task

Naomi Enzinna and Sam Tilsen*

Abstract. In this paper, the Boards for Automated Referential Communication Task (BARCoT) experimental program is described. BARCoT emulates the referential communication task used in Hwang et al. 2015, in which participants interact with model talkers to fill in words on a game board. BARCoT was created in order to allow for this task to be conducted in two distant locations (Miami, Florida, and Ithaca, New York), with participants interacting with the same two model talkers in each location. The program also allows for quicker data analysis by automating the labeling and storing of data in a format suitable for use with the Montreal Forced Aligner (McAuliffe et al., 2017). How BARCoT works—including details related to the stimuli, materials, task, and more—are described in this paper, along with information on how this program may be used in the future.

1 Introduction

The purpose of this paper is to describe the Boards for Automated Referential Communication Task (BAR-CoT) program used in Enzinna 2018. BARCoT is an experimental program that was built to emulate the referential communication task used in Hwang et al. 2015. In the task, participants interact with a model talker, using game boards (Figure 1) as a reference during the interaction. The game boards are made up of 6x6 squares, and in some of the squares there are pairs of words. The *word-pairs* consist of two words, which appear in two squares of the same color. The model talker's game board is incomplete, in that the model talker's board is missing words from the word-pairs that the participant's board is not. The participant's goal in the referential communication task is to help the model talker complete their boards. To do this, the model talker asks the participant about the words missing from their boards, and then the participant responds by telling them the words that are missing. The data provided by BARCoT is interactive speech data over an extended period of time (determined by the experimenter and design) and is useful for studies on phonetic accommodation and priming.

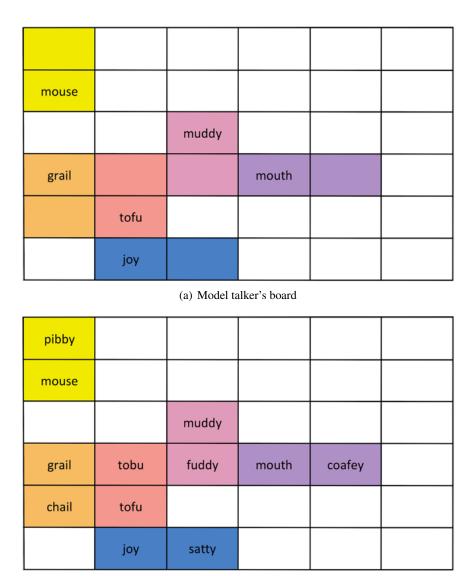
Using Figure 1 from Enzinna 2018 as an example, the model talker asks, "What is by the word TOFU?" Both TOFU and TOBU are in orange squares and next to each other on the participant's board. In response, the participant says, "TOBU is by the word TOFU." The word TOFU is a *prime word*, in that the model talker says it just prior to the participant's response. The word TOBU² is a *target word*, in that only the participant says the word and it contains some target feature that the experimenter is measuring. In Enzinna 2018, the target feature being examined is Voice Onset Time (VOT) in English. TOBU contains a voiceless stop at the beginning of a stressed syllable, and thus a long-lag VOT is expected. The prime word, TOFU, also contains a voiceless stop at the beginning of a stressed syllable. Thus, the word-pair TOFU-TOBU tests for VOT priming effects, because the model talker's VOT in TOFU primes the participant's VOT in TOBU. However, there are also word-pairs on the board that do not test for priming effects. For example, for the word-pair MOUSE-PIBBY, the model talker asks, "What is by the word MOUSE?" and the participant responds, "PIBBY is by the word MOUSE." In this case, MOUSE does not contain a voiceless stop, and the model talker does not prime the participant's VOT in PIBBY.

The BARCoT program runs this referential communication task on a computer in MATLAB, using prerecorded model talker voices for audio stimuli. Participants interact with the pre-recorded model talkers while completing the task, using a headset with audio and recording capabilities. The program then labels the audio data according to various variables (e.g., participant identification number, model talker, word,

^{*}Corresponding author email: nre23@cornell.edu

¹The colors used on the boards are color-blind friendly.

²Some of the stimuli words are nonce words, which will be explained further in Section 2.2.



(b) Participant's board

Figure 1: In the task, the participant helps the model talker fill in the missing words on their board.

linguistic variable, time, trial number, priming, etc.), storing the data in tables. Then, using a Praat script, the audio data can be converted into sound files and Praat Text Grids, which can then be used with the Montreal Forced Aligner (McAuliffe et al., 2017) for automatic segmentation.

BARCoT was created for purposes relevant to the experiment run in Enzinna 2018. In her dissertation, Enzinna examines how participants vary their speech (specifically examining VOT), depending on the participant's linguistic background, the model talker's linguistic background, and the participant's speech community. In particular, Enzinna investigated how English monolinguals and Spanish-English bilinguals from a majority monolingual community (Ithaca, New York) and a majority Spanish-English bilingual community (Miami, Florida) adjusted their VOT when interacting with an English monolingual model talker and a Spanish-English bilingual model talker. Additionally, Enzinna (2018) examined the time-course of accommodation, measuring how participants adjusted their VOTs throughout short-term interactions with each model talker.

Because BARCoT was designed to meet the needs of Enzinna's (2018) experiment, it has the following benefits: First, BARCoT allows for the task to be conducted in distant locations. In Enzinna 2018, the task needed to be conducted in Miami and Ithaca. By using pre-recorded model talker voices, participants in different locations were able to interact with the same model talkers. Second, using pre-recorded model talker voices helps to control the data. All participants in Enzinna 2018 interacted with the same stimuli; thus, any variation in the model talker's speech was the same for all participants, and did not differ on a day-to-day basis like the speech of an in-person model talker. Third, the program was built to automate data processing, which then speeds up data analysis and allows for more data collection. This allowed for an immense amount of data to be collected and analyzed in a relatively short period of time. Finally, our hope in creating the BARCoT program was for future researchers to benefit from the program's capabilities once it is made available for public download.

In the remainder of this paper, we describe how the BARCoT program works (Section 2)—specifically, we provide an overview of the BARCoT program, the task, the materials and stimuli, the randomization and ordering, and the data processing procedures. We conclude by briefly discussing our hopes for the future of the BARCoT program and how readers may acquire the program themselves (Section 3).

2 BARCoT methodology

An overview of how BARCoT works is as follows: In order to run BARCoT, a user must provide the program with instructions, a word-stimuli spreadsheet, and model talker recordings. The word-stimuli spreadsheet must include information about the word-pairs used in the experiment, such as the target words, the prime words, the features being examined (e.g., VOT), etc. The model talker recordings must include sound files and their corresponding labelled Praat Text Grids. Using the word-stimuli spreadsheet and model talker recordings, BARCoT generates a table of randomized sets of boards for the number of participants that the user tells it to produce. The table includes information that will be used for all boards for all participants, such as what squares on the board each target word and prime word will appear in, the corresponding model talker sound file information, etc.

Once all instructions, model talker recordings, and randomized board information have been provided, the referential communication task is ready to run. At this point, the experimenter tells the program the participant's identification number, board set number, and location (if relevant). All corresponding data is then labelled with this information. The task then runs as follows: First, instructions are shown to the participant. Then, the participant completes the practice trials. After the practice trials, the participant is shown another, shorter set of instructions, preparing them to begin the task with the first model talker. Next, they complete the task with the first model talker. Afterward, if there is more than one model talker, the participant is shown another short set of instructions.³ Once they are ready to begin again, they complete the task with the second model talker. After completing the task with all model talkers, they are notified that the experiment is complete.

During the task, the model talkers ask the participants questions about the words on the board, and the participants respond. After a participant responds, they must click on the square that corresponds to the word related to their response. Their click time is then recorded, and the program is triggered to move on to the next trial. If participants need the model talker to repeat the question, they can right click anywhere on the screen and the model talker will repeat the question. The right clicks are also recorded. All participant speech data, board information, and click information are then stored into a data table, which is then used to process the data. A Praat script uses the click times to create Text Grids that can then be used with the Montreal Forced Aligner (McAuliffe et al., 2017) to segment the speech data.

³In Enzinna 2018, the participant was allowed to take a break at this time, if necessary, but they were not allowed to speak to anyone.

In the remainder of this section, details regarding the BARCoT methods are provided—specifically, a description of the task (Section 2.1), word-pairs (Section 2.2), model talker recordings (Section 2.3), non-linguistic materials (Section 2.4), randomization and ordering patterns (Section 2.5), and data processing methods (Section 2.6).

2.1 Task

As described in Section 1, the task employed by BARCoT is a referential communication task similar to the task used in Hwang et al.'s (2015) accommodation study. In the task, participants see a board with 6x6 squares (Figure 2b) on a computer screen. In some of the squares there are word-pairs. Participants are asked about the words on the board by a pre-recorded model talker over a headset with audio and recording capabilities. In the instructions, participants are told that the model talkers have incomplete boards, which they need the participant's help to complete. (Sample instructions from Enzinna 2018 are provided in Appendix A.) To complete the boards, the model talkers ask participants what words are on their boards, referencing other words on the boards to indicate which empty square they need help with. The reference word is always next to the square being asked about and is always in a square of the same color.

Using Figure 2 as an example, the model talker asks, "What is by the word MOUSE?" Both MOUSE and PIBBY are in yellow squares and next to each other on the participant's board. In response, the participant says, "PIBBY is by the word MOUSE." After responding, the participant clicks on the square containing the answer (PIBBY). The model talker then asks about another word on the board. Once the participant has been asked about all of the words on the board, a new board begins. If there is more than one model talker, each participant completes all of the boards with each model talker, one model talker at a time. For example, in Enzinna 2018, a participant would complete all trials (36 trials) with the English monolingual model talker, and then all trials (the same 36 trials) with the Spanish-English bilingual model talker afterward—or vice-versa.

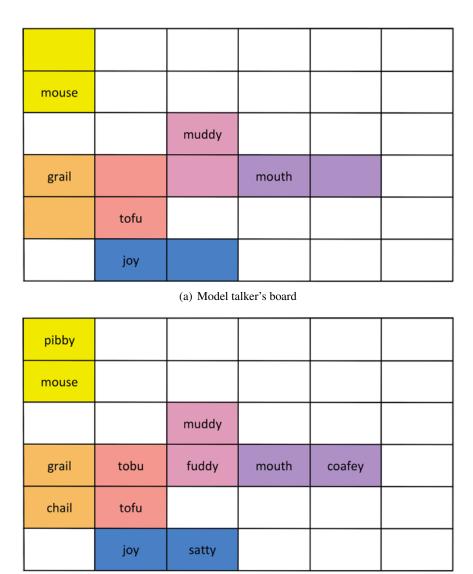
2.2 Word-pairs

The words appear on the boards in pairs (*word-pairs*). The word-pairs consist of a prime word and a target word. Using the boards in Figure 2 as an example, the model talker asks, "What is by the word MOUSE?" and the participant responds, "PIBBY is by the word MOUSE." In this example, the prime word is MOUSE and the target word is PIBBY. The target words are the words missing from the model talkers' boards, which means the participant does not hear the model talkers say the target words. Only the participant says the target words. Both the participant and the model talker(s) say the prime words. What variables are examined and how priming is used within BARCoT is determined by the experimenter, through use of the word-stimuli spreadsheet.

In Enzinna 2018, for example, the word-pairs were designed to allow for examination of the following dependent variables: duration of VOT after a voiceless stop, velarization of word-final /l/, duration of intervocalic /t/ and /d/ (flapping), vowel quality differences for /r ϵ æ α i e α o u/, rhythm, and pitch (henceforth referred to as *target variables*). These target variables were selected because they differ in English and Spanish. In the dissertation, only VOT was examined—specifically the VOTs in the target words. A complete list of the word-pairs containing voiceless stops is provided in Table 1.

108 target words total were used in the experiment in Enzinna 2018. All of the target words contained two of the aforementioned target variables each: one target consonant and one target vowel. 54 of the target words contained a voiceless stop. All of the target words containing a voiceless stop were disyllabic, with the voiceless stop both word-initial and phrase/sentence-initial. All of the target words were nonce

⁴The remaining target variables will be examined in future studies.



(b) Participant's board

Figure 2: In the task, the participant helps the model talker fill in the missing words on their board.

words or very-low frequency words (in cases where there were no nonce options) to increase likelihood of accommodation (Goldinger, 1998; Stollenwerk, 1986; Babel, 2010; D'Imperio et al., 2014). The target word PIBBY, for example, begins with a voiceless stop, contains vowel /ɪ/, is disyllabic, and is a nonce word.

All target words occurred once with a *target prime* and once with an *unrelated prime*. The target primes contained the same target variables as the target word. The unrelated primes did not contain any of the target variables. All of the priming words were real words. The target primes were low-frequency words that share the same target vowel (/ $1 \epsilon \approx \alpha$ i e α o u/) and target consonant (/p t k l t d/) as the target word they were paired with, differing from the target word as little as a possible. For example, for the target word TASSY, the target prime was TAFFY. These words share the target vowel / α / and the target consonant /t/, differing only in place of articulation for the second consonant. The unrelated primes were words that do not contain a target consonant or vowel, and word frequency was not restricted. For example, for the target word TASSY, the unrelated prime was ROY.

Table 1: Word-pairs containing voiceless stops.

TARGI	ET VARIABLES	PRIMED W	VORD-PAIRS	UNPRIMED	Word-Pairs
STOP	VOWEL	PRIME WORD	TARGET WORD	PRIME WORD	TARGET WORD
/p/	/ɪ/	pitchy	piffy	mouse	pibby
/p/	/٤/	peggy	pessy	choice	petchy
/p/	/æ/	patchy	paggy	gore	passy
/p/	/Λ/	puffy	pubby	mime	putchy
/p/	/i/	peachy	peagy	force	peasy
/p/	/e/	pacey	pafey	shout	pabey
/p/	/a/	posse	pobby	why	poffy
/p/	/o/	pogo	pogey	wifi	poachy
/p/	/u/	poofy	pooby	sight	poogy
/t/	/I/	tizzy	tibby	good	tiffy
/t/	/ε/	techy	teggy	house	tessy
/t/	/æ/	taffy	tatchy	roy	tassy
/t/	$/\Lambda/$	tubby	tussy	sort	tuggy
/t/	/i/	teeny	teeby	wood	teefy
/t/	/e/	taser	taber	four	tafey
/t/	/a/	toffee	tossy	out	totchy
/t/	/o/	tofu	tobu	ground	toasu
/t/	/u/	toothy	toosy	rye	toochy
/k/	/I/	kissy	kibby	short	kiffy
/k/	/ɛ/	kegger	kessy	door	keggy
/k/	/æ/	canny	caffy	south	cabby
/k/	/Λ/	cubby	cuffy	voice	cussy
/k/	/i/	keesha	keesy	foot	keechy
/k/	/e/	casey	cafey	hood	cabey
/k/	/a/	coffin	cobby	buy	cossy
/k/	/o/	kobe	coasey	mouth	coafey
/k/	/u/	cougar	coogy	bore	coosy

In addition to the word-pairs used in the experiment trials, there are word-pairs used in the practice trials. These word-pairs should not contain the target variables being examined in the study, for the purpose of not priming participants' speech during this portion of the study.⁵

⁵In Enzinna 2018, there were three practice boards at the start of the experiment. Only the first practice board did not contain any target variables; the following two contained target variables in the target words only (not in the prime words). This was initially done for the purpose of using these values as baseline values, to compare with those in the experiment trials. However, participants made too many mistakes during the practice trials for those values to not be misleading. For this reason, Enzinna does not recommend including target variables in the practice trials. Additionally, after conducting the study and finding that accommodation can be influenced by a speaker's previous interaction, she does not suggest using baseline values as comparisons at all: "I do not believe that there is a such thing as a baseline value, as there is evidence that speech is influenced by recent interactions, changes in frequency, changes in social factors, and so forth. Instead, I recommend analyzing changes and differences in speech within and between interactions in order to better understand accommodation effects" (Enzinna, 2018, p. 51).

2.3 Model talker recordings

For each model talker and each prime word, the experimenter must provide BARCoT with a recording of the model talker asking "What is by the word [PRIME WORD]?" In Enzinna 2018, the model talker recordings were obtained by asking both model talkers to read a list of sentences. All of the sentences in the list were the same frame sentence, varying only by a single word: "What is by the word [PRIME WORD]?" For example, the model talkers read "What is by the word MOUSE?" with MOUSE alternating with other words: TIZZY, POGO, etc. Each sentence repeated twice in the list, presented in a randomized order. Then, those alternating words (e.g., MOUSE) were extracted from the frame sentences and spliced together with one version of the frame sentence. Thus, only one frame sentence was heard by all participants. This reduced additional variation in the recordings, which may have influenced the trials. Additionally, to ensure that the new spliced recordings sounded natural, Enzinna listened to them and selected the most natural sounding version for each word for use in the experiment trials, and two colleagues tested the full experiment and said that they did not notice the splicing.

Additionally, regarding the practice trials, the experimenter should carefully consider who to record for the model talker during this portion of the study. In Enzinna 2018, the model talker that participants heard during the practice trials was the pre-recorded voice of the experimenter (Naomi Enzinna). Her voice was used during the practice trials for several reasons: First, even though this could influence the participants' speech (Hay et al., 2006), all participants had to interact with her before the study in order to receive instructions, and thus using her voice did not add any new factors. Further, she was both a Miami native and an Ithaca resident at the time of the study, and therefore she fell into both of the target speech communities.

2.4 Non-linguistic materials

In order to run BARCoT, the following materials are required: a computer (a laptop if recording in different locations), a headset with audio and recording capabilities, a mouse,⁶ MATLAB,⁷ Praat, and the Montreal Forced Aligner (McAuliffe et al., 2017).

2.5 Randomization and ordering

BARCoT was designed to randomize and counterbalance the stimuli in ways relevant to the purposes of the experiment in Enzinna 2018. First, BARCoT counterbalances the order in which participants hear model talkers, when there is more than one model talker. In Enzinna 2018, for example, half of the participants heard the monolingual model talker first, followed by the bilingual model talker; and half of the participants heard the bilingual model talker first, followed by the monolingual model talker.

Second, one word-pair containing each target consonant appears on each board. In Enzinna 2018, there were six target consonants: /p, t, k, l, t, d/. Thus, there were six word-pairs on each board. Third, half of the word-pairs on each board are primed with a target prime, and half are primed with an unrelated prime; and priming of a target consonant alternates by board. For example, in Enzinna 2018, if /p/ was primed by a target prime on one board, then on the following board it would be primed by an unrelated prime. Last, all words were randomized for each participant. This randomized order was then repeated with each model talker.⁸

⁶A laptop's trackpad can be used, but is less preferable.

⁷The program was written with MATLAB R2017b.

⁸I asked several participants, after they were done with the entire experiment, whether they could tell that the words appeared in the same order for both model talkers, and they said that they could not tell because there were too many words to keep track of.

Additionally, in Enzinna 2018, the ordering of the word-pairs was also counterbalanced. In the word-pairs, each target consonant occurs twice with each target vowel (/I & & A i & a o u/), creating two target words. As discussed in Section 2.2, these two target words are then primed by a target prime and an unrelated prime, creating in total 4 word-pairs. For example, /p/ and /a/ occur in the target words POBBY and POFFY, which are both primed by POSSE (target prime) and WHY (unrelated prime). These combinations create the following 4 word-pairs: POSSE-POBBY, WHY-POFFY, WHY-POBBY, and POSSE-POFFY.

These word-pairs were then split across two halves of the study, so that each target word occurred only once per half. For example, if POSSE-POBBY and WHY-POFFY occurred in the first half of the experiment, then WHY-POBBY and POSSE-POFFY occurred in the second half (and vice-versa). This was done to create some distance between each time a participant produced a target word. The ordering of the halves of the stimuli were counterbalanced, so that half of all participants saw POSSE-POBBY and WHY-POFFY first, and the other half saw WHY-POBBY and POSSE-POFFY first. The BARCoT program is designed to randomize word-pair stimuli to meet these criteria, but can be altered for other needs.

2.6 Data processing

BARCoT records all participants' speech, board information, and click times. The click times and board information are saved in tables, which can then be used with a Praat script to create Text Grids with boundaries after each response. Those Text Grids and their matching sound files, along with a dictionary containing all of the words used in the study (i.e., all of the words in the word-stimuli spreadsheet, plus those in the frame sentences) and their pronunciations, can then be used with the Montreal Forced Aligner (MFA) to segment the speech. Enzinna (2018) used the MFA with a pre-trained acoustic model trained on English. After checking the alignments, the segments can then be extracted from the Text Grids and analyzed in MATLAB, R, etc.

3 Future use

In the future, we hope to utilize the BARCoT program further in our own research. We plan to use BARCoT to examine phonetic accommodation in new locations, with different model talkers, by different participant groups, and with different linguistic variables. We also plan to further examine the time-course of accommodation by analyzing the data collected with BARCoT in new ways. Additionally, we hope that other researchers adapt the BARCoT program for their own research needs. If you are interested in acquiring or learning more about BARCoT, please contact the corresponding author.

Acknowledgments

We are grateful to Draga Zec, Abby Cohn, and Marie Huffman—Naomi Enzinna's dissertation committee members (in addition to co-author Sam Tilsen)—for their contributions to her dissertation research. Thank you to the Cornell Linguistics Department, the Cornell Phonetics Lab, Bruce McKee, and the Cornell Linguistics graduate students, for their contributions to this work, in regards to helping with the recordings, testing the experiment, providing feedback on the design, and more. A special thanks to Daniel Scarpace for helping to write the Praat script described in Section 2.6.

⁹It should be noted that this model does not segment VOT separately from the preceding stop. Thus, in her dictionary, she added an /h/ (HH) after all voiceless stops, and the MFA segmented the aspiration as if it were an /h/. After the MFA aligned the speech, she checked all of the alignments for errors. She does not recommend this method for automatic alignment of VOTs, but chose this method because it allowed her to align all of her speech data, which will be useful for future analysis of all of the remaining dependent variables.

References

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. Language in Society 39:437–456.

D'Imperio, Mariapaola, Rossana Cavone, and Caterina Petrone. 2014. Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology* 5:1226.

Enzinna, Naomi. 2018. Automatic and social effects on accommodation in monolingual and bilingual speech. Doctoral dissertation, Cornell University.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. Psychological Review 105:251.

Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458–484.

Hwang, Jiwon, Susan E Brennan, and Marie K Huffman. 2015. Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language* 81:72–90.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of Interspeech*, 498–502.

Stollenwerk, Debra A. 1986. Word frequency and dialect borrowing. In *OSU WPL 34*, 133–141. Ohio State University. Department of Linguistics.

Appendix A

Provided below are images of the instructions that participants received, which were displayed on a computer screen through the BARCoT program.

Click the NEXT button to begin.
NEXT

On your screen, you The colored squa occurring next t	res will appe	ear in pail er. These s ple scree	rs, with tv square pa	vo square irs will be	s of the s
		fubby			
mace	βλ	north			
migh	nt		morph	sibby	meeb
					about
		died		seg	noisy
		chab			
	ВАСК			NEXT	

In this study, there are two roles: the role of the Helper and the Matcher.
You have been assigned the role of the Helper.

You will complete a task with the Matcher.
The Matcher will speak to you over the headset.

BACK

NEXT

	Matcher's				
		north			
might			morph		
					about
		died			noisy
	ВАСК			NEXT	

Your goal is to help the I your screen. To do this						
			north			
m	might			morph		
						about
			died			noisy
	ВА	ск			NEXT	

For example, if your screen were the one below, you would hear the Matcher ask, "What is by the word NORTH?" Nearby NORTH is the word FUBBY so you would respond by saying, "FUBBY is by the word NORTH." fubby north macey morph sibby might about died noisy chab ВАСК NEXT

Note that NORTH and FUBBY are in squares of the same color. This means that these words are paired. \\

The Matcher will always ask you for a word by referencing the other word in a pair. This means you should respond with the word that shares the same color square as the one referenced.

	fubby			
macey	north			
might		morph	sibby	meeb
				about
	died		seg	noisy
	chab			

ВАСК

NEXT

Some of the words on the board are real words, and some are not. know how to say a word, don't worry. Just say it however feels nature fully fully macey north might morph sibby meeb died seg noisy chab						
macey north might morph sibby meeb about died seg noisy						
macey north might morph sibby meeb about died seg noisy	Some of the words o	on the boa word, don'	rd are rea 't worry. J	al words, ust say it	and some however	are not. feels nat
macey north might morph sibby meeb about died seg noisy						
macey north might morph sibby meeb about died seg noisy						
might morph sibby meeb about died seg noisy		Τ	fubby			
died seg noisy	macey		north			
died seg noisy	might			morph	sibby	meeb
						about
chab			died		seg	noisy
			chab			
	F	ВАСК			NEXT	
BACK NEXT]		
BACK						

If you accidentally click on the wrong square, the board will flash red three times. If this happens, click on the correct square to move on to the next word. $\label{eq:correct}$ If you accidentally said the wrong word, please repeat the phrase with the correct word (e.g., "FUBBY is by the word NORTH") before clicking. If you would like the Matcher to repeat what they just said, right click anywhere on the screen. fubby north macey sibby might morph about died noisy chab BACK NEXT

To practice this, you will first complete 3 practice boards. The Matcher for the practice boards is the experimenter, so you will hear her voice over the headset.

Then you will complete 72 boards with 2 different Matchers (36 boards per Matcher).

NEXT

ВАСК

When you are ready, click the NEXT button to begin the practice boards. Note: The board may take a few seconds to load. Please be patient.	
BACK	

Department of Linguistics Cornell University 203 Morrill Hall 159 Central Ave. Ithaca, NY 14853 nre23@cornell.edu tilsen@cornell.edu