Segmental targets versus lexical interference

Production of second-language targets on first exposure and the result of minimal training

Susanne E. Carroll & Joseph W. Windsor
University of Calgary

This chapter reports on a study of Anglophone adults exposed to German for the first time. Production data are examined and compared to the stimuli participants were exposed to. Data provide clear evidence of lexical influences on speech behaviour; L1 words interfere in production and provide a context for sub-lexical contextual effects at the syllabic and foot level, as well as at the segmental level. We discuss vowels and /r/-sounds in particular. Our data are consistent with claims that speech perception involves a dynamic interaction between objective properties of the speech signal, the structure of the L1 lexicon, and the L1 phonetics-phonology interface. Crucially, we show that on first exposure, learners are already sensitive to L2 sounds not part of the L1. This will serve as a basis for learning novel phones, but lexical competition clearly constrains the learning process.

Keywords: transfer; first exposure learners; phonological features; syllable positions; rhotics

1. Introduction

It is widely assumed that accurate perception of L2 speech sounds is a prerequisite for accurate L2 speech production (Escudero 2005, 2006; Flege 2003). In other words, learners must form sound representations on the basis of the input they are exposed to. The study of the acquisition of an L2 sound system thus proceeds on several fronts: perception and production studies, descriptive work on L1 and target language phonetics and phonology, as well as developmental research which seeks causal accounts of behavioural changes in learners in terms of mechanisms of perception and learning. One common thread is that experience interacts with L1 knowledge in shaping what L2ers acquire and when. In the very large literatures that have focused on L2 phonetic and phonological acquisition, perhaps the topic that has been the least studied to date is how experience shapes learner

DOI 10.1075/hsld.4.03car

behaviours at the earliest stages of acquisition. In this chapter, we present some initial production data from "first exposure learners", gathered in response to specific stimuli that allow us to make a modest contribution to this topic. Our goal is to show how the L1 lexicon constrains perception of the stimuli and interacts with L1 phonological knowledge in determining what the learner produces.

Many studies have shown that knowledge of a specific sound system shapes the perception of non-native phonetic and phonological distinctions. For example, adult non-native listeners, exposed to a novel language, have difficulty categorizing and discriminating phonetic contrasts that are not distinctive in their L1 (Lisker & Abramson 1967; Polka 1992; Werker & Tees 1984; among many others). However, contrastiveness turns out not to be a good predictor of the ability to perceive phonetic distinctions; non-native listeners can detect some L2 contrasts even better than listeners who have the contrast as part of their L1 phonological repertoire (Best & Strange 1992; Best et al. 2003, inter alia). For a review of the literature on non-native perception, see Best & Tyler 2007, de Jong et al. 2009.

Many studies demonstrate that word recognition is language specific and that various factors play a role in how we "hear" words (Cutler 2012). For example, native language prosody can play an important role in word recognition. Speakers of a language where stress patterns differentiate words (e.g. English *CONtent* – Noun – versus *conTENT* – Adjective) become sensitive to the patterns of stress in early infancy (Curtin et al. 2005; Jusczyk 1997; Jusczyk et al. 1993) while speakers of a language where stress is not distinctive may exhibit "stress deafness" and be unable, on first exposure to another language, to recognize words that differ only in stress placement (Dupoux et al. 1997; Peperkamp & Dupoux 2002). Moreover, this problem can persist even after substantial amounts of exposure to the target language as individuals acquire more advanced levels of L2 proficiency (Tremblay 2008).

In addition, knowledge of the language-specific patterns of consonant and vowel sequences (phonotactics) also determine how we "hear" words. Infants early on acquire sensitivity to the distribution of consonants and vowels in words and syllables (Hohne & Jusczyk 1994; Jusczyk et al. 1994; Mattys & Jusczyk 2001a, b; Mattys et al. 1999; Chambers et al. 2003). Phonotactics play an important role in explaining patterns of word recognition in native-speaking adults (Brent & Cartwright 1996a,b; Li 1996; McQueen 1998; Mattys et al. 2005). Phonotactics also play a role in speech production (Vitevich & Luce 1999; Goldrick & Larson 2008). Moreover, native-language phonotactics affect the perception of L2 segmental contrasts (Flege & Wang 1989; Tench 2003; Weber & Cutler 2006, among many others), and they constrain distributional learning of novel distinctions (Finn & Hudson Kam 2008).

Despite such strong evidence for L1 effects in perception and production, L1 knowledge is not a straitjacket preventing adults from becoming proficient users of an L2. However, the relationship between L1 knowledge, amount of exposure and phonetic and phonological learning is anything but clear. As we shall see, even on first exposure to an L1, after minimal amounts of input and no prior practice, adults can segment novel sound forms and attempt to produce novel L2 sounds. How learners progress from this state to a target-like pronunciation is not well understood. It is known that, even after substantial amounts of exposure to an L2, the amount of daily L1 language use of an L2 learner plays a role in the development of native-like ability to distinguish L2 vowels and consonants, (Flege et al. 1995; Flege et al. 1997; Flege et al. 1999; Guion et al. 2000; Piske et al. 2001; Piske et al. 2002). This suggests limitations on the ability of even child L2 learners to keep separate the segmental categories of their two languages. See Piske 2007 for a critical review. This is consistent with a claim by Best and Tyler (2007) that perceptual learning occurs "early on" in L2 learning, before learners have acquired substantial amounts of L2 vocabulary. In addition, Best and Tyler speculate that exposure to words (especially exposure to minimal lexical pairs) plays a crucial role in leading learners to establish functional equivalences for distinct sounds that ought to lead to the emergence of new phonological categories. Determining how early "early on" is, is a task still facing the field.

In this chapter we explore through a small multiple-case study the initial production of German words produced by first exposure learners trained on a precise set of stimuli. First exposure learners are *not* non-native listeners or "functional monolinguals" in the sense of Best and Tyler (2007). While our participants had had no systematic prior exposure to the target language and were not studying it at the time of the experiment, they were mostly not monolingual Anglophones. Thus, "prior" knowledge includes not only the L1 but also any phonetic or phonological categories, and prosodic words that form their knowledge of other languages.[1] Discussion of the possible effects of prior exposure to other languages goes beyond the limits of this study but familiarity with other L2s can offer learners phonetic variants as possible targets for segments recognized as functionally equivalent, such as, e.g. Spanish tapped [r] for the German uvular fricative [ʁ].

We illustrate three distinct sources of L1 influence on initial L2 production: words, syllables, and the repertoire of segments (see Figures 1 and 2 below for our assumptions of the phonemic inventories of English and German). Each of these factors has been studied separately in research dealing with fluent bilinguals (see Cutler 2012). We shall demonstrate that first exposure learners on hearing *cognate words*, defined here as words that sound similar to specific L1 words and are functionally equivalent to them, tend to produce the L1-word sound form even

1. The emphasis on monolingual participants in many perceptual studies is a matter of aiming for methodological rigour that one can hardly object to. However, if it is true that bilingualism and multilingualism are a fact of life in most parts of the world, second language acquisition work should include participants who reflect that reality and explore the consequences.

when it deviates substantially from the stimulus. Some of the production data provides evidence that L1 words have been activated by the L2 stimuli and the sound forms of these L1 words are the basis for the sound sequences the participants are attempting to produce. This means that the learners may be less accurate in their production of cognate items than non-cognate items even when the segments of the words consist of functional equivalents of L1 sounds (and thus should be relatively easy to perceive and produce).

	Bi-labial		Labio-dental		Inter-dental		Alveolar		Post-Alveolar		Palatal		Velar		Uvular		Glottal	
Plosive	/p/	/b/					/t/	/d/					/k/	/g/			/ʔ/	
Nasal		/m/						/n/						/ŋ/				
Fricative			/f/	/v/	/θ/	/ð/	/s/	/z/	/tʃ/ /ʃ/	/dʒ/ /ʒ/	/ç/				/χ/	/ʁ/	/h/	
Approx.		/w/						/ɹ/				/j/						
Lateral								/l/										

Figure 1. German[2] and English phonemic consonant inventory

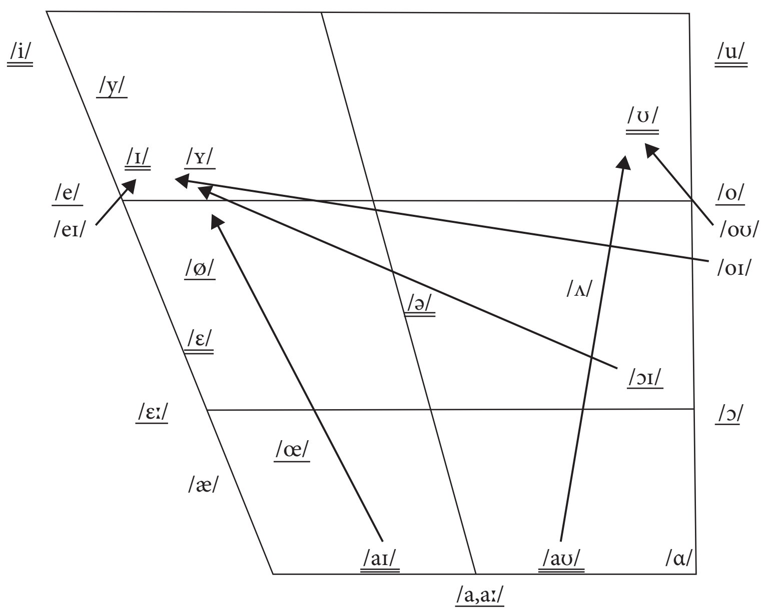

Figure 2. German and English phonemic vowel inventories

2. The German phonemic system is based on Kohler (2010).

Phonemes which have a double-underline exist in both languages, those with a single underline are phonemes in German which English does not possess, unmarked phonemes exist in English, but not in German.[3]

In order to pronounce non-cognate words, by hypothesis, learners must create novel phonological representations because their L1 lexicon has no lexical equivalent. However, on first exposure, learners are presumably not in a position to create novel words from unfamiliar L2 segments. We shall provide evidence that our participants struggled with German rhotics. This is not a surprising result because, while the English rhotic is relatively stable in its realization as [ɹ], German rhotics vary by syllable position and are rarely – if ever – realized in this manner (Hall 1993; Eisenberg 2004).[4] German also has front rounded vowels; our stimuli include instances of [ø] and [œ]. Since Canadian English does not include front rounded vowels, these caused pronunciation difficulties and the participants in this study showed a wide range of repair strategies – some trying to maintain the [FRONT] feature and relying on an L1 vowel, others trying to maintain the feature [ROUND] and specifically shifting their pronunciation away from an L1 category. This variably resulted in [ɪ], [ɛ], or [ɞ], for example. However, it should not be assumed that where the L2 provided close equivalents to L1 phones that these were approximated. We shall provide some data that suggests that even in the case of the non-cognate forms, the sound forms of extant L1 words influence the production of a novel L2 word. Thus, we find lexical effects even in the case of the non-cognate words which cause some of the learners to deviate from the stimuli in ways not predictable from a simple comparison of consonant and vowel repertoires. Indeed, some of our data are consistent with the idea that we store sub-lexical sound sequences (syllables or syllabic sequences) in long-term memory and phonological planning draws on such representations (Levelt 1999; Cholin et al. 2004).

3. Canadian raising is not indicated here. The phonemes of German are taken from Kohler (2010). The reduced vowel in German is often claimed to be 'deep schwa' ([ɐ]), however, for ease of comparison, we use the vowel space outlined by Kohler.

4. These are taps [ɾ] or approximates [ʁ̞] inter-vocalically, uvular fricatives [ʁ] in onset position and in certain consonant clusters which may be realized as voiceless [χ] after voiceless obstruents, or, post-vocalically or word-finally, may be vocalized to deep schwa [ɐ] with little-to-no 'r'-colouring or rhoticization of the vowel (see Kohler 2010 for a brief discussion of the German phonological system). There does not seem to be consensus in the literature over whether this vocalic allophone is part of a branching nucleus or not – while some researchers claim that this allophone is non-syllabic, Kohler (2010) maintains that the vocalization of this rhotic leads to a diphthong.

Other productions show consistent L1 effects within syllables in terms of L1 phonotactics. We will also observe novel, L1-inconsistent syllables, presumably arising from attempts to come up with a compromise between L1 gestures and the sounds detected. Thus, several of our participants produced alveolar tapped [r]-sounds on exposure to the German uvular fricative. The result was not stimulus-true but it was not a typical English approximant [ɹ] either. These data suggest that the learners are aware of the need to produce something else. Of course, in attempting to produce words containing rhotics and fronted vowels, our participants tend to be less consistent in their pronunciations than in their productions of words where the entire word, entire syllables, or individual segments are assimilated to L1 forms. The development of a consistent target-like pronunciation might well require modelling, feedback, or meta-linguistic training (Pisoni & Lively 1995; Bradlow et al. 1997).

2. The study of "first exposure" learners

Much mainstream SLA research investigates L2 learners at intermediate, advanced, or even native-like levels of proficiency. There has been far less attention to the initial stage of L2 acquisition. It would be misleading, however, to say that we know little about the initial-stage learner. This is because, as noted above, the literature on non-native perceptual learning is large and directly relevant to assumptions we might make about the initial state of knowledge and how it guides learners in processing input on first exposure. Still, there are caveats to make. The literature on non-native perceptual learning systematically draws on individuals who are not only naïve to the target L2 but also naïve to any language other than their L1. We suspect that such individuals may not be typical of the population of second language learners. More importantly, the standard methodology used in studies of perceptual learning involves stimuli that bear little resemblance to the kind of continuous speech that language learners typically learn from. First exposure learners are not, not even in the foreign language classroom, presented with words in isolation (something that would facilitate forming a representation of the sounds of the word).

Nor are they typically asked to discriminate minimal-pairs (something that would make listening very difficult). Thus, we should exercise caution in generalizing from such experimental studies to make inferences about the course of L2 acquisition. Accordingly, within mainstream SLA, a number of researchers have begun to explore learning at the very initial stage of L2 acquisition in order to understand how both prior knowledge of a language and exposure to L2 input constrains L2 learning.

So far most of the first exposure studies have dealt with the L2 learner's ability to segment prosodic words from continuous speech in some kind of meaningful context. Using different methodologies for the presentation of L2 stimuli, different L1/L2 language pairs, and different means of measuring learning, a small body of work has repeatedly shown that first exposure learners can indeed segment L2 sound forms on the basis of very little input, and regardless of the degree of similarity between the sound systems of the target L2 and the L1 of the learners (Rast 2008, 2010; Gullberg et al. 2010, 2012; Carroll 2012). First exposure learners are simultaneously attentive to gesture (when visually processing the situations in which speech is produced) and to prosodic cues to focus since gesture facilitates segmentation (Gullberg et al. 2010). The initial or final words of an utterance are recognized much more readily than words presented utterance-medially (Shoemaker & Rast 2013). Shoemaker and Rast (2013) attribute this to sensitivity to the edges of prosodic constituents, meaning that even first exposure learners are constructing higher-order prosodic constituents. Results from the current study are consistent with this claim. However, none of this work has yet asked the question of how phonological contrasts of the sort typically studied in the perceptual learning literature begin to emerge from the stored representations (phonetic or phonological) that result from initial segmentation of the L2 signal. How much lexical acquisition might be required for such contrasts to emerge is an important question raised by Best and Tyler (2007). Equally important is the question of how the structure of the L1 lexicon, for example, the density of lexical neighbourhoods, might constrain both perception and production of the L2 (see Luce et al. 1990, and Vitevitch & Luce 1999 on lexical neighbourhoods in spoken word recognition).

The existing literature on first exposure learners provides considerable evidence for lexical effects on segmentation. An L2 word that phonologically resembles an L1 word (*cognate words*) will activate that word, leading to faster responses on a word recognition task (Carroll 2012).

Participants are also more accurate in identifying target words on a forced-choice word identification task (see below) when those words are cognate. Shoemaker and Rast (2013) report similar results.[5] Thus, the organization of the

5. A comment on terminology is needed here. Rast (2008, 2010), and Rast and Dommergues (2003) preferred to speak of words that are "phonologically transparent". We prefer to speak of "cognate" versus "non-cognate" words because the terms will be familiar to many psycholinguists who have studied patterns of lexical activation among second language learners and bilinguals. In this literature, there is no assumption that "cognate words" must be historically related (indeed, the term has been used of nonce forms, see, for example, Costa et al. (2000). Secondly, even Rast has assumed more than phonemic relatedness. Words that she treats

L1 lexicon plays a crucial role in segmenting continuous speech in real time. Our production data will provide evidence that it also plays a crucial role in forming the initial representations of the L2 that get stored in long-term memory, representations that presumably will form the basis for the incremental acquisition of the target system's consonants and vowels.

3. The current study

3.1 Methodology

In this paper we analyse production data gathered during a study focusing on participants' receptive abilities (Carroll 2012, 2014). The original study involved participants responding to stimuli with button presses; we measured their response times and accuracy. After initial training and testing, we also gathered production data at three different points in time. It is a subset of this production data that will be the focus of this paper. Since we are crucially interested in the relationship between the input and what the participants said, we will provide some detail on the procedures of the original study.

3.2 Procedures and stimuli

Participants were tested individually at a computer station. On arrival, they filled out a background questionnaire and were then assigned to one of two experiments (the order of which was counter-balanced across participants).[6] In the "cognate" experiment, participants were exposed to stimuli containing cognate target words. In the "non-cognate" experiment, participants were exposed to stimuli containing non-cognate target words. See Appendix 1 for examples of the stimuli used in both experiments of the study. The main task of each experiment was a forced-choice word identification task. In the instructions given at the beginning of the main task (presented aurally and in writing on the computer screen in English), participants were told that they would see pictures of individuals and they were to learn their names. Participants then heard 20 presentational sentences that both contained

as phonologically "transparent" are precisely (Polish) L2 word forms that are semantically related to L1 (French) forms, a relationship that allows Rast's participants to guess correctly the meanings of the target (Polish) words. Moreover, phonological relatedness in her studies is a gradient notion and it is not clear what "closeness" is required for phonological transparency. This strikes us as potentially circular. See Friel & Kennison (2001).

6. The questions asked were about their home language, any languages they had studied, their regular language use, time spent living in a country or region where another language was spoken, and so on.

target words and introduced the people depicted in the pictures. These statements were immediately followed by 20 questions (also containing the target words) to which they responded by pressing a dedicated button on the computer keyboard. This constituted the first training trial. Training was followed by a test where participants heard only the questions. However, before being permitted to proceed to the test, participants had to correctly match all 20 names to the appropriate pictures. Some participants were able to proceed to test after a single exposure to the statements and questions; some required 10 training trials; most fell in-between. Following the test, participants had a two-week hiatus and then returned to the laboratory for two more exposures (re-test1 and re-test2). We gathered production data at test, re-test1 and re-test 2. See Table 1 for background data on the participants and the number of training trials each one required.[7]

Table 1. Participant information

Participant	Age	Birthplace	Home language in childhood	Languages other than English	# of training trials (cognate items)	# of training trials (non-cognate items)
01F051107	19	Canada	English	French	3	7
01F081107	23	Canada	English	French	4	5
01F151107	26	Canada	English	French, Spanish, Italian	2	5
01F261107	22	Canada	English	Latin	5	7
02F061107	22	United Kingdom	English	Dutch	2	3
04F021107	26	Canada	English + Mandarin	Mandarin, French	3	6
Mean	23			1.5	3.2	5.5

7. Participant labels indicate sex ("F" = female; "M" = male) and the date of testing. "01F051107" was a female participant tested on November 5th, 2007. The "01" at the left edge indicates that this participant was the first one tested on that date. Many of the participants who were born in Canada were, in fact, born and raised in Alberta. Our recruitment notices stipulated that the participants in our sample should not know German. Our questionnaire confirmed that none had been exposed to German in the home, through instruction, or by living in a German-speaking part of the world. As Table 1 shows, none of the participants in our sample were "functional monolinguals". Indeed, among the 33 individuals we tested, only two were monolingual English-speakers, a pattern typical of younger Canadians more generally. However, our question asking about "other languages" asked about any language to which the participant had been exposed without requiring proficiency. Many of our participants were not highly proficient in the "other" languages they had studied.

As Table 1 shows, participants required more training trials to correctly map all 20 non-cognates to the pictures. Split-half reliabilities conducted on the complete data set from Training Trial1, showed that participants were just as accurate on the first stimuli they were exposed to as the last of the list. Inspection of the data showed that most participants who required more training were making only one or two errors in the lists. The production data confirm this. We infer from these results that segmentation of the target words was easy.

Since we are interested in examining the effects of exposure on production, we should explain how to interpret the right-most columns of Table 1. A participant who took 2 training trials to criterion had heard the target forms four times when they performed the test that preceded their first recording. By re-test2, this participant had heard the target forms seven times. A participant who needed three training trials to criterion had heard the target forms six times when they performed the test. By re-test2, this participant had heard the target forms nine times. And so on.

The two experiments used both visual stimuli and auditory stimuli. Visual stimuli consisted of 80 coloured line drawings that were designed to depict 40 individuals (two pictures of each individual).[8] All pictures were stored as TIFF files which were then used in E-Prime programming. 40 pictures were selected for each experiment (= 20 individuals). No picture was repeated across experiments. See Appendix 1.

As noted, auditory stimuli consisted of two kinds of utterances: statements and questions. All stimuli were first written out and then verified by a native speaker of standard German for accuracy. This individual also recorded all of the statements and questions in one of the laboratories of the Language Research Centre of the University of Calgary. Recordings were verified by the first author and particular statements or questions were re-recorded when necessary. Recordings were then digitalized.

Both statements and questions were used during the training trials. Only the questions were used at test and re-tests. As noted, the training trials consisted of two phases: participants heard a list of 20 sentences while looking at the appropriate picture for that individual. They then heard the corresponding questions to these sentences (again looking at the appropriate picture) and they then responded to each question by keying in a response: F1 if the target name was the first name in the conjunction, F12 if it was the second name in the conjunction. In all cases, one of the names matched the picture, meaning that the listener's task was to select

8. In some cases, an individual was shown in two postures, for example, a frontal exposure and an exposure in profile. In other cases, the individual was shown holding an object versus not holding that object. See Appendix 1.

the correct sound form and reject the inappropriate one. Responses were automatically recorded and at the end of the 20th question an accuracy score appeared on-screen. If participants correctly identified all 20 names, they proceeded to the test phase. If they failed to correctly match all 20 names to the 20 pictures, they re-did the training trial to a maximum of 10 attempts. If a participant failed on the 10th training trial, they were released from the experiment (and no production data were collected from these participants).

During the test, participants looked at the second picture of the individuals and heard the same 20 questions. When participants returned for Session 2, they saw the original pictures and heard a different version of the questions with the order of the names reversed. Thus, across the three major phases of each experiment (training trials, test and re-test), the stimuli pairs were always unique (either the pictures or the auditory stimuli varied). This was to encourage more abstract processing of the stimuli. There were repetitions of the same exemplars from one training trial to another and from re-test1 to re-test2. In all presentations of the stimuli lists, utterance-picture pairs were fully randomized.

Each statement contained one target name in one of four syntactic constructions used to introduce people or objects into a conversation. See (1) which contains examples of cognate words (C) used in the cognate experiment (1a, b) and non-cognate words (N) used in the non- cognate experiment (1c, d).

(1) a. Hier ist Agnes (C)
[ˌhiːɐ ɪst ˈagnəs]
'Here is Agnes'

b. Da steht Claudia (C)
[ˈdaː ʃteːt ˌklaʊdia]
'There stands Claudia'

c. Hier sehen Sie Lutz (N)
[ʻhiːɐ zeːən zi ˌlʊts]
'Here see you Lutz'

d. Das ist Annegret (N)
[das ˌɪst ʻanəgʁet]
'That is Annegret'

In each question used to evoke a forced-choice decision, the target name was embedded in a conjunction along with a name that shared some phonological properties with it. Note, however, that the foils were not minimal pairs.[9] See (2).

9. Thus, our methodology differed significantly from that typically used in perceptual learning experiments where non-native listeners are required to discriminate acoustically minimal pairs. A subsequent study in which we used minimal pairs that differed only in stress

(2) a. Ist hier Agnes oder Angela?
Is here Agnes or Angela?

b. Steht da Charlotte oder Claudia?
Stands there Charlotte or Claudia?

c. Sehen Sie Hier Lutz oder Ludo?
See you here Lutz or Ludo?

d. Ist das Annegret oder Annika?
Is that Annegret or Annika?

Production data were recorded on an Edirol R-09H digital recorder located on the desk in front of the participant. Sound quality was good.

3.3 Initial analysis of production data and stimuli

Production data were subsequently transcribed by an undergraduate student of linguistics using the International Phonetic Alphabet and information on German and English sound systems from the phonological and phonetic literature and various other sources.[10] Accuracy of transcriptions was checked by a different student of linguistics. Once a subset of 6 participants had been selected for this study, the transcriptions were checked again by the first author. This resulted in 720 data points. There were numerous discrepancies across the transcribers. Accordingly, for our case studies, we used any transcription where two of the three transcribers agreed.[11] Where each of the transcribers used a unique transcription, the transcription of the first author was selected for analysis.

placement or phonotactics showed that none of the participants were able to match all 20 names to the pictures. Given the high performance level with the items described in the text, this suggests that minimal pairs are highly confusable for non-native listeners (and make severe demands on working memory) while target pairs that differ metrically in several ways can be readily distinguished.

10. For example, in doing the third transcription, the first author repeatedly checked participant pronunciations against the sound files from the University of Los Angeles website: ⟨http://www.linguistics.ucla.edu/people/hayes/103/Charts/VChart/#TheVowels⟩

11. Across the three transcribers, we have at least two identical transcriptions for approximately 60% of the data. For the remaining cases, many of the discrepancies involved missing data because Transcriber 1 did not transcribe any response that she judged to be the "wrong response". This was normally a pronunciation of the foil. Transcriber 2 also failed to record a transcription for many responses. Where there were actually three transcriptions, discrepancies involved the systematic use of an inappropriate rhotic symbol or disagreements as to whether the final syllable of targets like *Anja* or *Heike* involved [a] or [ə]. We readily acknowledge that transcription of the approximative pronunciations typical of L2 learners involves

Transcriptions were also made of the stimuli. Again, there were three transcriptions made by the same individuals who transcribed the participants' productions. In this case, there were fewer discrepancies among the transcribers with agreement on 70% of the words. Again, discrepancies involved the correct way to transcribe German rhotics, especially in syllable-final position. Where discrepancies did exist in the transcriptions, we used the same procedure as with the participant data.

These transcriptions were then used to code for various factors in two separate statistical models – one for rhotics, and one for vowels. The production data of our six subjects for rhotics were coded for set (cognate or non-cognate), trial number (test, retest 1, or retest 2), whether or not the response was correct, syllable position of the rhotic in the target,[12] the place and manner of the target, whether the response was on-target or not, whether there was any rhoticism evidenced in the production, and whether or not the response was English-like or not (i.e. whether their production was an alveolar approximate, or if they produced something non-English like a tap, trill, or uvular fricative). For the vowels, this data was coded for set (cognate versus non-cognate), test (test, retest1, retest2), whether or not the vowel belonged to a stressed syllable or not, whether the target syllable had a onset, whether or not the target syllable had a coda, phonological features of the target vowel ([FRONT], [HIGH], [LOW], [RTR] (retracted tongue root), or [ROUND]), whether or not the target syllable contained a rhotic, whether or not the target vowel was an English phoneme, and finally – whether or not the production was target-like. In these tests, we were interested in finding out if there were any significant predictors of the last data point in the above list: whether productions were target-like. Using these factors, we ran an analysis of variance (ANOVA) on a generalized linear model (GLM) to look for any significant main

precisely the same kinds of perceptual distortions discussed in the literature on non-native speech perception, due to lexical effects and segmental category assimilations.

12. We assume that vocalized rhotics are the result of a [RHOTIC] feature in the syllable margin (see Note 4), and so these instances of rhotics were coded as codas. It is useful to note that where [ɐ] is an allophone of the underlying coda rhotic /ʁ/ little-to-no rhoticization (lowering of F3) was observed on any of these vowels. Despite this fact, the subjects in this study were often able to recover the [RHOTIC] feature, often leading to pronunciations that contained rhoticized vowels such as the variable pronunciation of *Eberhard* as [ˈe.bɚ.hɑɹt], [ˈe.bəɾ.hɑɹt], or [ˈe.bəɹ.hɑɹt] when the target form that was heard was [ˈe.bɛ.haɐt] (we could not notice any rhoticization on the target [ɛ] vowel which facilitated our subjects reconstructing a rhotic for this syllable. It is more likely that this vowel was half long ([ɛˑ]), which listeners interpreted as a "dropped r" as may be the case for some British or New-England varieties of English.

effects or interactions of these factors in determining how target like our subjects productions were, and what influenced those correct productions.

3.4 Results

Earlier work (Carroll 2012, 2014) showed that segmentation of various kinds of German words by English-speakers using this methodology is easy. In the case of the cognate/non-cognate experiments, accuracy rates from all 33 participants on Training Trial1 were significantly above chance on both cognate items (82.7%) and non-cognate items (70.6%); one-sample means comparison tests for cognates: [$t(23) = 36.89$, $p < 0.001$]; for non-cognates: [$t(32) = 30.97$, $p < 0.001$]. A two-sample t-test on the difference in means on the cognate versus non-cognate items showed that the participants were significantly more accurate on the cognate items [$t(55) = -4.09$, $p < 0.001$]. This pattern is partly reflected in the responses of our subset of participants (cognates = 79%; non-cognates = 84.2%). All but two participants obtained higher accuracy scores on the cognates than on the non-cognates.[13]

Out of the original 33 participants, 10 were not able to learn all 20 items on the non-cognate experiment. We used data from the remaining 23 participants in looking at how many trials were needed to learn all 20 names in both experiments. It took the 23 participants 3.04 trials on average to correctly map all 20 cognate names to the pictures. In contrast, it took the same participants 5.56 training trials on average to correctly map all 20 non-cognate names to the relevant pictures. A paired t-test on the number of training trials to criterion showed that this difference in means was significant [$t(22) = -5.61$, $p < 0.001$]. The number of training trials for our subset of participants is virtually identical. See Table 1. From these data we can conclude that participants were initially processing the cognate items differently from the non-cognate items and did so in a way that facilitated the word recognition task. However, once participants had learnt all 20 items, there were no differences in their performance on the two word types. See Carroll (2012) for details.

In short, we can assume that the input was perceptually segmented, that our participants accordingly had a perceptual basis for producing the words, and that they were processing the cognate words differently from the non-cognate words. See Dijkstra et al. (1999) for discussion of how cognate words are processed.

Inspection of the transcriptions reveals that the accuracy data from the button presses is confirmed by the attempts to pronounce the words. Participants were overwhelmingly accurate in selecting the target item to pronounce, as opposed to

13. Participant 02F061107 scored 80% on the non-cognate first training trial and 70% on the cognate words; participant 04F021107 obtained the same score on both lists (90%).

the foil. Ignoring segmental errors and the occasional stress retraction (from e.g. *An'dreas* to *'Andreas*) participants produced recognizable target words 90.5% of the time. In the case of the cognate words, there were 28 errors which included pronouncing the foils, misparsing the target (e.g. *Kai* [kaɪ] became [məkaɪ]), and other miscellaneous errors. In the case of the non-cognate words, there were 40 errors involving the same categories (12 misparses, 4 pronunciations of the foils and 24 miscellaneous errors). Pronunciation of the foils mostly correlated with errors in button presses. The production data thus confirms that our participants had segmented L2 words, indeed not only the target words but also the foils.[14]

We tabulated accuracy responses based on the transcriptions of the test production data, first for stressed and unstressed syllables, and then for the words as a whole. Table 2 presents this quantitative data, organized in terms of the cognate/non-cognate status of the words, number of syllables, whether the syllable was stressed or not, where the error (if any) occurred, and what kind of error it involved (wrong vowel, weakening a vowel such as [a] to schwa, strengthening a schwa to [a], wrong consonantal onset or coda, and so on).

Table 2 shows that the least number of pronunciation errors occurred in the strong syllables of words that are phonological equivalents to English syllables, e.g. [kaɪ], [haɪ], [dit], [ni], [li], [gi], and so on. As noted, learners had problems with rhotics and front rounded vowels so words containing these sounds tend to cluster towards the bottom of the table with very low accuracy scores.

What the quantitative analyses fail to reveal, however, are pronunciations that clearly suggest that L1 equivalent words have been activated. Here is where we notice variability among the respondents. For example, on the cognate items, participant 01F051107 systematically pronounced them with an English accent. This participant was the only person to pronounce German *Johanna* [jo.ˈha.na] with an initial affricate [dʒo]. She and a second participant (04F021107) used a low fronted vowel on the second syllable [hæ], sounds readily explained if the German input activated the English name *Joanne* [dʒo'æn] or *Joanna* [dʒo'ænə].

The target German name *Josef* [ˈjo.səf] was also pronounced by 01F051107 with an initial affricate [dʒ], explicable if the input activated the corresponding English word *Joseph*. This participant was also the only one to pronounce target German *Carolina*, [kaʁolinə] whose third syllable is [li], with the English syllable

14. These data also confirm that although the task was very easy, it did require retention over time of the association of a given name to a picture mapping as provided by the declarative sentences. Thus, participants could not accurately learn to criterion just by attending to the questions that preceded their button presses. Recall that these declarative sentences were heard *only* during the training trials.

Table 2. Accuracy scores on words and syllables + error types

Name & Word Type	Mean Accuracy on Word	Mean Accuracy on Strong Syll	Error type*	Mean Accuracy on Weak Syll1	Error type	Mean Accuracy on Weak Syll2	Error type	Mean Accuracy on Weak Syll3	Error type
KAI NC	88.8%	88.8%	Vowel; onset C						
CLAUdia C	75.8%	66.6%	Vowel; misc	94.4%	Misc	66.6%	[a] for schwa		
joHAnna C	75.8%	77.7%	Vowel	88.8%	Onset C	61.1%	Vowel weakening		
HEIke NC	69.4%	61.1%	Onset-raising	77.7%	Wrong syllable; /a/ for schwa				
BEnno NC	66.6%	66.6%	Vowel; misc	66.6%	Coda C, dipthongization				
GEorg C	62.5%	66.7%	Vowel	58.3%	Vowel, Coda, R				
geZIna NC	59.2%	83.3%	Misc	0%	Vowel, onset C, misc.	94.4%	[a] for schwa		
JOsef C	58.3%	94.4%	Onset C	22.2%	Vowel, onset C				
EDmund C	55.5%	94.4%	Misc.	16.6%	Vowel weakening; coda voicing				
Eva C	55.5%	66.6%	Vowel	44.4%	Onset; vowel weakening				
FEMke NC	55.5%	33.3%	Vowel; coda, misc	77.7%	[a] for schwa				

(*Continued*)

Name & Word Type	Mean Accuracy on Word	Mean Accuracy on Strong Syll	Error type*	Mean Accuracy on Weak Syll1	Error type	Mean Accuracy on Weak Syll2	Error type	Mean Accuracy on Weak Syll3	Error type
reGIna NC	53.6%	100.0%		16.6%	Vowel weakening; /r/	44.4%	Vowel weakening		
SENta NC	52.7%	66.6%	Onset C; misc	38.8%	Vowel weakening; onset				
HEIdrun NC	52.7%	88.8%	Vowel onset weakening; onset C	16.6%	Vowel, /r/, coda				
aNIta C	51.8%	100.0%		22.2%	Vowel	33.3%	Vowel weakening, onset C		
DIETmar NC	50%	100.0%		0%	Vowel, /r/				
ANja NC	49.9%	33.3%	Vowel	66.6%	Vowel weakening, wrong syllable				
BRUno C	47.2%	0%	/r/	94.4%	Dipthongization				
AGnes C	44.4%	11.1%	Vowel; misc	77.7%	Vowel, misc.				
caroLIna C	43.0%	83.3%	Vowel	50%	Vowel	0%	/r/; vowel	38.8%	Vowel weakening;
Otto C	38.8%	11.1%	Vowel	66.6%	Dipthongization				

(Continued)

Table 2. Accuracy scores on words and syllables + error types (Continued)

Name & Word Type	Mean Accuracy on Word	Mean Accuracy on Strong Syll	Error type*	Mean Accuracy on Weak Syll1	Error type	Mean Accuracy on Weak Syll2	Error type	Mean Accuracy on Weak Syll3	Error type
BÄRbel NC	38.5%	0%	Vowel, /r/, coda, misc.	77.%	Vowel, misc.				
SÖNke NC	33.3%	5.5%	Vowel, coda, misc.	61.1%	Misc.				
SONja C	33.3%	22.2%	Onset, vowel, misc.	44.4%	Vowel weakening, misc.				
anDREas C	33.3%	0%	/r/, vowel	38.8%	Vowel	61.1%	[a] for schwa		
Eberhardt NC	27.7%	83.3%	Vowel	0%	Vowel, /r/	0%	Vowel, /r/		
KArin C	25%	50.%	Vowel, misc.	0%	/r/, vowel, misc.				
LUTZ NC	22.2%	22.2%	Vowel						
ARnim NC	22.2%	0%	Vowel, /r/	44.4%	Onset, vowel, misc.				
GISbert NC	19.4%	38.8%	Coda, misc.	0%	/r/, vowel, misc				
HArald C	17.3%	22.2%		12.5%	Vowel, Coda				
Annegret NC	35.3%	16.6%	Vowel	88.8%	Misc.	0%	/r/		

(*Continued*)

Name & Word Type	Mean Accuracy on Word	Mean Accuracy on Strong Syll	Error type*	Mean Accuracy on Weak Syll1	Error type	Mean Accuracy on Weak Syll2	Error type	Mean Accuracy on Weak Syll3	Error type
margaREte C	22.2%	0%	/r/, vowel	33.3%	Vowel	0%	/r/, vowel	55.5%	[a] for schwa
LAUra C	16.6%	33.3%	Vowel	0%	/r/, vowel weakening				
CORdula NC	12.9%	0%	Vowel, /r/, misc.	22.2%	Vowel, onset C, misc.	16.6%	Onset, vowel, misc.		
ALbert C	8.3%	16.6%	Vowel, coda, misc	0%	Vowel, /r/, misc				
HERmann C	8.3%	0%	Vowel, /r/	16.6%	Vowel weakening, coda				
HARTmut NC	5.5%	0%	Vowel, /r/	11.1%	Vowel, coda				
FRANK C	0%	0%	/r/, vowel,						
JÖRG NC	0%	0%	Vowel, /r/						

Stressed syllables are capitalized;
* foil pronunciations are included in the % but not in the error type

[laɪ], which is explicable if the input has activated the English word *Carolina* (as in "North and South Carolina"). Similarly, participant 04F021107 pronounced target [efa] as [iva], suggesting that *Eva* had activated the English name *Eve* [iːv].

The only word on the cognate list that appears not to have readily activated a functionally equivalent word (and its sound form) is the German *Georg*, a word whose spelling might be assimilatable to the English *George* but whose pronunciation all of our participants treated as a novel form. Thus, while linguistically a cognate word, *Georg*, target = [ˈge.joɐk], does not activate the sound form of *George*.

We also find suggestive evidence among the non-cognate words that the stimuli have activated English lexical items that might be influencing pronunciation. Although two participants pronounced *Anja* (target = [ˈan.ja]) with a low-vowel in the first syllable [an], most of the participants used a low fronted vowel [æn], which makes sense if *Anja* has activated the English name *Ann* [æn]. Similarly, most of the participants used the same low fronted vowel in *Annegret* (target = [ˈanə.gʁet]), which suggests activation of the English *Anna*. Subject 01F051107 pronounced *Arnim* as *Adam* and a couple of participants pronounced *Jörg* (target = [jöɐk]) as *York* ([jɔɹk]) – showing great variability in the realization of both the initial vowel, and rhotic. In short, there is considerable evidence of lexical effects from activated English words on the pronunciation of the target names.

At the same time, lexical activation of English words interacts with perceptual representations. Learners can combine L1 units with sounds that are not part of that L1 word to try to imitate the input. Most of the participants were accurate on German *Johanna* (target = [jo.ˈha.na]), which suggests that even if it activates the English equivalents, the learners could choose to attend to the input and try to produce target-like sounds. 04F021107 is, again, interesting because one of her productions involved combining the initial syllable [jo] with the low front vowel [hæ], again suggesting that the participant was computing syllabic strings in a creative way, based in part on the stimulus and in part on an activated English word. Similarly, while most of the participants pronounced *Frank* (target = [fʁaŋk]) with a completely English pronunciation [fɹæŋk], three participants used a low back vowel [ɑ] to attempt to approximate the stimulus. Most participants pronounced *Edmund* with a final voiced alveolar stop [d] but a couple of participants used a voiceless [t], in conformity with the stimulus. In short, participants were able to creatively combine sounds from the L1 repertoire even in cognate words suggesting that the L1 sound form of the cognates was open to influence from a perceptual representation.

Considering the data now below the word, we found L1 phonological effects. A couple of errors involved introducing a nasal velar into the coda of a Syllable1 position immediately before a [k]. Interestingly, the non-cognate name *Heike* [ˈhaɪkə] was systematically pronounced with a raised diphthong, evidence of

"Canadian-raising" (Chambers 1975, 1989; Moreton & Thomas 2007). Unstressed syllables in word-final position were often pronounced as schwa, in conformity with the English sound system. At the same time, participants were obviously sensitive to the repeated appearance of [a] in word final position because some of the errors in words like *Gesine*, *Sönke,* or *Claudia* and *Margarita* (both pronounced by our German native speaker with a final schwa) involved introducing [a] into the position where a schwa should have occurred.

Not surprisingly, the pronunciation of front rounded vowels caused problems and led to various solutions. The name *Jörg* (target = [jöɐk]) was replaced by various back vowels that preserve lip-rounding, including vowels that assimilate the word to *York*. Some participants produced a mid-centralized round vowel [ɞ]. *Sönke* (target = [ˈzœŋ.kə]) is replaced by various back vowels but also by the front vowels [ɪ] and [ɛ] which involve maintaining tongue position, but eliminate lip-rounding. The [ʊ] vowel of *Lutz* (target = [lʊts]) occurs in English (in words like *could*) but our participants preferred a centralized unrounded [ɞ] (perhaps influenced by trying to make this word sound "foreign" as it was a non-cognate form, but crucially, still maintaining the features [BACK], [RTR] (retracted tongue root) and, [ROUND]).[15]

The final issue to discuss is what phonological cues either hindered or helped the six participants in this multiple-case study in realizing the target forms. In order to figure this out we looked specifically at rhotics and vowels which we will now discuss.

Looking first at the production of rhotics, these data were examined in an attempt to answer three questions: (i) Was there a significant effect of cognate status? (ii) Was there a significant effect of syllable position? and, (iii) Did subjects attempt to shift their rhotic productions towards a non-English production? This latter question was important because, as we know, many individuals lack the ability to produce some rhotics without substantial phonetic training, such as alveolar or uvular trills. So even if subjects did not have the fine-motor skill necessary to create target-like uvular fricatives, for example, any shift away from an alveolar approximate was considered to be de-Anglicizing their production.

One note to make here is that incorrect responses were removed from the data set in all instances before statistics were run on that data to remove any potential confounds. The first measure considered was: Did subjects produce a rhotic in their responses when presented with one in the stimuli? An ANOVA was run

15. Figure-skating fans will know this word as a common noun, as in *triple lutz jump*. The pronunciation of that word is typically heard as [lʌts], clearly not the target our participants were aiming for.

on a GLM which determined that set and syllable position were both significant predictors of whether or not a subject produced a rhotic. Set was evaluated as [$G^2(403) = 31.433$, $p < 0.001$] with a main effect showing that rhotics were produced significantly more often in response to cognate forms. Syllable position was evaluated as [$G^2(402) = 56.262$, $p < 0.001$] with a main effect showing that rhotics were produced significantly more often if the stimuli had a rhotic in onset position. No significant interaction was found for these (or any other) main effects.

The second measure considered was: Did subjects shift their productions away from an English-like rhotic? Once more, an ANOVA on the GLM was run considering the various factors to look for significance which should be included in the model. In this instance, no significant main effects or interactions were found. Given this finding, the data were divided into two subsets: subjects who were actively attempting to mimic the German accent, and those who were not. The subjects ($n = 4$) who were not attempting to mimic the German accent (classified as "Very Anglicized") were then found to have a significant main effect of coda position on whether or not they shifted their productions to a non-English form [$G^2(223)$ 6.4730, $p = 0.039$] as well as showing a significant main effect of non-cognates as being a better predictor of shifting their production away from a standard English articulation [$G^2(225) = 9.5124$, $p = 0.002$]. Possibly because the other subjects were actively attempting to mimic the German productions, no significant effects were found to predict when they would or would not shift their productions away from [ɹ].

Once again looking at all six subjects, the final measure investigated was whether or not the German rhotic target was produced (not just displaying a production shifting away from [ɹ] in some way). As with the "very Anglicized" group, both coda position and non-cognate status were found to be significant indicators of whether or not subjects produced an on-target rhotic. For non-cognates, the result was [$G^2(330) = 5.6315$, $p = 0.018$] and for codas, the result was [$G^2(328) = 20.3028$, $p < 0.001$]. In none of the models was target rhotic type (manner or place) found to yield significant effects, and so they were not included in the models which ANOVAs were run on. Only main effects were found to be significant in this data set, no significant interactions were found.

In terms of the vowel data, the same considerations were taken with the data such that incorrect responses were not considered. Additionally, in tokens where either a syllable was epenthesized or deleted, every effort was taken to compare only the syllables that had correlates with the input. For example, subject 02F061107 once produced [ʌ.ˈhəɪ.kə] in response to the target [ˈhaɪ.kə]. In this instance, it is not clear that the initial [ʌ] represents a misparse of the input, or an extra-linguistic expression of trying to decide between the forms presented or to retrieve from memory the correct association prior to making a (correct) decision.

In instances such as these only the syllables [ˈhəɹ.kə] were considered – though it was noted that the stressed syllable represented a medial syllable in the response, not an initial one.

For this test, several significant main effects and interactions were found, and so there was no need to sub-divide the subjects into those who were or were not actively attempting to mimic the accent they were attending to. In attempting to ascertain if there were any important predictors of whether or not productions were target-like, several main effects were found to be significant. Unsurprisingly, if the target was an English phoneme, subjects were significantly more likely to produce a target-like response [$G^2(623) = 263.444$, $p < 0.001$] – this effect also features in most of the significant interactions as well. Onsets were also found to be a significant factor in predicting a target like production such that if a syllable had an onset, the vowel was significantly more likely to be target-like [$G^2(621) = 5.216$, $p = 0.024$]. If the syllable was stressed, it was also significantly more likely that the production would be target-like [$G^2(620) = 8.338$, $p = 0.004$]. Several significant interactions with the target being an English phoneme were also found: If the English phoneme had both an onset and a coda, it was significantly more likely to be target-like [$G^2(609) = 12.141$, $p < 0.001$], as well as if it had a coda and was stressed [$G^2(608) = 4.189$, $p = 0.041$]. If, however, the English phonemic vowel was followed by a coda and belonged to the non-cognate set, this interaction was found to be a significant predictor of a non-target-like production [$G^2(606) = 11.444$, $p < 0.001$]. Also, if there was an English phonemic vowel followed by a coda – but importantly having no onset and not being stressed – this was found to be a significant predictor of a non-target-like production [$G^2(618) = 12.594$, $p < 0.001$]. Finally, one significant interaction was found not involving having an English phonemic vowel in the target, and that is between non-cognate forms and stressed syllables in which it was found that this interaction was also a significant predictor of subjects producing a target-like vowel [$G^2(610) = 7.469$, $p = 0.006$].

3.5 Discussion

Interpreting these results, we suggest that the patterns observed in our data show that rhotics are more likely to be produced by participants if they appeared either in onset position or in cognate forms, subjects were significantly more likely to produce non-English rhotics if they were in non-cognates, or in coda position. It is well known that phonetic material which appears in the initial positions of a prosodic unit, or that is stressed, is more salient to listeners, so the first effect is unsurprising. Further, because L2 learners already have representations for cognate forms which include rhotics, it should also be no surprise that rhotics were produced more readily in those positions (though, in onset, an [ɹ] was more likely to be produced than a target-like rhotic). But, how can we explain why subjects who were not trying to mimic the German accent were more likely to shift their production in some way away from [ɹ] in non-cognates and in coda position, and why all listeners were more likely to be on-target in non-cognates or if the rhotic appeared in coda position?

With regards to why subjects performed better in producing target-like rhotics in non-cognates, one answer is readily suggested: Since subjects did not possess representations for the non-cognate forms, their ability to perform better on this set shows that they do not have competing lexical interference from pre-existing lexical representations. This further suggests that first-exposure learners quickly create representations in their L2 allowing them to map sound-meaning correspondences with minimal training.

But what is special about coda position and the ability to produce non-English sounds significantly more often? For this, there are two possible explanations. The first explanation is the robustness of phonetic cues. Wright (2004) argues that for obstruents, there are more salient cues provided in onset position because of the release burst and formant transitional cues going into a vowel. However, for rhotics, Hamann (2009) shows that the primary transitional cues for perception are based in the third formant, thus creating better transitional cues for post-vocalic rhotics. Thus, if listeners are getting more salient cues in this position, it could potentially facilitate more target-like productions. One thing this cannot answer, however, is why this is not only the case for participants who are attempting to mimic the German accent, but also for those who are not.

Another possible explanation for this effect of coda position facilitating significantly more target-like productions is one of under-specification in phonological content. In many English dialects, rhotics in coda position may be dropped, or present as rhoticized vowels rather than realizing a full alveolar approximate in coda position. This indicates that rhotics in this position are somehow less important in distinguishing a contrast from other competing representations. Thus, in coda position, only one of two things is really important to maintaining a lexical contrast. In r-less dialects, it is the moraic weight of the syllable that must be retained – so that when an [ɹ] is dropped, the preceding vowel becomes long. In dialects where the vowel is simply rhoticized, it seems that place and manner features in the syllable coda are lost, but the feature [RHOTIC] is maintained. If this is the case, one could hypothesize that if the only distinguishing feature of rhotics in coda position is, in fact, [RHOTIC], where a lexical representation existed, it would, by default, become an alveolar approximate in English. However, if place and manner features are not specified in this position, when listeners are presented with additional information to override the default articulation in a form that they do not possess a robust representation for, those vacant featural slots can be 'filled in' from the input – thus resulting in a target-like pronunciation. This is exactly what is observed in the data for this study.

Combining the two possibilities: We get stronger cues for place and manner features of rhotics in coda position, where in English rhotics seem under-specified. This leaves blank slots that positive evidence could influence or, fill in, to become more target-like in an L2.

These results may seem at odds with the conclusions of Colantoni & Steele (2007) who investigated the English acquisition of a French L2 /ʁ/, concluding that place and manner features were more accurately produced in onset position, and voicing features in coda. However, their test was conducted on intermediate and advanced learners of French who have had more time to flesh out their representations and acquire the non-native contrast and articulatory gestures (see the earlier section on *The study of "first exposure" learners* for relevant discussion). This is exactly the reason that in this study we chose to investigate learners on first exposure – so that we could understand what the initial representations of L2 learners looked like, and what they encoded. While these representations undoubtedly change with experience and reinforcement, we conclude here that the initial representation of L2 learners utilizes the features and prosodic categories of the L1. Where these representations are robust in the L1, learners will exhibit lexical interference. Where these representations are defective or non-existant in the L1, learners will be prone to acquiring other phonological features of place and manner, allowing more target-like productions without lexical interference. This is the reason that even those learners who were not actively attempting to mimic the German accent produced target-like articulations significantly more often for non-cognates, or in coda position.

With regards to the vowels, the patterns observed in our data suggest that more salient cues facilitate target-like productions. Belonging to the L1 phonemic inventory, having an onset, an English phoneme being framed between an onset and coda to offer transitional cues, English phonemes with coda consonants in non-cognate forms, or being stressed – a stressed English phoneme with a coda, or the stressed syllable in a non-cognate form, all help. (See Curtin 2009 on the saliency of stress and the importance of phonotactics, or Wright 2004 on transitional cues and perception.) More interesting is what is going on with the significant predictors of non-target-like productions: English phonemes which have a coda, or a functional equivalent in a non-cognate form. As we can see from the data, subjects were generally good at matching a target [a] or monophthong [e] or [o] (compared to the Canadian English [æ/a], [eɪ], and [oʊ]). However, when there is a voiceless coda consonant for a diphthong (i.e. /aɪ/), the production usually resulted in Canadian Raising to [əɪ] (Chambers 1975, 1989; Moreton & Thomas 2007), which would result in a non-target-like production. This was also evidenced in many of the productions of target /ˈhaɪ.kə/, indicating that those subjects parsed what they heard as [ˈhaɪk.ə] or [ˈhaɪk̩ə] with an ambi-syllabic [k] which produced the [əɪ] diphthong. Support for this analysis comes from the fact that subject 01F081107 twice produced the Canadian Raising variant, but not on re-test2 where she faithfully produced the [aɪ̠] target, but mis-pronounced the following consonant as [d] – which would not trigger the phonological effect. Other factors in the coda position which inhibited a subject from producing a target-like vowel include nasals (i.e. /ˈhaɪ.dʁun/ being realized as [ˈhaɪ.drʊn] or [ˈhaɪ.dɹən]) and rhotics (i.e. /ˈgiz.bɛʁt/ being realized as [ˈgis.bəɹt] or [ˈgis.bɚt]) – both of which affect formant structure, and could lead to vowel reduction.

4. Conclusions

In this paper, we have analysed production data from a small group of native speakers of English, exposed to German for the first time. We have provided some quantitative data showing that learners are most accurate on syllables and segments that conform to the L1 repertoire but that their pronunciation is heavily influenced by the apparent activation of L1 words that sound like (a part of) the target word. We expected this kind of influence in the case of the cognate words, but we found evidence of a "words-within-words" effect even in the non-cognate words (Cutler & Norris 1988; Cutler 2012). This suggests that lexical effects on L2 pronunciation among more advanced learners might be more widespread than the literature suggests. Acquiring an accurate pronunciation entails not only constructing an accurate representation of the input (presumably based on repeated exposure to targets) but also inhibiting L1 words that might provide alternative representations of syllables or syllable sequences. This is especially important in the case of cognate words (Dijkstra et al. 1999).

Our participants also showed effects of the L1 sound system in terms of some limited phonotactic and prosodic effects. Non-target segments (front-rounded vowels and rhotics) were problematic, although we observed considerable variation among the participants in their attempts to produce sounds that fit the input. Of specific note is the fact that even when productions did not match the target vowels, either because they did not possess the target vowel in the L1, or when – we hypothesize – they attempted to make non-cognate words "more foreign", phonological features (specifically for height, front/backness, rounding, and [RTR], or some subset thereof, were frequently maintained. This was observed in the substitution of [ɪ] or [ɛ] for /ø/ and /œ/ and [ɞ] and [ɒ] for /ʊ/ and /ɔ/. Most phonemic consonants in German are also available in English (there were no palatal/velar fricatives in our data set to be able to observe the repair strategies learners might employ, or how successful our participants would be in producing these targets faithfully). The exception to this being that some participants, as mentioned earlier, often produced [dʒ] for a word-initial /j/ in the target, especially when that word was a cognate with English such as *Josef* or *Johanna*. This did not extend to medial /j/ where the glide was produced faithfully, as in *Sonja*. Throughout we have evidence of various sorts that participants are in fact sensitive to distinctions present in the signal, even when those distinctions are not part of the L1 system.

Proponents of "enriched lexical representations" (Curtin 2002), namely the proposal that we store in long-term memory redundant acoustic information, rather than stripping off all information not needed by the phonological system, offer a mechanism for explaining how our participants could negotiate representations that are a compromise between the L1 pronunciation of a word and the novel representations that learners ended up producing. Such representations might serve as the basis for emerging categories as learners process more and more of the L2. How input actually drives that process is a topic for further research.

5. Acknowledgements

This research was made possible through generous funding from the Social Sciences & Humanities Research Council of Canada through both its Canada Research Chairs program (950- 202408) and its general grants program (410-2006-0323). We also wish to thank research assistants Lindsay Hracs, Danica MacDonald, Melissa Raymond, and Silke Weber who carried out essential tasks during various stages of this research. An anonymous reviewer and our Editor Hagen Peukert ensured that the final version of this chapter is in better shape than the initial draft; we gratefully acknowledge their suggestions and comments. Authors are listed alphabetically and contributed equally to this work.

References

Best, C.T. & Strange, W. 1992. Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–30.

Best, C.T., Traill, A., Carter, A., Harrison, K.D. & Faber, A. 2003. !Xóõ click perception in English, Isizulu and Sesotho listeners. In *Proceedings of the 15th International Congress of Phonetic Sciences*, M.J. Solé, D. Recasens & J. Romero (eds), 853–56. Barcelona: Causal Productions.

Best, C.T., & Tyler, M.D. 2007. Non-native and second-language speech perception. Commonalities and complementarities. In *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production* [Language Learning & Language Teaching 17], M.J. Munro & O.-S. Bohn (eds), 13–34. Amsterdam: John Benjamins. DOI: 10.1075/lllt.17.07bes

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R.A. & Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101(4): 2299–2310. DOI: 10.1121/1.418276

Brent, M.R., & Cartwright, T.A. 1996a. Distributional regularity and lexical access. *Journal of Psycholinguistic Research* 26: 363–375. DOI: 10.1023/A:1025032825951

Brent, M.R. & Cartwright, T.A. 1996b. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61: 93–125. DOI: 10.1016/S0010-0277(96)00719-6

Carroll, S.E. 2012. Segmentation on first exposure to an L2: Evidence for knowledge-driven, top-down processing. In *Multilingual Individuals and Multilingual Societies* [Hamburg Studies on Multilingualism 13], K. Braunmüller, C. Gabriel & B. Hänel-Faulhaber (eds), 23–45. Amsterdam: John Benjamins. DOI: 10.1075/hsm.13.04car

Carroll, S.E. 2014. Processing 'words' in early-stage SLA: A comparison of first exposure and low proficiency learners. In *Input Processing in Second Language Acquisition*, Z.H. Han & R. Rast (eds). Cambridge: CUP.

Chambers, J.K. 1975. Canadian raising. In *Canadian English. Origins and Structures*, J.K. Chambers, 83–100. Toronto: Methuen.

Chambers, J.K. 1989. Canadian raising: Blocking, fronting, etc. *American Speech* 64(1): 75–88. DOI: 10.2307/455114

Chambers, K.E., Onishi, K.H. & Fisher, C. 2003. Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87: B69–B77. DOI: 10.1016/s0010-0277(02)00233-0

Cholin, J., Schiller, N.O. & Levelt, W.J.M. 2004. The preparation of syllables in speech production. *Journal of Memory & Language* 50(1): 47–61. DOI: 10.1016/j.jml.2003.08.003

Colantoni, L., & Steele, J. 2007. Acquiring /ʁ/ in context. *Studies in Second Language Acquisition* 29: 381–406. DOI: 10.1017/S0272263107070258

Costa, A., Caramazza, A. & Sebastian-Galles, N. 2000. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(5): 1283–96. DOI: 10.1037/0278-7393.26.5.1283

Curtin, S. 2002. Representational Richness in Phonological Development. Ph.D. dissertation, University of Southern California.

Curtin, S. 2009. Twelve-month-olds learn word-object associations differing only in stress patterns. *Journal of Child Language* 36: 1157–1165. DOI: 10.1017/S0305000909009428

Curtin, S., Mintz, T.H. & Christiansen, M.H. 2005. Stress changes the representational landscape: Evidence from word segmentation. *Cognition* 96: 233–62. DOI: 10.1016/j.cognition.2004.08.005

Cutler, A. 2012. *Native listening. Language Experience and the Recognition of Spoken Words*. Cambridge MA: The MIT Press.

Cutler, A. & Norris, D.G. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113–21. DOI: 10.1037/0096-1523.14.1.113

de Jong, K.J., Silbert, N.H. & Park, H. 2009. Generalisation across segments in second language consonant identification. *Language Learning* 59(1): 1–31. DOI: 10.1111/j.1467-9922.2009.00499.x

Dijkstra, A., Grainger, J. & van Heuven, W.J.B. 1999. Recognizing cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory & Language* 41: 496–518. DOI: 10.1006/jmla.1999.2654

Dupoux, E., Pallier, C., Sebastián-Gallés, N. & Mehler, J. 1997. A destressing 'deafness' in French? *Journal of Memory & Language* 36: 406–21. DOI: 10.1006/jmla.1996.2500

Eisenberg, P. 2004. *Grundriß der Deutschen Grammatik. Das Wort*, 2nd edn. Stuttgart: J.B. Metzler.

Escudero, P. 2005. Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorisation. Ph.D. dissertation, University of Utrecht.

Escudero, P. 2006. Second language phonology: The role of perception. In *Phonology in Context*, M.C. Pennington (ed.), 109–34. Houndmills: Palgrave MacMillan.

Finn, A. & Hudson Kam, C.L. 2008. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition* 108: 477–99. DOI: 10.1016/j.cognition.2008.04.002

Flege, J.E. 2003. Assessing constraints on second-language segmental production and perception. In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, N.O. Schiller & A.S. Meyer (eds), 319–55. Berlin: Mouton de Gruyter.

Flege, J.E., Frieda, W.M. & Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–86. DOI: 10.1006/jpho.1996.0040

Flege, J.E., Munro, M.J. & MacKay, I.R.A. 1995. Effects of age of second language learning on the production of English consonants. *Speech Communication* 16: 1–26. DOI: 10.1016/0167-6393(94)00044-B

Flege, J.E. & Wang, C. 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics* 17: 299–315.

Flege, J.E., Yeni-Komshian, G.H. & Liu, S. 1999. Age constraints on second language acquisition. *Journal of Memory & Language* 41: 78–104. DOI: 10.1006/jmla.1999.2638

Friel, B.M. & Kennison, S.M. 2001. Identifying German-English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language and Cognition* 4: 249–74. DOI: 10.1017/S1366728901000438

Goldrick, M. & Larson, M. 2008. Phonotactic probability influences speech production. *Cognition* 107: 1155–64. DOI: 10.1016/j.cognition.2007.11.009

Guion, S., Flege, J.E. & Loftin, J.D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42. DOI: 10.1006/jpho.2000.0104

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(Suppl. 2): 5–24. DOI: 10.1111/j.1467-9922.2010.00598.x

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 in Language Teaching* (*IRAL*) 50: 239–76.

Hall, T.A. 1993. The phonology of German /R/. *Phonology* 10(1): 83–105. DOI: 10.1017/S0952675700001743

Hamann, S. 2009. The learner of a perception grammar as a source of sound change. In *Phonology in Perception,* P. Boersma & S. Hamann (eds), 111–49. Berlin: Mouton de Gruyter.

Hohne, E.A. & Jusczyk, P.W. 1994. Two-month-old-infants' sensitivity to allophonic differences. *Perception & Psychophysics* 56: 613–23. DOI: 10.3758/BF03208355

Jusczyk, P.W. 1997. *The Discovery of Spoken Language*. Cambridge MA: The MIT Press.

Jusczyk, P.W., Cutler, A. & Redanz, N. 1993. Infants' preference for the predominant stress patterns in English. *Child Development* 64: 675–87. DOI: 10.2307/1131210

Jusczyk, P.W., Luce, P.A. & Charles-Luce, J. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language* 35(5): 630–645. DOI: 10.1006/jmla.1994.1030

Kohler, K. 2010. German. In *Handbook of the International Phonetic Association: A Guide to the use of the International Phonetic Alphabet*, I. Maddieson & M. Barry (eds), 86–89. Cambridge: CUP.

Levelt, W.J.M. 1999. Producing spoken language: A blueprint of the speaker. In *The Neurocognition of Language*, C.M. Brown & P. Hagoort (eds), 83–122. Oxford: OUP.

Li, P. 1996. Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory & Language* 35: 757–74. DOI: 10.1006/jmla.1996.0039

Lisker, L. & Abramson, A.S. 1967. Some effects of context on voice onset time in English stops. *Language & Speech* 10: 1–28.

Luce, P.A., Pisoni, D.B. & Goldinger, S.D. 1990. Similarity neighbourhoods of spoken words. In *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives,* G.T.M. Altmann (ed.), 122–47. Cambridge MA: The MIT Press.

Mattys, S.L. & Jusczyk, P.W. 2001a. Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception & Performance* 27(3): 644–55. DOI: 10.1037/0096-1523.27.3.644

Mattys, S.L. & Jusczyk, P.W. 2001b. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78: 91–121. DOI: 10.1016/S0010-0277(00)00109-8

Mattys, S.L., Jusczyk, P.W., Luce, P.A., & Morgan, J.L. 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* 38(4): 465–94. DOI: 10.1006/cogp.1999.0721

Mattys, S.L., White, L. & Melhorn, J.F. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134(4): 477–500. DOI: 10.1037/0096-3445.134.4.477

McQueen, J.M. 1998. Segmentation of continuous speech using phonotactics. *Journal of Memory & Language* 39: 21–46. DOI: 10.1006/jmla.1998.2568

Moreton, E. & Thomas, E.R. 2007. Origins of Canadian raising in voiceless-coda effects: A case study in phonologization. In *Laboratory Phonology* IX, J. Cole & J.I. Hualde (eds), 37–64. Berlin: Mouton de Gruyter.

Peperkamp, S. & Dupoux, E. 2002. A typological study of stress 'deafness'. In *Laboratory Phonology* VII, C. Gussenhoven & N. Warner (eds), 203–240. Berlin: Mouton de Gruyter.

Piske, T. 2007. Implications of James E. Flege's research for the foreign language classroom. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* [Language Learning & Language Teaching 17], M.J. Munro & O.-S. Bohn (eds), 301–314. Amsterdam: John Benjamins. DOI: 10.1075/lllt.17.26pis

Piske, T., MacKay, I., & Flege, J.E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 181–215. DOI: 10.1006/jpho.2001.0134

Piske, T., Flege, J.E., MacKay, I.R.A., & Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59: 49–71. DOI: 10.1159/000056205

Pisoni, D.R., & Lively, S.E. 1995. Variability and invariance in speech perception. A new look at some old problems in perceptual learning. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, W. Strange (ed.), 433–59. Timonium MD: York Press.

Polka, L. 1992. Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics* 52: 37–52. DOI: 10.3758/BF03206758

Rast, R. 2008. *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.

Rast, R. 2010. The role of linguistic input in the first hours of adult language learning. *Language Learning* 60 (Suppl.2): 64–84. DOI: 10.1111/j.1467-9922.2010.00601.x

Rast, R. & Dommergues, J.-Y. 2003. Towards a characterisation of saliency on first exposure to a second language. *EUROSLA Yearbook* 3: 131–56. DOI: 10.1075/eurosla.3.09ras

Shoemaker, E., & Rast, R. 2013. Extracting words from the speech stream at first exposure. *Second Language Research* 29(2): 165–83. DOI: 10.1177/0267658313479360

Tench, P. 2003. Non-native speakers' misperception of English vowels and consonants: Evidence from Korean adults in UK. *International Rreview of Applied Linguistics in Teaching* (*IRAL*) 41: 145–73.

Tremblay, A. 2008. Is L2 lexical access prosodically constrained? On the processing of word stress by French Canadian L2 learners of English. *Applied Psycholinguistics* 29: 553–84. DOI: 10.1017/S0142716408080247

Vitevitch, M.S. & Luce, P.A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language* 40: 374–408. DOI: 10.1006/jmla.1998.2618

Weber, A. & Cutler, A. 2006. First-language phonotactics in second language listening. *Journal of the Acoustical Society of America* 119: 597–607. DOI: 10.1121/1.2141003

Werker, J.F. & Tees, R.C. 1984. Cross-language speech perception: Evidence for perceptual re-organisation during the first year of life. *Infant Behavior & Development* 7: 49–63. DOI: 10.1016/S0163-6383(84)80022-3

Wright, R. 2004. A review of perceptual cues and cue robustness. In *Phonetically based Phonology*, B. Hayes, R. Kirchner & D. Steriade (eds), 34–57. Cambridge: CUP. DOI: 10.1017/CBO9780511486401.002

Appendix 1: Stimuli

Cognate Sentences Training Trials

1. Hier ist Agnes.
2. Das ist Anita.
3. Hier sehen Sie Beatrice.
4. Da steht Claudia.
5. Hier ist Eva.
6. Das ist Johanna.
7. Hier sehen Sie Karin.
8. Da steht Laura.
9. Hier ist Margarete.
10. Das ist Sonja.
11. Hier sehen Sie Albert.
12. Da steht Andreas.
13. Hier ist Bruno.
14. Das ist Edmund.
15. Hier sehen Sie Frank.
16. Da steht Georg.
17. Hier ist Harald.

18. Das ist Hermann.
19. Hier sehen Sie Otto.
20. Da steht Josef.

Cognate Questions Training Trials/Re-tests

1. Ist hier Agnes oder Angela?
2. Ist das Anna oder Anita?
3. Hier sehen Sie Beatrice oder Berta?
4. Steht da Charlotte oder Claudia?
5. Ist hier Eva oder Edwina?
6. Ist das Julia oder Johanna?
7. Sehen Sie hier Karin oder Kirsten?
8. Steht da Lauren oder Laura?
9. Ist hier Margarete oder Martina?
10. Ist das Sandra oder Sonja?
11. Sehen Sie hier Albert oder Alexander?
12. Steht da Adolf oder Andreas?
13. Ist hier Bruno oder Bertram?
14. Ist das Erik oder Edmund?
15. Sehen Sie hier Frank oder Franz?
16. Steht da Gregor oder Georg?
17. Ist hier Harald oder Harry?
18. Ist das Hubert oder Hermann?
19. Steht da Josef oder Johannes?
20. Sehen Sie hier Oskar oder Otto?

Non-cognate Sentences Training Trials

1. Hier ist Anja.
2. Das ist Annegret.
3. Hier sehen Sie Bärbel.
4. Da steht Cordula.
5. Hier ist Femke.
6. Das ist Gesine.
7. Hier sehen Sie Heike.
8. Da steht Heidrun.
9. Hier ist Regina.
10. Das ist Senta.
11. Hier sehen Sie Arnim.
12. Da steht Benno.
13. Hier ist Dietmar.
14. Das ist Eberhardt.
15. Hier sehen Sie Gisbert.
16. Da steht Hartmut.
17. Hier ist Jörg.

18. Das ist Kai.
19. Hier sehen Sie Lutz.
20. Da steht Sönke.

Non-cognate questions Training Trials/Retests

1. Ist hier Anje oder Anka?
2. Ist das Annika oder Annegret?
3. Hier sehen Sie Bärbel oder Bella?
4. Steht da Cornelia oder Cordula?
5. Ist hier Franke oder Femke?
6. Ist das Gesine oder Gisela?
7. Sehen Sie hier Heide oder Heike?
8. Steht da Heidrun oder Helga?
9. Ist hier Rike oder Regina?
10. Ist das Senta oder Silka?
11. Sehen Sie hier Arno oder Arnim?
12. Steht da Bernd oder Benno?
13. Ist hier Dietmar oder Detlef?
14. Ist das Ekkehard oder Eberhardt?
15. Sehen Sie hier Gisbert oder Günther?
16. Steht da Helmut oder Hartmut?
17. Ist hier Jörg oder Joachim?
18. Ist das Kurt oder Kai?
19. Sehen Sie hier Lutz oder Ludo?
20. Steht da Sigmund oder Sönke?

Example of stimulus picture