

The Role of Linguistic Knowledge in the Encoding of Words and Voices in Memory

Stephen Winters¹, Karen Lichtman², and Silke Weber¹

¹University of Calgary and ²Northern Illinois University

1. Introduction

There is a longstanding tradition in phonetics to conceive of speech as containing both “indexical” and “linguistic” components (Abercrombie, 1967). The linguistic properties of speech support the identification of linguistic (phonemic, lexical, etc.) contrasts, while the indexical properties of speech “reveal personal characteristics of the speaker”, such as their dialect or group membership, their gender, their age, their emotional or mental state, or particular vocal idiosyncrasies unique to that speaker. In Abercrombie's conception, speakers provide a personal "medium" for linguistic messages; this means that certain indexical properties in this medium may be "extra-linguistic".

Exemplar theories of speech perception (Johnson, 2007) have emphasized the cognitive utility of not breaking down the signal into independent linguistic and indexical streams. Instead, exemplar theories conjecture that listeners store in memory unanalyzed representations of speech that are “rich” with informative detail. Linguistic representations might, for instance, include indexical or talker-specific information, just as indexical representations might include linguistic information. Generalizations are then calculated on the fly, from summed activations of similar exemplars. In its strong form, exemplar theory claims that listeners store in memory every detail of every speech event they experience in their lifetimes.

The core evidence for the exemplar view comes from known interactions between indexical and linguistic properties in speech perception. Linguistic information affects the identification of the indexical properties of speech in that listeners can more easily identify someone speaking a language that they know (Goggin et al., 1991; Perrachione et al., 2009). The reverse of this equation is also true--indexical information affects the processing of the linguistic content of speech. Palmeri et al. (1993), for instance, demonstrated that voice information influences the recognition of repeated words in a "continuous word recognition" experiment. Listeners in this experiment heard a series of words, spoken in different voices, and identified whether or not each word was “old” or “new” in the list. It was easier for listeners to recognize repeated words if they were spoken in the same voice as on the first presentation, suggesting that listeners were storing word and voice information together in memory. Nygaard et al. (1994) also showed that there was a "familiar talker advantage" in word recognition. In this study, listeners learned to identify talkers over a series of nine days. On the tenth day of the study, the listeners transcribed words spoken by both trained (familiar) talkers and untrained (unfamiliar) talkers. The listeners identified a significantly higher percentage of words produced by the familiar talkers, indicating that knowledge of the familiar talkers' voices in memory contributed to the processing of novel stimuli.

However, such evidence of indexical influences on linguistic processing primarily comes from English-only language studies. Do listeners encode the same amount of indexical detail in memory from exemplars of second language (L2) speech? There is some reason to suspect that they might not, since listeners often exhibit perceptual insensitivity to fine-grained phonetic details of L2 speech. For

* The authors would like to thank Isabelle Darcy, Lourdes Ortega, and the other audience members at the SLRF 2011 conference for their comments on a presentation of this work, two anonymous reviewers for their comments and suggestions on an earlier version of this paper, and Christine Shea and Stephanie Archer for their assistance in refining the methodology used in the word recognition experiment.

example, English listeners perceptually collapse non-native VOT distinctions in Thai (Abramson & Lisker, 1970) and Japanese learners of English famously have difficulty acquiring the English /l/ - /ɾ/ distinction (Strange & Dittman, 1984).

There is also evidence to suggest that language may not affect the processing of indexical information when listeners hear voices speaking in a language they do not understand. In Winters et al. (2008), monolingual English listeners were trained to identify bilingual talkers from words produced in one of two languages: English or German. After learning to identify these talkers in one language over a series of four days, the listeners were then tested on their ability to identify the same talkers, speaking in the other language. Listeners who were trained to identify the talkers from English stimuli displayed a drop-off in identification accuracy when they heard the same talkers speaking in German. Listeners who were trained to identify the talkers speaking in German, however, exhibited no such drop-off in accuracy when they heard those talkers speaking in English. This pattern of results suggested that the German-trained listeners had developed language-independent representations of the talkers' voices, while the English-trained listeners had encoded language-specific information in their cognitive representations of the talkers' voices.

In Levi et al. (2011), monolingual English listeners were similarly trained to identify bilingual talkers speaking either English or German. After training, both groups of listeners were tested on their ability to identify English words spoken by both those (familiar) talkers and a group of unfamiliar talkers in noise. Results showed that only the English-trained listeners displayed a word recognition benefit in this task. This replicated the "familiar talker advantage" found by Nygaard et al. (1994), but also revealed that this advantage does not transfer across languages.

The results of these two studies displayed two consistent patterns. English-trained listeners displayed a perceptual interaction between linguistic and talker information in both voice learning experiments. The German-trained listeners, however, did not display a perceptual interaction between linguistic and talker information in either experiment. The English-trained listeners thus seemed to develop richly detailed, exemplar-like representations of voices--in which linguistic and indexical information intersect--while German-trained listeners developed sparser, language-independent representations of voices. Training listeners to perform an indexical task in an L2 does not, therefore, seem to yield exemplar-style representations that integrate indexical and linguistic information.

This study asks whether the same is true of listeners performing a linguistic task in a second language. Will these listeners also separate linguistic and indexical information when learning to identify linguistic items from an unknown language? There is some evidence from previous work that listeners can and do process the linguistic and indexical information in L2 words in an integrated fashion. Bradlow and Pisoni (1999) found, for instance, that listeners have more difficulty transcribing a list of L2 words produced by multiple talkers than by a single talker. Trofimovich (2005) also found an advantage for same-voice repetitions over different-voice repetitions in an auditory word naming task.

This experiment looked at the effects of indexical information on L2 word recognition by engaging three different groups of listeners in a continuous word recognition task: native German speakers, monolingual English speakers, and native English speakers who were learning German as a second language. Each of these groups heard two series of German words produced by a variety of native German speakers. Each word in a list was repeated once in that list; the participants' task was to listen to each word and indicate whether it was a "new" item in the list or a "repeated" item. Half of the repetitions were produced in the same voice as on the initial presentation, and the other half of the repetitions were produced in a different voice. Based on the results of Palmeri et al. (1993), it was expected that native German listeners would correctly recognize repeated words produced in the same voice as "repeated" items more often than repeated words produced in a different voice. What was unknown was whether the non-native speakers of German would exhibit the same perceptual pattern. If they were to store the L2 stimuli in an integrated, exemplar-style fashion in memory, then they should more readily recognize a word repeated in the same voice as on its initial presentation, since it would exactly match an exemplar in memory. If they were to store the linguistic and indexical content of each word separately in memory, however, they should exhibit no recognition benefit for words produced in the same voice as on the initial presentation.

2. Methods

2.1. Materials

Stimuli for the experiment were produced by 10 German-L1/English-L2 bilinguals. Five of these speakers were male, and five were female; all of them hailed from roughly the same dialect region of central Germany. These talkers produced 360 monosyllabic German CVC words, which ranged from high to low frequency. These words effectively formed all of the monosyllabic CVC words in the CELEX German database (Baayen et al., 1995); the fact that they ranged in frequency was thus a natural outgrowth of the comprehensive nature of the original stimulus list. All stimuli were recorded in a quiet, sound-attenuated booth. Individual words were spliced from the original recordings into separate sound files, each of which was normalized to 65 dB intensity. These stimuli were originally produced for a different study (Levi et al., 2007); further details on their production may be found there.

2.2. Procedure

Listeners heard two series of 160 single-word trials. Within each series, there were 80 "new" tokens (words which had not been presented before) and 80 "old" tokens (repetitions of words which had been presented before). Of the 80 "old" tokens in a series, 40 were produced by the same voice as in the initial presentation (old-same), and 40 were produced by a new voice (old-different). After hearing each word, listeners indicated whether they thought the word was "new" or "old" in the list by clicking on one of two on-screen buttons. Both response time and the listener's response were saved into an output data file. After registering a response, listeners then clicked on another button to play the next word in the series; the entire experiment was thus self-paced.

Note that, since the experiment was run on different machines in different testing venues, the consistency of response time measurements could not be guaranteed, and thus will not be presented in depth in the analysis section. A reviewer has suggested, however, that enabling participants to proceed at their own pace could have allowed some participants to "rehearse" items in memory more than others, thereby providing them with a performance advantage on the task. We therefore used the timing data to investigate a possible relationship between overall time spent on the task and total correct identification percentage for each listener. Total time spent on the task ranged from 773 seconds to 1583 seconds, and the percentage of correct responses given by each participant ranged from 64% to 90%, but there was no clear relationship between the two. The two scores actually yielded a correlation that suggested a relationship in the opposite direction of what the reviewer hypothesized might happen ($r = -.111$), indicating that slower participants had more difficulty with the task, but this relationship was not significant ($p = .433$).

The tokens in each series were all produced by five different talkers, all of the same gender (male or female). The two series that each listener heard differed in both the gender of the voices producing the words, and the words presented in the series. That is, no word presented in the series produced by one gender was also presented in the series produced by the other gender. Each talker produced 32 of the tokens in the series--16 old and 16 new. Each talker thus produced 8 repeated items that were initially presented in their voice, and 8 repeated items that had initially been presented in a different voice (2 for each of the other 4 voices). Once a listener had finished one series, they took a short break and then worked through the second series, produced by the other gender talkers. The time in between sessions was, on average, a little over a minute (69 seconds). This was effectively as much time as it took for the experimenter to take a seat at the computer and set up the second half of the experiment. The order of presentation of genders was counterbalanced across listeners.

A different selection of 320 German words (from the larger set of 360) was chosen randomly for each listener. The stimuli were played to listeners in a quiet room, at a comfortable listening level, over headphones, via a customized Supercard stack running on a Macintosh laptop. The entire experiment lasted about 30 minutes.

2.3. Participants

Three sets of listeners participated in this experiment. The first group (Eng-Mono) consisted of 17 monolingual English listeners, with no knowledge of German. This group was recruited from an introductory psycholinguistics class at the University of Illinois at Urbana-Champaign (UIUC) and received partial course credit for their participation in the experiment. These participants had a mean age of 20.4 years. Most were from Illinois and had parents who were also from Illinois, and all grew up as monolingual English speakers. On average, they had studied 1.4 foreign languages (mainly Spanish) starting at age 14.4, but none of them reported any prior exposure to the German language.

The second group (Eng-Bi) consisted of 19 English-L1/German-L2 listeners, who were recruited from a fourth-semester German class at UIUC. These listeners also received partial course credit for participating in the study. These listeners had a mean age of 21.0 years old. All of them had learned only English as their native language from birth; the majority of them grew up in Illinois and had parents who were also from Illinois. On average, they started learning German at age 16.3, and rated themselves 2.8 at speaking German, 3.1 at writing German, and 3.3 at reading German, using a scale where 1 is poor and 5 is excellent. On average, they spoke 1.2 foreign languages besides German--mostly Spanish, with a mean starting age of 15.8. None had ever lived in Germany.

The third group (German) consisted of 16 German-L1 listeners. These listeners were tested at Universität Paderborn in Germany, where they were recruited from the general student populace and were paid 7 Euros each for their participation in the study. For this group of listeners, the instructions were translated into German, and the entire study was run and presented to them in the German language.

3. Analysis + Results

3.1. Responses to all stimuli

An analysis of variance (ANOVA) was then conducted to determine the possible effects of Lexical Frequency and Listener Group (English-Mono, Eng-Bi, German) on the percentage of correct responses given by each listener. In order to investigate the effects of lexical frequency on listener responses, all of the words in the database were divided into three frequency bins--high, mid and low--with 120 words in each bin. One reviewer also expressed interest in cognates in the word list, and whether listeners might have processed or recognized these words in a different fashion than non-cognate words. Strictly speaking, there were a few cognate words presented to listeners in the experiment--German words that were similar in both phonological form and meaning to English words, like "mein" or "Netz"--but they were so few in number that it is not possible to conduct a meaningful statistical analysis of how listeners responded to them in contrast to other, non-cognate words. Casting the net a bit wider, however, we did break the production list down into words that did not sound like English words (e.g., "Loch" or "null") and words that did sound like English words (e.g., "Pein", "Bett"), even though they did not necessarily mean the same thing (in German) as the word they sounded like in English. The "English-likeness" of the German words thus became the third independent factor in this Analysis of Variance.

This ANOVA yielded significant main effects of Listener Group ($F(2,16628) = 94.39; p < .001; \eta^2 = .011$) and English-likeness ($F(1,16628) = 12.85; p < .001; \eta^2 = .001$). There were also two significant interactions: Listener Group * Frequency Bin ($F(2,16628) = 7.47; p < .001; \eta^2 = .001$) and Listener Group * English-likeness ($F(2,16628) = 4.47; p = .011; \eta^2 = .001$). Post-hoc analysis of the main effect of Listener Group indicated that German listeners correctly classified more stimuli (84.8%) than both the monolingual English listeners (74.2%; $t(10558) = 13.5; p < .001$) and the bilingual English listeners (77.1%; $t(11198) = 10.28; p < .001$). The English bilingual listeners also correctly classified more stimuli than the monolingual listeners ($t(11518) = 23.64; p < .001$). With respect to the effect of English-likeness, English-like words were identified correctly more often (81.2%) than non-English-like words (78.1%) ($t(16638) = 3.57; p < .001$).

Post-hoc analysis of the significant Listener Group*English-likeness interaction indicated that this main effect was due entirely to the performance of the two groups of native English speakers. The monolingual English listeners identified English-like words correctly 77.9% of the time, compared to

73.5% of non-English-like words ($t(5438) = 2.71$; $p = .007$), while the bilingual English listeners identified English-like words correctly 81.7% of the time, compared to 76.3% for the non-English-like words ($t(6078) = 3.56$; $p < .001$). For the German listeners, however, there was no significant difference between the two groups of words (84.1% and 84.9%, respectively).

Post-hoc analysis of the Listener Group*Frequency Bin interaction revealed that the effect of Frequency on the percentage of correct responses exhibited three different patterns, depending on the Listener Group providing the responses. For the English monolingual listeners, Frequency had no effect on response accuracy. For the English bilingual listeners, however, there was a positive relationship between Frequency and response accuracy; these listeners correctly identified high frequency words 78.5% of the time, but low frequency words only 75.1% of the time [$t(4072) = 2.59$; $p = .01$]. Medium frequency words (77.9%) were also recognized more accurately than low frequency words [$t(4076) = 2.08$; $p = .037$], but there was no significant difference in response accuracy for high and medium frequency words. For the German listeners, on the other hand, both medium (86.9%) and low frequency (85.3%) words were identified significantly more accurately ($t(3422) = 3.79$; $p < .001$ and $t(3396) = 2.47$; $p = .014$, respectively) than the high frequency words (82.2%). A plot of this interaction is shown in Figure 1.

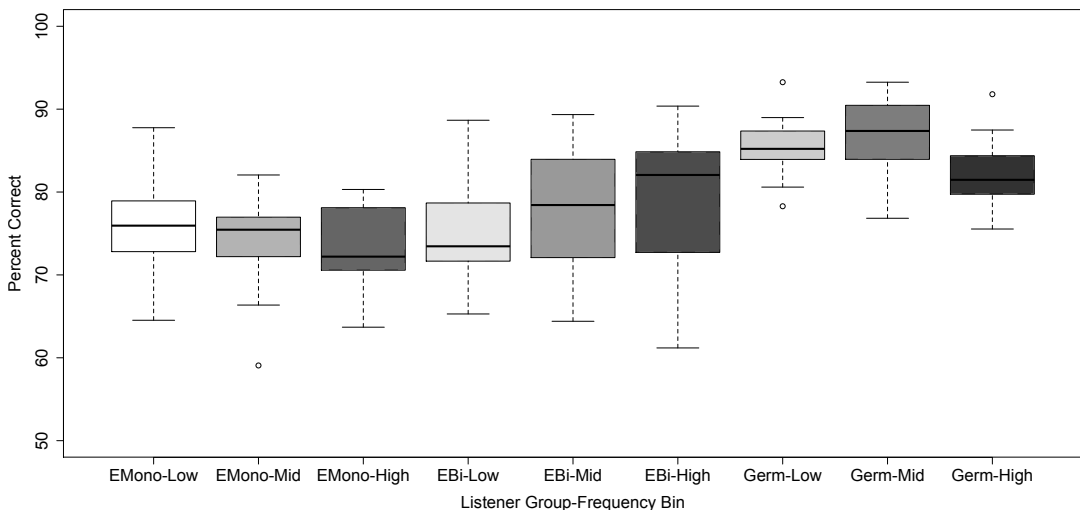


Figure 1. Percentage of correct identifications of all stimuli, by listener group and frequency bin

3.2. Responses to repeated stimuli

In addition, a four-way ANOVA was also run on the response data from only the repeated stimuli presented in the experiment. The independent factors in this ANOVA included Listener Group, Frequency Bin, English-likeness and Repeat Type (same voice, different voice). Significant main effects included English-likeness ($F(1,8296) = 5.96$; $p = .014$; $\eta^2 = .001$) and Repeat Type ($F(1,8296) = 63.31$; $p < .001$; $\eta^2 = .007$). (Note, interestingly, that there were no significant main effects of either Listener Group or Frequency.) There were also significant interactions between Listener Group * Frequency Bin ($F(2,8296) = 16.38$; $p < .001$; $\eta^2 = .004$), Listener Group * English-likeness ($F(2,8296) = 3.50$; $p = .03$; $\eta^2 = .001$), and Frequency Bin * Repeat Type ($F(1,8296) = 5.23$; $p = .022$; $\eta^2 = .001$).

Since the patterns of the significant Listener Group, Frequency Bin and English-likeness factors and interactions closely resemble those found in the previous ANOVA, they will not be discussed in further detail here. The Repeat Type effects and interactions are new, however, and therefore deserve further elaboration. The main effect of Repeat Type was clear: repeated words presented in the same voice (80.8%) were correctly classified more often than repeated words presented in a different voice (73.5%; $t(8318) = 7.94$; $p < .001$). Post-hoc analysis of the Frequency*Repeat Type interaction revealed that this general effect of voice held for all three frequency bins. There was a twist, however, in that correct recognition of the words repeated in a different voice depended on frequency:

identification accuracy was higher for repeated medium (74.8%) and high-frequency words (74.5%) than for low frequency words (71.1%; [t(2804) = 2.01; p = .044] and [t(2734) = 2.17; p = .03], respectively). There were no corresponding lexical frequency effects on words repeated in the same voice--post-hoc t-tests revealed no significant differences between the recognition of high (80.1%), medium (80.6%) and low (81.6%) frequency words.

3.3. Phonetic Breakdown

The number of correct and incorrect responses was also broken down according to the phonetic characteristics of the stimulus words. This analysis focused primarily on phonetic features of the CVC words which are found in German but not English:

- words containing the velar fricative (/x/), such as *sich* or *Loch*
- words containing the affricates /pf/ or /ts/, such as *Pfiff* or *Zahn*
- words containing an initial /r/ (phonetically [ʀ]), such as *Rad* or *Rock*
- words containing a final /r/, (phonetically [ʀ] or [ʁ]), such as *Bier* or *leer*
- words containing a final "clear" /l/, such as *Ball* or *toll*
- words containing a front, rounded vowel, such as *kühl* or *schön*
- words containing a long mid-vowel, such as *Boot* or *Weg*

Since the particular stimuli presented to each listener were randomly selected, they were not balanced across these various phonetic categories. This meant it was not feasible to conduct a rigorous statistical analysis of the effects of phonetics on correct and incorrect responses; hence, the breakdown reported here will focus on prominent descriptive patterns in the results.

The first of these patterns is found in responses to words containing velar fricatives. The monolingual English listeners found these stimuli particularly difficult to recognize in repeated stimuli, with no apparent difference in recognition accuracy with respect to repetitions in the same voice or a new voice. The bilingual English listeners performed more or less as well with these repeated stimuli as did the native German listeners, however. A breakdown of the correct response percentages for these listener groups and the two repeated stimuli types can be found in Figure 2.

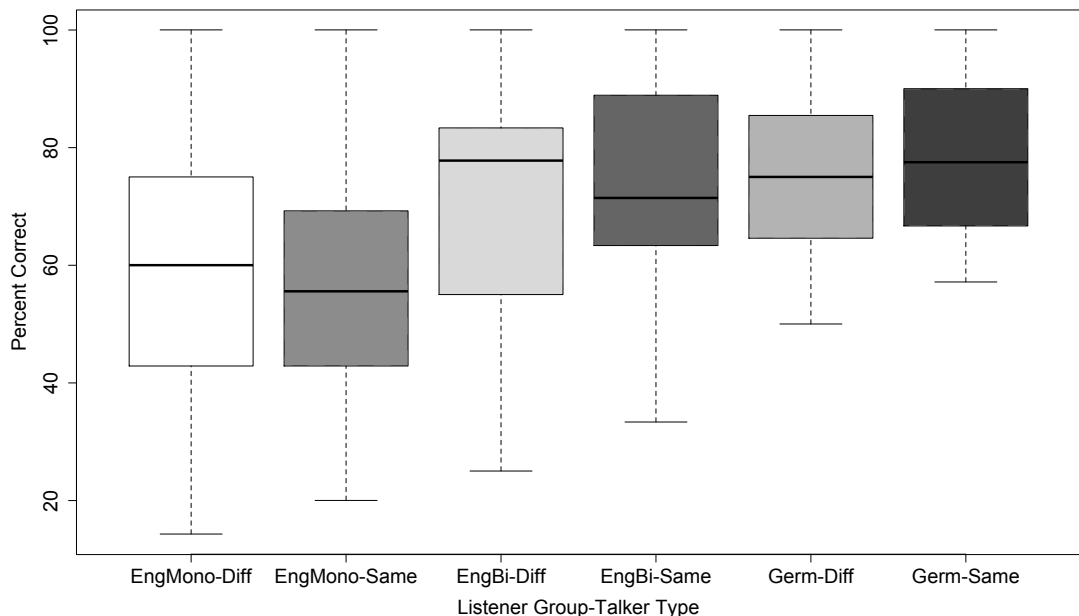


Figure 2. Percentage of correct identifications of repeated stimuli containing velar fricatives, by repetition type and listener group

The second pattern can be seen in both Figures 3 and 4. Here, native English listeners exhibited more difficulty in recognizing repeated words produced by a different voice than repeated words produced by the same voice as on their initial presentation. This pattern emerged for words containing both final /r/ and long, mid-vowels. The German listeners did not exhibit a similar drop-off in recognition accuracy across the two different repetition types for the same sets of words. Moreover, both groups of native English listeners appeared to have an advantage over the German listeners in recognizing repeated words with mid vowels produced by the same speaker.

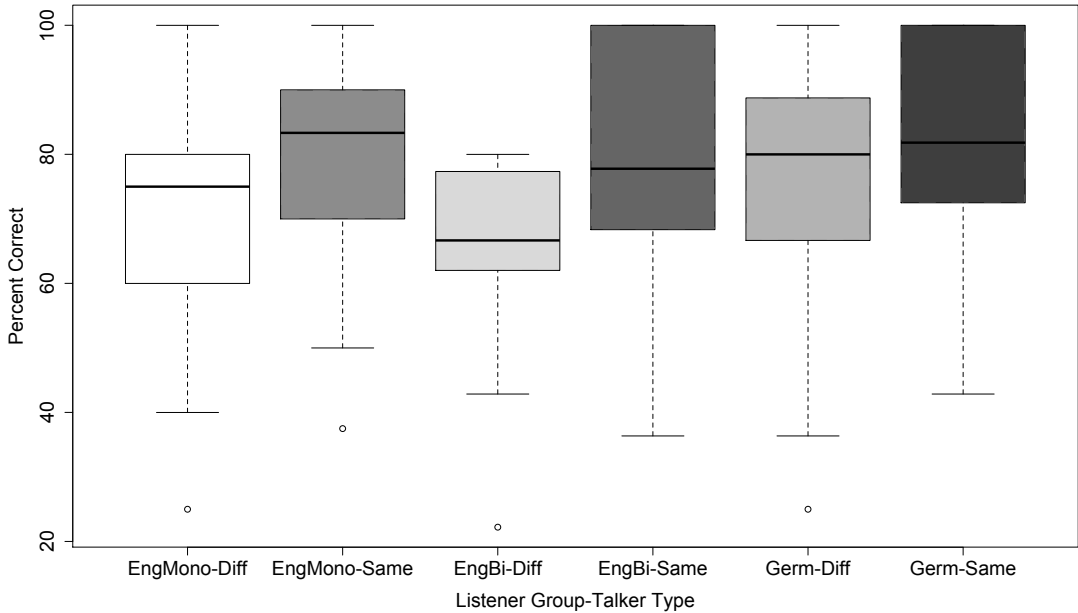


Figure 3. Percentage of correct identifications of repeated stimuli containing final /r/, by repetition type and listener group

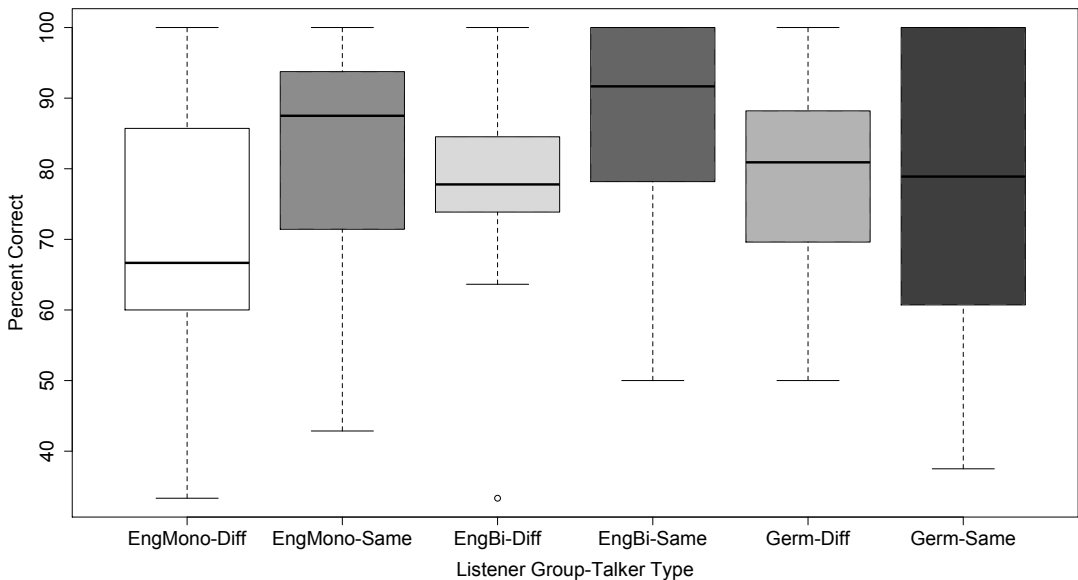


Figure 4. Percentage of correct identifications of repeated stimuli containing mid vowels, by repetition type and listener group

4. Discussion

The results of this study show that voice information does affect word recognition even when listeners hear words from a language they do not know. Both monolingual English listeners and native English-speaking learners of German correctly recognized repetitions of German words more often when they were repeated in the same voice as on their initial presentation, rather than in a different voice. This finding echoes the earlier result of Trofimovich (2005), who found that L2 learners of Spanish have shorter response latencies in a repetition task for Spanish words repeated in the same voice, rather than in a different voice. The German listeners in this study also replicated this earlier continuous word recognition finding from Palmeri et al. (1993). Together, these results indicate that exemplar-style processing of L2 stimuli is possible, and crucially, it seems to emerge when listeners focus on the linguistic content of the signal, rather than its indexical properties. This combination of results thus yielded a shift between exemplar- and non-exemplar-style processing that was not found by Trofimovich (2005), where listeners exhibited a same-voice processing benefit regardless of whether they focused on the semantic or phonetic properties of a test word on its first presentation. The key to inducing different processing styles seems to be shift attention not between different aspects of the linguistic signal itself (meaning or sound), but between the indexical "medium" and the linguistic message.

Above and beyond the conjecture that listeners encode indexical details in their memories of individual presentations of words, exemplar theory holds that listeners recognize words by comparing them to these stored exemplars in memory. The activation that words induce is proportional to their similarity to particular exemplars in memory, and that activation is summed over all similar exemplars in memory. It is this summation function that enables frequency effects to follow straightforwardly (at least in principle) from the basic precepts of exemplar theory. Frequency effects emerged in the results of this study, but they painted a fairly complicated picture of how stored exemplars of speech might become involved with the process of word recognition, in both an L1 and an L2. When considering listener responses to all stimuli presented in the experiment (both old and new), the English bilingual listeners exhibited a positive effect of word frequency on their recognition scores: the more frequent a word, the more likely they were to correctly classify it as old or new. In exemplar terms, this likely occurred because they were able to link frequent words to more exemplars in memory, thus inducing more activation and making the words easier to recognize. No such frequency effect occurred for the English monolingual listeners, of course, because they had no similar store of experiences to draw upon in performing the German word recognition task. A similar frequency effect did not, however, emerge as expected for German listeners. Instead, these listeners counterintuitively posted the worst recognition accuracy scores for the most frequent German words. This result suggests that exemplar theory's basic story of how word recognition works might be an oversimplification. The theory might possibly be amended by conjecturing that the large number of exemplars stored in memory for highly frequent words may make it more difficult to isolate individual instances of those word as having been presented in the study. While this is just speculation, the interaction effect is telling: word frequency affects word recognition in a different manner for native speakers than it does for second language learners.

Interestingly, main effects of both lexical frequency and listener group washed out when words were presented to listeners more than once. Even within this subset of stimuli, however, lexical frequency did still interact with the type of voice in which repeated words were presented to listeners. It was significantly easier for listeners to recognize high- and mid-frequency words repeated in a different voice than it was for listeners to recognize low frequency words. No corresponding frequency effects were found for words repeated in the same voice. This finding suggests that low frequency words have a more fragile, talker-specific representation in memory. From an exemplar-theoretic point of view, this likely occurred because there are fewer similar exemplars in memory to connect those stimuli to, and the individual stimulus token heard in the experiment thus forms a more prominent part of the activated representation of the word. A presentation of the same word in a different voice would thus be harder to match up to this overly individualized representation than it would be to a more general representation that resulted from a larger sum of similar exemplars in memory. The second "non-effect" of frequency is interesting in that it shows that repetition of a word in the same voice effectively washes out frequency effects. In this case, it seems that recognition is based effectively on

the token of the word that was (most recently) heard in the experiment, rather than being distributed across the many traces of those words stored in memory that the listener heard prior to the experimental session.

The effects of German words' "English-likeness" on response accuracy also fall in line with the basic predictions of exemplar theory, in that they show the importance of listeners being able to relate words to previous linguistic experience in word recognition. Just as those listeners who spoke German were better able to remember high frequency German words, native English listeners were better able to remember words that sounded like English words. For the German listeners, however, this distinction made no difference: all of the words just sounded like they were from German.

More generally, these results also indicate that a listener's memory for words in a language increases in proportion to the listener's proficiency in that language. Native German listeners exhibited the highest level of word recognition accuracy out of all the participants, and English-speaking learners of German performed better than the monolingual English listeners. Proficiency also affected listeners' ability to encode particular types of sounds in memory. The monolingual listeners, in particular, had difficulty recalling words including velar fricatives; these sounds may have just been too exotic for these listeners to understand or recall. For other exotic sounds, non-native listeners exhibited more sensitivity to token-specific phonetic details. In particular, this was the case for tokens with more vocalic information, such as those including long mid vowels or ending in /r/ (which is often realized as a [ɐ] or [ə] off-glide in German). It is possible that non-native listeners encode highly talker-specific information for exotic vowels, or this result may reflect the inherent variability in the way German speakers produce these segments. In either case, this effect resembles the findings of Goldinger (1996), in that it was easier for the native German listeners to generalize across different tokens of these words because they could draw on a larger and more variable population of similar exemplars in memory.

5. Conclusion

In combination with the results of earlier studies such as Winters et al. (2008) and Levi et al. (2011), these findings suggest an apparent asymmetry in the way linguistic and indexical information interact in L2 speech perception. When learning to identify voices, listeners seem to ignore L2 linguistic information, but when learning to identify L2 words, listeners remember the voices in which they were spoken. This asymmetry poses a challenge to exemplar models which assume that both linguistic and indexical information are stored in unanalyzed, integrated form in memory. Rather than simply putting indexical and linguistic information on the same level, listeners may interpret indexical information as more fundamental to the task of speech perception, and use it as the "medium" through which linguistic information may be understood. The results of this study also demonstrate the utility of testing the precepts of exemplar theories of speech perception with second language stimuli. Determining how listeners process tokens from the linguistic unknown provides an effective way of investigating the phonetic, semantic and task-based limits on which details of speech listeners are able to encode in--and retrieve from--memory.

References

- Abercrombie, David (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University.
- Abramson, Arthur S. and Lisker, Leigh (1970). Discriminability along the voicing continuum: cross language tests. *Proceedings of the 6th International Congress of Phonetic Sciences, Prague*, 569-573.
- Baayen, R. Harald, Piepenbrock, Richard, and Gulikers, Leon (1995) *The CELEX lexical database (Release 2) [CD-ROM]*. Linguistic Data Consortium, University of Pennsylvania [Distributor], Philadelphia, PA.
- Bradlow, Ann R., and Pisoni, David B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, 106(4), 2074-2085.
- Goggin, Judith P., Thompson, Charles P., Strube, Gerhard and Simental, Liza R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, 19 (5), 448-458.
- Goldinger, Stephen D. (1996). Words and Voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 1166-1183.

- Johnson, Keith A. (2007). Decisions and mechanisms in exemplar-based phonology. in *Experimental Approaches to Phonology*, In Honor of John Ohala (M.J. Sole, P. Beddor & M. Ohala, eds.) Oxford University Press, 25-40.
- Levi, Susannah V., Winters, Stephen J., and Pisoni, David B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, 121, 2327-2338.
- Levi, Susannah V., Winters, Stephen J., and Pisoni, David B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *Journal of the Acoustical Society of America*, 130 (6), 4053-4062.
- Nygaard, Lynne C., Sommers, Mitchell S. and Pisoni, David B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- Palmeri, Thomas J., Goldinger, Stephen D., and Pisoni, David B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 309-328.
- Perrachione, Tyler K., Pierrehumbert, Janet B. and Wong, Patrick C. (2009). "Differential neural contributions to native- and foreign-language talker identification." *Journal of Experimental Psychology - Human Perception and Performance*, 35, 1950-1960.
- Strange, Winifred and Dittman, Sibylla (1984). Effects of discrimination training on the perception of /r/-/l/ by Japanese adults learning English. *Perception & Psychophysics*, 36, 131-145.
- Trofimovich, Pavel. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, 26, 479-504.
- Winters, Stephen J., Levi, Susannah V., and Pisoni, David B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*, 123 (6), 4524-4538.

Selected Proceedings of the 2011 Second Language Research Forum: Converging Theory and Practice

edited by Erik Voss,
Shu-Ju Diana Tai, and Zhi Li

Cascadilla Proceedings Project Somerville, MA 2013

Copyright information

Selected Proceedings of the 2011 Second Language Research Forum:
Converging Theory and Practice
© 2013 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-458-4 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Winters, Stephen, Karen Lichtman, and Silke Weber. 2013. The Role of Linguistic Knowledge in the Encoding of Words and Voices in Memory. In *Selected Proceedings of the 2011 Second Language Research Forum*, ed. Erik Voss et al., 129-138. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2911.