OPEN-SET IDENTIFICATION OF NON-NATIVE TALKERS' LANGUAGE BACKGROUNDS

Elizabeth A. McCullough

Ohio State University
eam@ling.ohio-state.edu

ABSTRACT

Listeners are skilled at detecting native talkers of a language, but can they identify specific non-native language backgrounds? Open-set identification was used to explore this question. Eighty monolingual American English-speaking listeners labeled the language backgrounds of 30 talkers with 5 different native languages (L1s) on the basis of syllable- and word-length samples of English. As expected, listeners often identified L1 American English talkers correctly, despite hearing extremely short auditory stimuli. While listeners were sometimes unwilling to assign labels to L1 Korean, L1 Spanish, and to some extent L1 Mandarin talkers, L1 Hindi talkers were labeled frequently, and often the labels were correct. Responses revealed that listeners perceived many more language backgrounds than were actually represented by the talkers, a result which cannot be conveyed by the closed-set identification tasks commonly used. Impacts of listeners' perceptions of talker language background on cross-cultural communication are discussed.

Keywords: speech perception, accent identification, foreign accent, L2 speech

1. INTRODUCTION

When individuals speak in a language that is not their native language, they are often said to pronounce it with a foreign accent. Native listeners of the target language are skilled at distinguishing native from non-native talkers ([1], [2]). However, while this binary distinction receives much attention, listeners' ability to distinguish among non-native talkers of different backgrounds is less commonly investigated.

One approach that has been used to explore listeners' sensitivity to different varieties of non-native speech is free classification ([3], [4]), in which listeners have proven to be more accurate at grouping together talkers from some language backgrounds than from others. However, this approach does not reveal what listeners believe the talkers' language backgrounds to be, a question which is generally addressed using closed-set identification tasks. For instance, English-speaking listeners tested by Derwing and Munro [5] heard phrase-length stimuli in English and identified whether each talker's native language was Cantonese, Japanese, Polish, or Spanish. Performance was consistently above chance, ranging from 41% correct on L1 Japanese talkers to 63% correct on L1 Cantonese talkers. Vieru et al. [6] conducted a similar task with French-speaking listeners, who heard 10-second excerpts of French and chose whether each talker's native language was Arabic, English, German, Italian, Portuguese, or Spanish. Responses were 52% accurate overall, ranging from 25% correct on L1 Portuguese talkers to 77% correct on L1 Arabic talkers. In a similar task that also included native French talkers, listeners correctly identified the native talkers 96% of the time.

Listeners' skill at correctly identifying native talkers of their own language is unsurprising, given their ability to accurately detect and group together native talkers in other tasks. However, correct identification rates for particular non-native backgrounds vary. Closed-set identification limits responses in that the labels available to the listener are only those chosen by the experimenter, which tend to be those that actually describe the stimuli. If listeners believe that they know a talker's language background but do not see it represented among the labels provided, their answers cannot appropriately reflect their perceptions. Additionally, if listeners have no idea what a talker's language background is but choose the correct response option by process of elimination, their knowledge may be overestimated by their performance.

In the present study, each listener volunteered talker language background labels freely, rather than choosing from a set list. This approach provides an important counterpoint to closed-set language background identification tasks: without suggestions, can listeners accurately identify talkers' language backgrounds at all, or are hints from a fixed list important for their success? The auditory stimuli in this study consisted of isolated consonant-vowel sequences and short words, and were substantially shorter than stimuli used previously ([5], [6]). Overall, the task described in this work presented a difficult speech perception challenge.

2. METHODS

An open-set identification task was included in a larger experiment to collect listeners' responses regarding talkers' native language backgrounds.

2.1. Listeners

Eighty monolingual native speakers of American English participated in this experiment for partial course credit.

2.2. Stimulus materials

An English word list of disyllabic trochees beginning with all combinations of the stop consonants /b, d, g, p, t, k/ followed by the vowels /i, ɪ, ɛ, æ, ʌ, u, ɚ, eɪ, oʊ, aɪ/ was recorded by 3 female and 3 male talkers from each of the following native language (L1) backgrounds: American English, Hindi, Korean, Mandarin, and Spanish. Thus, the 30 talkers included native American English talkers as well as non-native talkers from 4 language backgrounds.

For the open-set identification task, one word was selected for each of the 30 talkers such that no word was repeated among the 30 talkers and no initial consonant or vowel was repeated among the 6 talkers from a given language background. The words thus selected are shown in Table 1. In the task, half of the listeners heard the entire disyllabic word, while the others heard only the consonant-vowel (CV) sequence extracted from the beginning of each word.

Table 1: Words selected for use as stimuli.

Talker L1	Words
L1 American English	beagle, duping, gable, pepper, tagging, kibble
L1 Hindi	buddy, dipper, geeky, purple, tubing, cackle
L1 Korean	bootie, Debbie, gutter, piking, turkey, keeper
L1 Mandarin	bidder, dirty, guiding, paper, Toby, kegger
L1 Spanish	bedding, dating, goading, pity, tiger, couple

2.3. Procedure

Each listener performed the task on a computer while wearing headphones. Talkers were represented by small rectangular icons on the screen, which the listeners could click on to hear the talker and could drag around the screen. Each listener was instructed to rearrange the icons on the screen such that talkers who had the same native language were grouped together. The grouping procedure and results are discussed further in [4]. After the grouping task was completed, each listener was asked to identify, if possible, the native language of each group created. While some listeners were unable to provide a response for every group they created, only 1 of the 80 participants assigned no labels at all.

2.4. Analysis

As the grouping analysis detailed in [4] is rather complex, the present work focuses only on the labels, and assumes that the label assigned to the group is the label that the listener would have assigned to each individual member of the group. As common group membership was meant to indicate shared language background, the results from this approach should not differ substantially from those that would be obtained from an individual talker labeling task. In some cases, however, talkers may have "inherited" labels from similar-sounding talkers. This issue is addressed in the discussion.

In order to discuss the results, listeners' open-ended responses must be organized in some way. Table 2 includes the categories used for analysis and the responses given by listeners that were included in each category. Note that there is a category for "East Asian" labels rather than for "Korean" and "Mandarin" separately. Many listeners seemed to have trouble narrowing their labels down to a single East Asian language, as demonstrated by the numerous responses listing multiple languages, as well as the generic label of "Asian."

Table 2: Categorization of listeners' labels.

Category	Listeners' labels
native	"English," "American English"
Indian/Hindi	"Indian," "Hindi," "Indian English," "Hindu," "a language in India"
East Asian	"Chinese," "Mandarin," "Korean," "Asian," "Chinese/Japanese," "Chinese/Japanese/Korean," "Chinese/Asian," "Chinese/Taiwanese," "Mandarin/Chinese/Korean," "a language in Asia"
Spanish	"Spanish"
other	"African," "Arabic/Middle Eastern," "Australian English," "British English," "Dutch," "French," "German," "French/German/European," "Irish," "Irish English," "Italian," "Russian," "Scottish," "Slavic," "Swahili," "Swedish," "Swiss," "electronic"

3. RESULTS

Listeners' open-set labeling responses were examined in light of two questions: how accurately talkers were labeled, and how often talkers were labeled at all.

3.1. Label accuracy

Rates of correct label assignment out of all response opportunities are presented in Table 3. Labels from the "East Asian" category were considered correct for talkers from both the L1 Korean and L1 Mandarin groups. It is important to remember that there is no chance level for open-set identification: every correct label was volunteered by a listener rather than chosen from a list.

To test the effects of language background and stimulus length on label accuracy, a 2-way repeated-measures ANOVA was performed on the correct label counts per talker, with language background as a between-subjects variable and stimulus length as a within-subjects variable. The main effect of language background was significant ($F(4,25) = 24.7$, $p < 0.001$). Post hoc Bonferroni-corrected t-tests for each possible pairing of language backgrounds revealed that correct labels were provided more often for L1 American English talkers ($M = 70\%$) than for L1 Korean, L1 Mandarin, and L1 Spanish talkers ($M = 13\%$, 17%, and 12%, respectively), while responses for L1 Hindi talkers ($M = 36\%$) were not different from any other group. The main effect of length was also significant ($F(1,25) = 12.8$, $p < 0.01$), with more correct labels for words ($M = 34\%$) than for CVs ($M = 25\%$). Finally, the interaction of language background and stimulus length was significant ($F(4,25) = 3.1$, $p < 0.05$). Paired t-tests indicated no significant differences based on stimulus length for any of the non-native talker groups, with marginally better performance for words than for CVs produced by L1 American English talkers ($p = 0.05$). Within most language backgrounds, however, the effect of stimulus length trended in the direction of the significant main effect.

Table 3: Percent of correct labels assigned to CVs and words for talkers from each background.

Talker L1	Correct label category	CVs	Words
English	native	57	83
Hindi	Indian/Hindi	33	40
Korean	East Asian	10	15
Mandarin	East Asian	13	21
Spanish	Spanish	12	12

For native talkers, the results are far from the nearly perfect performance observed previously for closed-set identification [6], likely because the auditory stimuli were substantially shorter than the 10-second excerpts used before. Nonetheless, listeners were reasonably accurate, with 57% correctly identifying native talkers based on CVs and 83% based on words. The labels provided for talkers from L1 Korean, L1 Mandarin, and L1 Spanish backgrounds were consistently less accurate, ranging between 10% and 21% correct. L1 Hindi talkers were correctly identified 33% of the time based on CVs and 40% based on words, neither significantly worse than native talkers nor significantly better than other non-native talkers. Listeners exhibited a range of performance across talker language backgrounds, just as in other tasks. Overall, longer stimuli were more often correctly identified; that is, hearing an additional syllable allowed listeners to better recognize a talker's language background.

As listeners did not often correctly identify the backgrounds of L1 Korean, L1 Mandarin, and L1 Spanish talkers, an important next step is to examine what responses were given instead. The most common responses for talkers from these groups were not any of the categories described in Table 2, but rather the absence of any label at all. In the standard forced-choice, closed-set approach to language background identification, not providing a label is generally not an option, although it was possible in the present task. In the following section, listeners' assignments of any label, whether correct or incorrect, are examined.

3.2. Label assignment

Rates of label assignment (including both correct and incorrect labels) out of all response opportunities are presented in Table 4. To test the effects of language background and stimulus length on label assignment, a 2-way repeated measures ANOVA was performed on the label counts per talker, with language background as a between-subjects variable and stimulus length as a within-subjects variable. The main effect of language background was significant ($F(4,25) = 11.7$, $p < 0.001$). Post hoc Bonferroni-corrected t-tests for each possible pairing of language backgrounds revealed that labels were provided more often for L1 American English talkers and L1 Hindi talkers ($M = 83\%$ and 77%, respectively) than for L1 Korean and L1 Spanish talkers ($M = 64\%$ and 63%, respectively). Label assignment frequency for L1 Mandarin talkers ($M = 68\%$) did not differ from that for any other group. The main effect of stimulus length was not significant ($p > 0.05$), while the interaction between language background and

stimulus length was significant ($F(4,25) = 5.0$, $p < 0.01$). Paired t-tests indicated no differences based on stimulus length for L1 Korean, L1 Mandarin, or L1 Spanish talkers, but more labels were provided for words than for CVs produced by L1 Hindi talkers ($t(5) = -3.4$, $p < 0.05$) and marginally for those produced by L1 American English talkers ($p = 0.06$).

Table 4: Percent of responses to CVs and words that included labels (correct or incorrect) for talkers from each background.

Talker L1	CVs	Words
English	77	89
Hindi	71	83
Korean	67	61
Mandarin	68	68
Spanish	66	60

This analysis examines an option generally unavailable in the forced-choice, closed-set tasks used previously: the ability to refrain from giving a label if uncertain about what the label should be. For L1 American English and L1 Hindi talkers, listeners employed this option relatively infrequently, providing labels at least 71% of the time. For L1 Korean and L1 Spanish talkers, this option was exercised significantly more often, in at least one-third of responses. L1 Mandarin talkers were often not assigned labels, although statistically the frequency did not differ from any other language background group. Additionally, for L1 American English and L1 Hindi talkers, hearing two syllables rather than one meant that listeners assigned labels more often. However, for L1 Korean, L1 Mandarin, and L1 Spanish talkers, hearing a word rather than a CV did not affect label assignment frequency. Overall, listeners seemed to have clear ideas about the language backgrounds of L1 American English and L1 Hindi talkers, while for the remaining groups, especially L1 Korean and L1 Spanish, the labeling task was more challenging.

4. DISCUSSION

In open-set identification, as in many other tasks, listeners are skilled at identifying native as opposed to non-native talkers. In contrast, as found in other tasks with multiple non-native backgrounds, their performance on non-native talkers varies considerably. Moreover, when listeners more frequently provided labels for talkers of some language background, those labels were often accurate, as shown by the results for L1 American English and L1 Hindi talkers in Sections 3.1 and 3.2. This pattern offers support to the assumption that listeners' willingness to provide a label was related to their confidence that the label was correct.

One question raised by the results presented above is why listeners performed so well with L1 Hindi talkers. As students, the listeners may have been accustomed to hearing Indian English, as India is the second most common country of origin of international students in United States universities [7]. However, as China and South Korea are first and third on the same list, respectively, the findings are unlikely to be wholly attributable to differences in familiarity with these non-native varieties of English. L1 Hindi talkers were also judged by American listeners to sound more accented than L1 Korean or L1 Mandarin talkers [8]. While identifiability of language background may be related to accentedness, further investigation would be required to detail the nature of any such link.

While the accuracy rates for L1 Korean, L1 Mandarin, and L1 Spanish talkers were rather low, due in part to listeners' reluctance to assign labels to these talkers at all, it should be remembered that the CV- and word-length stimuli were considerably shorter than the phrases or 10-second excerpts used in previous closed-set identification studies. As mentioned above, the present task was not purely a talker labeling task; listeners first grouped together talkers who were perceived to share a native language, then assigned a label to the entire group. This design might have made the task somewhat easier, in that individual talkers who might have been difficult to label independently could have "inherited" labels from talkers in the same group. Future work with longer auditory stimuli, in which each talker is individually assigned a language background label, will be more directly comparable to the closed-set tasks described above.

One detail that is impossible to observe in a closed-set task is the diversity of listeners' reported perceptions of talker language background. The various labels which were categorized as "other" in Table 2 show that listeners gave many responses which were not actually among the language backgrounds represented. While listeners perform well in closed-set identification tasks in that they often identify language backgrounds at rates above chance, their actual perceptions may be much more nuanced than a small, predetermined set of response options can capture.

In experiences outside the laboratory, listeners' perceptions of talker language background may influence how they treat talkers ([9]) and even how well they comprehend them ([10]). Thus, a better understanding of listeners' perceptions of language background is important in negotiating cross-cultural social interactions in today's world.

7. REFERENCES

[1] Flege, J. E. 1984. The detection of French accent by American listeners. *J. Acoust. Soc. Am*. 76, 692-707.

[2] Park, H. 2013. Detecting foreign accent in monosyllables: The role of L1 phonotactics. *J. Phonetics* 41, 78-87.

[3] Atagi, E., Bent, T. 2013. Auditory free classification of nonnative speech. *J. Phonetics* 41, 509-519.

[4] McCullough, E. A., Clopper, C. G. Auditory free classification of words and syllables in native and non-native English. Submitted.

[5] Derwing, T. M., Munro, M. J. 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Stud. Sec. Lang. Acq.* 20, 1-16.

[6] Vieru, B., Boula de Mareuil, P., Adda-Decker, M. 2011. Characterisation and identification of non-native French accents. *Speech Commun*. 53, 292-310.

[7] Institute of International Education. 2014. Open Doors report on international educational exchange.

[8] McCullough, E. A. 2013. Perceived foreign accent in three varieties of non-native English. *Ohio State U. Working Papers in Linguistics* 60, 51-66.

[9] Gluszek, A., Dovidio, J. F. 2011. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Pers. Soc. Psychol. Rev.* 14, 215-237.

[10] Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of non-native English-speaking teaching assistants. *Res. High. Educ.* 33, 511-531.