# EFFECTS OF TALKER ORDER ON ACCENT RATINGS

Charlotte Vaughn, Melissa Baese-Berk

University of Oregon
*cvaughn@uoregon.edu, mbaesebe@uoregon.edu*

## ABSTRACT

Despite the longstanding use of accent ratings in the literature, factors affecting those ratings have been relatively underexamined. Although accent ratings are often assumed to be stable properties of talkers, some factors like modality and task order have been shown to influence ratings. The present study asks how susceptible accent ratings are to contextual effects, particularly those arising from the ordering of talkers within the rating task.

Sixty-two listeners rated the accentedness of six talkers' English productions of a short read passage. We examine whether ratings of one L1-Chinese talker change across orderings. Results indicate that the target talker was rated as more accented when presented first than last, and that participants' ratings of the target were less consistent after being exposed to other talkers first. Talker order impacted accent ratings enough to change the rank ordering of non-native talkers' accentedness, raising methodological and theoretical implications.

**Keywords**: Accentedness, accent rating, non-native speech, speech perception

## 1. INTRODUCTION

Accent ratings are frequently used in the literature as a measure of a talker's proficiency, and as an independent verification of the relative proficiency of various talkers within an experiment (e.g., a "high accentedness" vs. "low accentedness" talker). However, despite the longstanding use of accentedness in these ways, factors affecting accent ratings are relatively underexamined.

A primary goal of accent rating tasks in the literature is to use a group of listeners' accent ratings from a single or small number of speech samples to represent a talker's level of accentedness, typically without consideration of factors like whether listeners' ratings may change during the course of the task. In this way, accentedness is implicitly assumed to be a stable property of a talker. However, several factors have been shown to impact accent ratings. For example, presentation modality of the stimuli (audio vs. audiovisual, [9]), the presence or absence of orthography [3], and other tasks being performed (e.g., multiple responses to a single stimulus [5];

multiple exposures to a single stimulus [4]) have been found to influence accent ratings. And, accent ratings have been shown to have poor interrater reliability [7]. Beyond task-based context, listeners' expectations about the speaker can also affect accent ratings [6]. Given these results, it seems clear that accent ratings are context-dependent, and may be affected by factors like the design of the task, the listeners, the speech samples used, and other considerations.

In the present study, we investigate one small piece of this question, and ask how susceptible accent ratings are to the order in which a target talker is presented to listeners. We present a non-native target talker either first or last among a series of native and non-native talkers, and investigate whether accent ratings significantly shift as a function of presentation order.

How could talker order affect accent ratings of a target non-native talker? It may be that it is listeners' implicit comparisons between talkers that matters; for example, if listeners incorporate native speakers into their accent rating scale, higher accent ratings might be assigned to a non-native target talker if heard last compared to if heard first. Or, it may be that there is a general task effect, with accent ratings increasing or decreasing over the course of the experiment, as suggested by [2] in a different domain. In terms of variability, if the target talker is presented last, participants may use the initial talkers to calibrate their rating scale in a consistent way, showing less variation in ratings than if presented first. Alternatively, participants' calibrations of the target to other talkers may not be systematic, resulting in greater variation in ratings of the target when presented last than when presented first. Regardless of directionality or its underlying cause, if accent ratings do shift depending on talker order, this indicates that talker order is a part of the contextual information that listeners factor into their assessments of talkers' accentedness.

## 2. METHOD

### 2.1. Participants

Participants were 62 listeners who were students at University of Oregon and received partial course credit for their time. All participants were either

native speakers of English ($N = 60$), or learned English at a young age (under 7, $N = 2$). No participants reported having uncorrected hearing loss.

## 2.2. Stimuli

Participants heard six male talkers from the ALLSSTAR corpus [1] produce the "North Wind and the Sun" paragraph in English. Talkers included four native Mandarin speakers (talkers 05, 21, 32, 35) and two native speakers of American English (talkers 50, 53). One native Mandarin speaker (talker 35) was chosen at random to be the target talker.

## 2.3. Procedure

Each participant listened to each of the six talkers read the "North Wind and the Sun" paragraph using a custom online presentation interface. Participants were requested to wear headphones. After each passage, listeners were asked to rate each talker's accentedness on a scale of 1 to 9, where 1 corresponded to "no accent" and 9 corresponded to "very strong accent". Participants could listen to each talker's reading passage only one time, but could adjust their rating before submitting it.

Participants were assigned to one of two conditions: Target Talker First or Target Talker Last, where the randomly chosen native Mandarin talker was presented either first or last among the six talkers, respectively. Thirty-one listeners were assigned to each condition. Within each condition, participants were assigned to a further randomization subcondition, which counterbalanced the order of the other five talkers ($N = 5$ or 6 per randomization order).

After the experiment, participants answered a series of demographic questions, as well as some questions about their reported experience with foreign-accented English (which is not discussed further in this paper for sake of space).

## 2.4. Analysis

Accent ratings were converted to z-scores on a by-participant basis to normalize for the range of values used by individual listeners.

## 3. RESULTS

To situate the target talker alongside the other five talkers as rated by listeners, Figure 1 presents a box plot of by-subject means of raw accent ratings for each of the talkers. Collapsing across all 10 conditions (Target First vs. Target Last x 5 talker randomization orders), the target talker (35) was rated as the second most accented talker in the study

overall. The target talker also shows relatively low variability in accent ratings, potentially because, by design, the target talker's order of presentation was less varied across conditions than all other talkers (discussed more below).

**Figure 1**: Raw accent ratings for all talkers across conditions. Non-native English speakers = dark blue, native speakers = light blue. Target = 35.
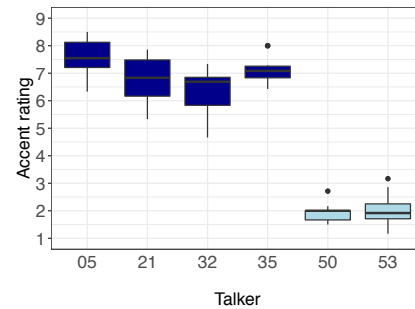


## 3.1. Target talker accent ratings when heard first versus last

Turning to our main question of interest, whether ratings of the target talker differed according to whether his sample was presented first or last, Figure 2 compares by-subject means of z-scored accent ratings for the Target Talker First and Target Talker Last conditions. Examining this figure, it is clear that participants in the Target Talker First condition (M = 0.79, SD = 0.35) reported the target talker as being more accented than listeners in the Target Talker Last (M = 0.54, SD = 0.40) condition. An ANOVA supports the observation that participants in these two conditions rated the talker differently, $F(1, 60) = 7.239$, $p = 0.009$, $\eta_G^2 = 0.108$.

**Figure 2**: Z-scored accent ratings for the target talker across conditions.
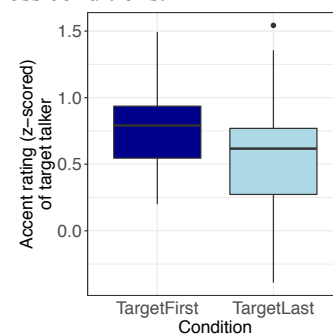


Interestingly, examining Figure 2 and the standard deviations reported above, not only do listeners in the Target Talker Last condition rate the target talker as less accented, but ratings of the target talker across listeners are also more variable when the target is last than when the target is first.

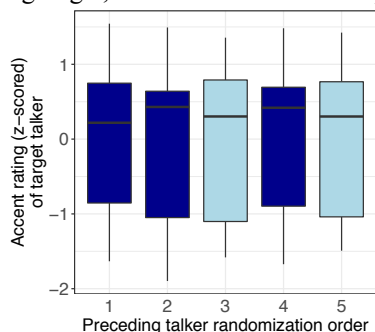## 3.2. Target accent ratings by preceding talker order

It is possible that the increased variability and lowered ratings in the Target Talker Last condition are a function of the specific ordering of talkers leading up to the target talker. Thus, we examine accent ratings within the Target Talker Last condition by randomization order. As described above and shown in Table 1, there were 5 possible randomization orders of talkers in the Target Talker Last condition (rows #1-5). One might expect that whether the talker immediately preceding the target was a native speaker or not may matter most; note that in randomization orders #3 and 5 in this condition, a native speaker was presented immediately before the target talker (in fifth position), and orders #1, 2, and 4 had a non-native talker in that position.

**Table 1**: Serial position of talkers for each randomization order in Target Talker Last condition. Light grey = native speakers (N), unshaded = non-target non-native speakers (NN), dark grey = target.

| Order | First | Second | Third | Fourth | Fifth | Sixth |
|-------|-------|--------|-------|--------|-------|-------|
| 1 | 50 (N) | 53 (N) | 32 (NN) | 21 (NN) | 05 (NN) | 35 (NN) |
| 2 | 53 (N) | 32 (NN) | 50 (N) | 05 (NN) | 21 (NN) | 35 (NN) |
| 3 | 32 (NN) | 05 (NN) | 21 (NN) | 50 (N) | 53 (N) | 35 (NN) |
| 4 | 21 (NN) | 50 (N) | 05 (NN) | 53 (N) | 32 (NN) | 35 (NN) |
| 5 | 05 (NN) | 21 (NN) | 53 (N) | 32 (NN) | 50 (N) | 35 (NN) |

To examine the possibility that the specific talker sequence preceding the target talker affected accent ratings, Figure 3 presents by-subject means of z-scored accent ratings in the Target Talker Last condition as a function of the five randomization orders of preceding talkers given in Table 1.

**Figure 3**: Z-scored accent ratings for the target talker according to preceding talker orders in Target Talker Last condition, corresponding to order numbers given in Table 1. Light blue = native speaker immediately preceding target, dark blue = non-native speaker.



Examining the plot, no clear patterns emerge, even when considering whether the talker immediately preceding the target talker was a native or non-native speaker. An ANOVA of accent ratings of the target in the Target Talker Last condition confirms no

significant effect of randomization order (F(4, 26) = 1.045, p = 0.403).

### 3.3. Accent ratings across talkers by serial position

If randomization order of preceding talkers is not straightforwardly responsible, what else may be contributing to the observed difference between accent ratings of the target talker in the Target Talker First and Target Talker Last conditions? One possibility is that there are general task effects, where participants become systematically more or less tolerant, rating all talkers as more or less accented as the experiment progresses. In order to test this possibility, we also examined the accent ratings of all talkers who were *not* the target talker, as a function of when in the experiment their recording was heard (i.e., their serial position, first-sixth). As our main result showed (displayed in Figure 2), the target talker was rated as less accented when heard last (sixth) than when heard first; we ask now whether there is evidence that other talkers were also rated as less accented later in the experiment. In this analysis we compare non-target talkers with each other, and not with the target talker, since each non-target talker was heard in each of the serial positions in the same number of conditions. (That is, each non-target talker was first in 1 out of 10 conditions, last in 1 out of 10 conditions, and second, third, fourth, and fifth in 2 out of 10 conditions, whereas the target talker occupied the first and final positions for half the conditions.) Figure 4 shows the five non-target talkers' average accent ratings according to when in the experiment they were heard (averaged across all 10 conditions).

**Figure 4**: Z-scored accent ratings for non-target talkers according to serial position in experiment.
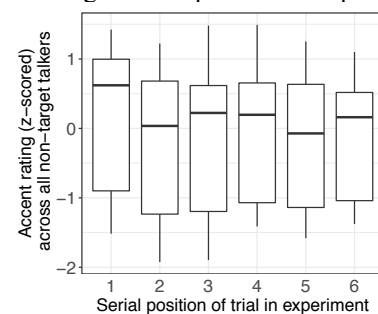


The pattern evident in the figure is supported by statistical tests. For instance, a linear mixed effects model with random intercepts for talker and for listener confirmed an effect of serial position (coded as a factor) on accent rating ($\chi^2 = 22.157$, $p < 0.0005$), and post hoc tests confirm that items rated in every serial position after the first were rated as significantly less accented than those heard first. In other words, regardless of talker, the item heard first

tended to be rated as more accented compared to those heard later in the experiment. Thus, it may be that at least part of the order effects observed for the target talker are due to this overall task effect, where listeners tend to penalize the first item the most.

### 3.4. Effect of talker order on accentedness rankings of all talkers

Whether the observed order effects are due to listeners' apparent tendency to become more tolerant after the first item, or whether they arise from implicit comparisons between preceding talkers that we have not uncovered here, it is clear that the order in which talkers are heard impacts their perceived accentedness. How much might these order effects matter? We note that one of the methodological uses of accent ratings in the literature is to as a tool to select talkers for a study; researchers obtain accent ratings for a group of talkers, and then select talkers with the highest or lowest ratings to represent "high" or "low" accentedness talkers more generally. Thus, the rank ordering of accentedness of talkers within a study can be consequential. We therefore ask how stable the rank orderings of the 6 talkers were across the 10 talker order conditions. Table 2 presents the order of talkers for each of the 10 conditions by accentedness.

**Table 2:** Rank ordering of the six talkers by condition. Light grey = native speakers (N), unshaded = non-target non-native speakers (NN), dark grey = target.

| Condition | Rank ordering of talkers from most accented (left) to least accented (right) | | | | | |
|---|---|---|---|---|---|---|
| TargetFirst 1 | 05 (NN) | 21 (NN) | 35 (NN) | 32 (NN) | 53 (N) | 50 (N) |
| TargetFirst 2 | 05 (NN) | 35 (NN) | 32 (NN) | 21 (NN) | 50 (N) | 53 (N) |
| TargetFirst 3 | 05 (NN) | 35 (NN) | 32 (NN) | 21 (NN) | 53 (N) | 50 (N) |
| TargetFirst 4 | 35 (NN) | 05 (NN) | 32 (NN) | 21 (NN) | 53 (N) | 50 (N) |
| TargetFirst 5 | 05 (NN) | 35 (NN) | 21 (NN) | 32 (NN) | 50 (N) | 53 (N) |
| TargetLast 1 | 21 (NN) | 05 (NN) | 32 (NN) | 35 (NN) | 50 (N) | 53 (N) |
| TargetLast 2 | 05 (NN) | 21 (NN) | 35 (NN) | 32 (NN) | 53 (N) | 50 (N) |
| TargetLast 3 | 35 (NN) | 05 (NN) | 32 (NN) | 21 (NN) | 50 (N) | 53 (N) |
| TargetLast 4 | 21 (NN) | 05 (NN) | 35 (NN) | 32 (NN) | 53 (N) | 50 (N) |
| TargetLast 5 | 05 (NN) | 32 (NN) | 35 (NN) | 21 (NN) | 50 (N) | 53 (N) |

Examining Table 2, we note that the target talker (35) changes position across conditions; the target is ranked as most accented in 2 out of 10 conditions, second-most accented in 3 of 10, third-most accented in 4 of 10, and fourth-most accented (i.e., least accented of all non-native speakers) in 1 out of 10. Further, the same rank ordering of talkers does not occur in more than 2 of the 10 conditions (and of those, once when the target talker was first, the other

when the target talker was last). These observations indicate that talker order has important and varied implications for the "meaning" of accentedness, and how it is interpreted, within a study.

### 4. DISCUSSION

The findings of this study add to previous results showing that accent ratings are not a stable property of a talker, but rather are context-specific judgments that take other factors into account [3, 4, 5, 6, 7, 9]. Specifically, the present results show that accent ratings are affected by the order in which a talker is presented in an experiment. In this study, when the randomly-chosen non-native target talker was presented first, they were rated as more accented than when that talker was presented last. This pattern was consistent for other talkers as well; talkers presented first were, in general, rated as being more accented than talkers presented later in the experiment. This finding suggests that studies may benefit from practice or "calibration" trials, though determining which talkers are used as practice is not a trivial matter. These results also suggest that the effect of talker order is not straightforwardly predictable listener-to-listener, since variability in ratings of the target talker increased when listeners had exposure to previous talkers, and since ratings of the target were not systematically shifted based on specific orderings of preceding talkers. The extent to which such effects would generalize to other target talkers is an important open question.

In future work, we plan to expand our examination of context effects to include which specific items were rated, whether those items are the same or different across talkers, and whether multiple shorter items from the same talker are blocked or randomized within the task.

In the meantime, the results presented here have both methodological and theoretical implications about accentedness as a construct. Theoretically, these results raise questions about the nature of accent ratings: What mechanisms result in listeners lowering accent ratings over time? And, to what extent do listeners' changing accent ratings during a task resemble the well-known process of adaptation in intelligibility? Is accentedness affected by all the same contextual factors as are known to impact intelligibility [8]? Methodologically, these results suggest that studies employing accent ratings should take into account the potential effects of talker order, potentially. Given that, in this study, the same set of talkers was ordered differently in their level of accentedness depending only on the order of presentation, accent ratings should be interpreted accordingly.

## 5. ACKNOWLEDGEMENTS

## 6. REFERENCES

[1] Bradlow, A. R., Ackerman, L., Burchfield, L. A., Hesterberg, L., Luque, J. S., Mok, K. (2010). ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings. Department of Linguistics, Northwestern University. Retrieved from *http://groups.linguistics.northwestern.edu/ speech_comm_group/allsstar/index.html*

[2] Dragojevic, M., Giles, H., Beck, A. C., Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385-405.

[3] Levi, S. V., Winters, S. J., Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *The Journal of the Acoustical Society of America, 121*(4), 2327-2338.

[4] Munro, M. J., Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 11*(3), 253-266

[5] O'Brien, M. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition, 38*, 587-605.

[6] Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511-531.

[7] Southwood, H., Flege, J. (1999). The validity and reliability of scaling foreign accent. *Clinical Linguistics & Phonetics*, *13*, 335-349.

[8] Tzeng, C. Y., Alexander, J. E., Sidaras, S. K., Nygaard, L. C. (2016). The role of training structure in perceptual learning of accented speech. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1-17.

[9] Yi, H. G., Phelps, J. E., Smiljanic, R., Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for non- native speech. *The Journal of the Acoustical Society of America, 134*(5), EL387-EL393.