Architekt or Architekt?
Perception of devoiced vowels produced by Japanese speakers of German

Frank Zimmerer1,2, Rei Yasuda1,3, Henning Reetz1

1 Institute of Phonetics, Goethe-University, Frankfurt, Germany
2 Institute of Phonetics, Saarland University, Saarbrücken, Germany
3 Graduate School of Language and Culture, Osaka University, Japan

zimmerer@em.uni-frankfurt.de, yasuda@em.uni-frankfurt.de, reetz@em.uni-frankfurt.de

Abstract

In Tokyo Japanese, vowel devoicing is a common process leading to the reduction of high, unstressed vowels, mostly when they occur between unvoiced consonants. Native Japanese speakers learning German show a strong tendency to produce these devoiced vowels in the foreign language, too, despite the lack of this regular process in that language, here in German. This study examines the extent to which this reduction process leads to perception problems by German listeners, who are confronted only rarely with devoiced vowels in their native language. Results of a phoneme detection task indicate that devoiced vowels may indeed lead to perceptual difficulties. German listeners seem to refrain from reconstructing the vowel completely, which also can add to a perceived foreign accent of Japanese productions by German listeners.

Index Terms: speech perception, devoiced vowels, L1-L2 interference

1. Introduction

When adults learn a foreign language, they rarely reach perfection in the pronunciation of that language. Their phonological system and their phonetic knowledge of their native language (L1) often interfere with the perception and production of sounds, words and sentences in the foreign language (L2) (e.g. [1], [2], [3]). Some of these processes creating interference can be completely absent in the target language (L2). This can lead to a foreign accent, which can even create problems when language learners talk with native speakers. Especially when L2 speakers produce L1 processes in the target language L2, where the process is not present, and subsequently listeners’ expectations for pronunciations of words (or segments, or sentences) are not met.

Different languages can have rather different vowel systems. In such cases, perceiving and producing vowels is affected by interference of L1 in L2-learning very often (e.g. [3]). Aside from vowel systems that can vary, (phonological or phonetic) processes that affect vowel pronunciations can also be different across two languages. One such example where the process of reducing vowels is common in L1 but very rare in L2 is high vowel devoicing in Japanese (in this case, L1) compared to German (here, L2). In unstressed syllables, Speakers of Tokyo Japanese tend to devoice the high vowels [i] and [u] regularly [e.g. 4,5,6] (subsequently, these vowels are referred to as <i> and <u>). In Tokyo Japanese, high vowels <i>, and <u>, are devoiced up to 90% of the time, when they occur in the correct context (e.g. [4,6]). On the other hand, there is no regular phonological process where high vowels are devoiced in unstressed positions in German.

Devoicing may occur, but it is not a regular process (e.g. [7]). The German vowel system consists of three high vowels (ignoring vowel length, and laxness), <i>, <y> and <u>. An investigation of Japanese learners of German indicates that Japanese speakers produce devoiced vowels in German words also quite regularly. A production experiment indicated that when Japanese speakers repeated words three times in a row, vowels were produced as devoiced 16% of the time. Some context showed a devoicing rate of 100%, other segmental contexts showed less reduction. Also, the rate of devoicing was depending on the vowel itself, with <i> being devoiced most often, <y> and <u> show this reduction pattern less frequently. Furthermore, since there are no [y] and [u] vowels in the Japanese vowel inventory, the latter is often mapped to [ui], the former to [i] when Japanese speakers produce German words containing those vowels. These results indicate that Japanese speakers show a clear interference of this L1 process when producing the L2 German ([8]).

The question that is investigated in this paper is, to what extent German listeners may have problems to understand words with devoiced vowels, because it is a process that occurs only rarely in their L1, other reduction processes are far more frequent [7,9]. To this end, a phoneme monitoring experiment was conducted. The items for this experiment were all German words (produced by German and Japanese speakers). Words as items were chosen, because listeners tend to be more lenient with imperfect input, when they can match it to a word (e.g. [1]). If German listeners have problems to perceive devoiced vowels, they should show different behavioral results between voiced and devoiced versions of words in this experiment. If there is a tendency towards poorer perception of items with devoiced vowels, Japanese language learners could be made aware of their production and could learn to produce vowels in the L2 with more voicing in order to be understood better and to reduce their foreign accent.

2. Method

2.1. Materials

For the phoneme monitoring experiment, 8 German words were chosen. Of these, 6 words contained the vowel /u/ (e.g. Zitrone ‘lemon’), 2 words had an /u/ (e.g. Milchzucker ‘milk sugar’). The different amount of items per vowel condition reflects the amount of vowel devoicing that can be observed by Japanese speakers in a production experiment, where results show that /u/ gets devoiced more often than /u/.

The eight words chosen for this experiment were presented to participants in three different conditions. They were produced by Japanese learners of German with either a devoiced (condition1) or a voiced vowel (condition2), and by
German native speakers with a voiced vowel (condition 3). For instance, the word Zitrone ('lemon') was uttered three times with a reduced vowel by a Japanese speaker, 2 times with a voiced /i/ vowel by a Japanese speaker and 2 times with a voiced /i/ by a German speaker. The voicing status of the vowels was assessed by the two authors of this article independently. Items were selected only when they agreed on the status of the voicing of the vowel. The assessment was rather conservative: vowels were rated as devoiced only when there was no voicing at all throughout the complete vowel.

Furthermore, for each word, two filler items were chosen (control), all produced by German speakers. These control items did not include the respective vowels and were used as baseline/control condition, to ensure participants understood the task and did not respond at random. Overall, 9 Japanese speakers and 3 German speakers produced the experimental items. All speakers were female, the words were originally recorded for a production experiment, thus, the items were not produced intentionally for this perception experiment. The Japanese speakers had studied German (Mean=2–4 years, ranging from 0.5–3.5, SD=0.9) and lived in Germany (Mean=19.2 months, ranging from 1–42, SD=18.2) at that time. They were on average 23.3 years old (ranging from 21–29, SD=2.8). The German speakers were speaking standard German. All speakers were recruited at the Goethe-University, Frankfurt. Overall, 72 words were chosen for the experiment, 24 items for condition 1, 16 items each for condition 2 and condition 3, as well as 16 control items. Table 1 shows the number of items that were chosen for each vowel and each condition.

Table 1. Number of items for each condition.

<table>
<thead>
<tr>
<th>Condition, L1 speaker</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/ reduced, Japanese</td>
<td>3*6 (18)</td>
</tr>
<tr>
<td>/i/ unreduced, Japanese</td>
<td>2*6 (12)</td>
</tr>
<tr>
<td>/i/ unreduced, German</td>
<td>2*6 (12)</td>
</tr>
<tr>
<td>/i/ – no /i/ German</td>
<td>2*6 (12)</td>
</tr>
<tr>
<td>/u/ reduced, Japanese</td>
<td>3*2 (6)</td>
</tr>
<tr>
<td>/u/ unreduced, Japanese</td>
<td>2*2 (4)</td>
</tr>
<tr>
<td>/u/ unreduced, German</td>
<td>2*2 (4)</td>
</tr>
<tr>
<td>/u/ – no /u/ German</td>
<td>2*2 (4)</td>
</tr>
<tr>
<td>SUM</td>
<td>9*8 (72)</td>
</tr>
</tbody>
</table>

Figure 1 shows the average duration of the words per condition and language (the non-vowel condition is ignored). Overall, items produced by German speakers had a mean duration of 657ms, the voiced items of Japanese speakers had a mean duration of 745ms, unvoiced versions were on average 725ms in length. Thus, the German productions were shorter than the Japanese ones. Concerning the Japanese voiced vowels, the <u> vowels were not produced with a clear [u], rather the words had an [u] – in accordance with the pronunciation of Japanese <u> (e.g. [6]).

For each vowel condition, a separate experimental list was created and every list was repeated 4 times with different (pseudo) randomizations. That is, each participant heard 288 items, and was expected to respond to 224 of them, and should not react to 64 items. The experimental lists were played subsequently to 10 participants. They were instructed to press a button as soon as they heard a certain vowel (e.g. /i/ in the first list, /u/ in the second), they were asked not to press any button when they did not hear the respective vowel. Reaction time measurement started with the beginning of the experimental item. An experimental trial consisted of a warning tone, 250ms of silence, the presentation of the word and another 1500ms of silence, before the next trial started.

In the experiment, 10 members of the Goethe-University, Frankfurt participated. They did not receive monetary reimbursement, but were paid in kind (coffee, tea, cookies). All participants were native speakers of German and did not report on any hearing problem. They were unaware of the purpose of the experiment.

**2.2. Factors of interest and predictions**

Japanese speakers show a clear tendency for L1 interference in the production of L2 words. This might lead to an audible foreign accent that impede the recognition of what has been said by German listeners. The first question thus is, to what extent words produced by Japanese speakers are harder to be perceived by German listeners than words produced by native German speakers. Since Japanese items with both voiced and devoiced vowels are part of the experimental lists, we expect that overall, words produced by German speakers are recognized easier than the words produced by Japanese speakers.

A second, related question is, whether there is a difference in perception when Japanese speakers devoice vowels compared to words of Japanese speakers with a voiced vowel. The process of vowel devoicing is rather rare for German listeners. This question is more specific than the first one. If there is a difference between voiced and voiceless vowels in the recognition rate, this could lead to adaptions in language learning courses for Japanese speakers in order to reduce their foreign accent.
Thirdly, the experimental set-up did enable us to investigate whether the unrounding (and slight fronting from \[u\] to \[ɯ\]) has an effect on vowel perception by German listeners. If there is a difference between the perception of the two vowels, we would expect that \(<u>\) items are recognized worse than \(<i>\) items, since the change (unrounding) in vowel is also not a common process in German.

3. Results

The responses of the participants were coded for CORRECTNESS. “0” was used for incorrect reactions (a wrong positive reaction occurred when there was a response to control items, or a wrong negative response was no response to a vowel condition) and “1” for correct reactions (no response to control items and a response to the vowel items). Concerning devoiced vowels, we labeled CORRECTNESS parallel to voiced ones, assuming listeners reconstruct a vowel. If they responded only on voicing and did not recognize a vowel at all, the response pattern should have been identical to the control condition, which, as can be seen below, was not the case. Generally, correct should not been interpreted literally, especially with respect to Japanese items, a non-response to a devoiced vowel or to a vowel \([u]\) instead of \([u]\) could also indicate that the vowel that was perceived did not concur with German listeners’ expectations concerning the acoustics of a given vowel. Or that the vowel was interpreted as absent, thus, listeners responded as they perceived the item. However this labeling enabled us to compare whether the words were treated the same or different as words produced by German speakers. Of the 10 participants, one had to be excluded since the error rate in control condition and condition3 (German voiced vowels) was higher than 15%. The data of the 9 remaining participants were analyzed further. Because the vowel lists were separated and the number of items was different in the lists, we first analyze each vowel condition separately before we combine the data. Figure 2 shows the percentage of correct responses for the different conditions and the different vowels, whereas Figure 3 gives an overview over the reaction time data, split up by conditions for the different vowels.

In the \(<i>\) list, participants had to respond to 1944 items and reached an overall mean accuracy of 85.13%. Concerning the different conditions, condition1 was correct in 62.2%, condition2 in 95.1%, condition3 in 95.6%, and the control condition in 99.1% of the responses. A linear mixed model was calculated for ACCURACY as dependent variable, with SUBJECT and ITEM as random variables, and CONDITION as factor. CONDITION was a significant factor (F(3,1927)=196.38, p<0.0001). A Student’s t post-hoc test indicated that all conditions were significantly different from each other. This shows that participants did not have problems with the task and were almost at ceiling in the “clear” conditions, where a voiced vowel was present (condition2 and condition3), or where the \(<i>\) was not there at all (control condition).

Next, a reaction time (RT) analysis was performed for the \(<i>\) items. In this analysis, the control condition was discarded, since the correct responses in this condition were those where no response was given and only correct responses were included (incorrect responses were those where no response was given in the other conditions – exclusion of 285 cases). Furthermore, responses faster than 200ms and slower than 1500ms were also discarded (22 cases). This left 1205 responses for the analysis. Overall, condition1 showed a mean reaction time of 865ms on average, condition2 of 866ms, and condition3 of 812ms. Again, the same model as for accuracy was calculated for RT as dependent variable. Here, as for accuracy, CONDITION was a significant factor (F(2,1191)=31.7, p<0.0001). A post-hoc Student’s t test indicated that all conditions were significantly different from each other.

Concerning \(<u>\) items, 648 responses went into the analysis. Again, accuracy is analyzed first, before the RT results are reported. For this experimental list, control items were responded to correctly in 92.3% of the time. Condition1 had a recognition rate of 49.1%, condition2 71.5%, and condition3 99.3%. As for \(<i>\>, a linear mixed model was calculated for ACCURACY with PARTICIPANT and ITEM as random factor and CONDITION as fixed factor. CONDITION was a significant factor (F(3,635)=71.4, p<0.0001). The post-hoc Student’s t test showed that the control condition was not significantly different from the German voiced condition (condition3), but that the Japanese items (condition1 and condition2) were significantly different.

Subsequently, the RT analysis was performed (as for \(<i>\>\)). Condition3 exhibited the fastest mean RT (850ms), condition1 was the slowest (978ms) and condition2 fell in between (939ms). The linear mixed model for RT (SUBJECT and ITEM as random factor, CONDITION as fixed) indicated that CONDITION was significant (F(2,341)=31.7, p<0.0001), and all conditions differed significantly from each other (Student’s t).

In the next step, \(<i>\>\> and \(<u>\>\> were analyzed together. In this analysis, the control condition was excluded. Condition1 had a mean accuracy rate of 58.9%, condition2 was responded to correctly in 89.2%, and condition3 in 96.5%. A linear mixed model for ACCURACY was computed. For this model, PARTICIPANT and ITEM were set as random factors. Vowel \(<i>\> or \(<u>\> and CONDITION (condition1, condition2, condition3) and the interaction of VOWEL and CONDITION were also entered into the analysis. Results indicate that CONDITION was a significant factor (F(2,1996)=214.8, p<0.0001), VOWEL was
not (p>0.05), whereas the interaction of VOWEL and CONDITION was significant (F(2,1996)=18.0, p<0.0001). Post-hoc Student’s t tests showed that all conditions differed significantly from each other. Planned comparisons concerning the interaction showed that devoiced Japanese vowels were treated identically, as were German voiced ones. However, while Japanese <i> voiced items patterned with German voiced vowels (<i> and <u>), Japanese voiced <u> did not, thus, Japanese voiced <i> and <u> were treated differently by German listeners.

The subsequent step was an analysis of the RT data without incorrect responses and without the control condition. Condition1 had a mean RT of 887ms, condition2 of 879ms, condition3 of 821ms. The linear mixed model (PARTICIPANT and ITEM as random factors, CONDITION, VOWEL and the interaction of VOWEL and CONDITION as fixed factors) was also calculated for RT. In this analysis, CONDITION (F(2,1537)=58.2, p<0.0001) and the interaction (F(2,1536)=5.6, p=0.01) were significant, VOWEL was not (p>0.05). The post-hoc test (Student’s t) indicated that all conditions differed significantly from each other. Planned comparisons investigating the interaction showed that German voiced vowels were not significantly different from each other, neither were Japanese voiced vowels, but Japanese devoiced vowels showed a significant difference in RT.

4. Discussion and conclusions

Our first factor of interest was whether German words produced by Japanese speakers were perceived differently by German speakers compared to German words uttered by German speakers. Results indicate that, overall, German listeners have a harder time to perceive Japanese speakers' productions. However, when the results are investigated in more detail, it becomes evident that if Japanese speakers produce vowels in a way that meets German listeners' expectations (i.e. voiced <i> condition), there is no difference between vowels produced by German speakers and vowels produced by Japanese speakers. Thus, regarding the second factor of interest (i.e. effect of devoicing), we found that it is vowel devoicing is crucial for perception by German listeners. Devoiced vowels, which are very rare in German, deteriorates perception by German listeners. This is true irrespective of which vowel gets devoiced (<i> and <u>).

Finally, the third factor of interest was the effect of changes in vowel quality. The results indicate that changes in vowel quality (i.e. from [u] to [ɯ]) also affect recognition by German listeners. Japanese <u> items were perceived worse than Japanese <i> items (even when they were voiced).

These results can help to improve language teaching for Japanese learners of German. Japanese does not have round high vowels (e.g. [6]). The fact that the phonological system of the native language leads to interference in the production of the foreign language is expected (e.g. [1] and [2]). Thus, special attention should be paid to the production of these vowels in language teaching. Rounding of vowels is important for German listeners, its absence leads to deterioration in understanding. Furthermore, Japanese learners of German could be made aware of the devoicing process of high vowels in Japanese and could be taught not to transfer this process to German (or other languages that do not have this process). This would decrease their accent in the L2 and would lead to a better comprehensibility.

5. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG – SPP 1234) and the German Academic Exchange Service (DAAD).

6. References