The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language*

Johanneke Caspers
Leiden University Centre for Linguistics

1. Introduction

According to Derwing & Munro (2005) there is generally not enough attention for pronunciation within the field of second language acquisition. This seems to be true, see for example *The Handbook on Second Language Acquisition* (Doughty & Long 2003) or *The Oxford Handbook of Applied Linguistics* (Kaplan 2002). Both are voluminous and comprehensive works on second language acquisition (SLA) research, but key notions such as pronunciation, prosody, intonation, stress and accent are — virtually or completely — missing from the index.

Pronunciation does seem an important aspect of second language acquisition, however, since incorrect pronunciation may harm successful communication. Especially with regard to suprasegmental phenomena such as word stress and sentence accent, knowledge of deviant pronunciation by second language speakers is limited, let alone its possible effects on intelligibility. The current paper tries to find a partial answer to the following general research question: how detrimental are suprasegmental errors to the intelligibility of nonnative speakers of Dutch?

In the recent literature on SLA that does pay attention to pronunciation the general view is held that errors with regard to prominence (e.g. word stress) are more important to intelligibility than segmental errors. This claim is tested in a perception experiment using data from French and Mandarin Chinese speakers of Dutch as a second language (DSL).
2. Background

Anderson-Hsieh, Johnson & Koehler (1992) investigated the influence of suprasegmental and segmental errors on the nonnativeness of L2 speakers. They were interested in the relationship between nonnatively sounding speech and actual deviances in segmental and prosodic characteristics. The pronunciation of sixty read-aloud speech samples from second language speakers of English with a wide range of mother tongues (L1s) were rated by trained jurors. Then the segmental errors in the samples were systematically analyzed. Finally, the prosodic structure of each sample was judged on a 4-point scale. The results showed that the prosodic variable had the strongest correlation with the global pronunciation ratings. This influential paper contributed to the current idea that suprasegmental errors affect intelligibility more than segmental errors. However, prosodic errors in the Anderson-Hsieh et al. study were not reliably labeled (they used only an overall impressionistic judgment), and the functional intelligibility of the L2 speech was not determined. This means that the empirical support for the claim that suprasegmental errors are more detrimental to intelligibility than segmental ones is not very strong. However, two recent studies by Hahn (2004, on sentence accent), and Field (2005, on word stress), both on English as a second language, do fill this gap to some extent.

Hahn (2004) examined the influence of erroneous sentence accents (Hahn uses the term ‘primary stress’) on native processing, comprehension and evaluation of L2 speech. She used a piece of discourse, recorded by an advanced nonnative (Korean) speaker of English in three versions: with correctly placed sentence accents, with incorrectly placed accents and without any accents. Results reveal no significant effect of the conditions on processing of the nonnative speech, but there is an effect on comprehension: the version with correctly placed accents can be recalled more accurately than the other versions, and the correct version is evaluated more positively.

Field (2005) tested the influence of misplaced word stress on intelligibility. He asked a native English speaker to pronounce disyllabic English words normally and with shifted lexical stress, sometimes accompanied by a change in vowel quality. The results show that the intelligibility for native listeners was compromised by stress shifting, and that the extent of degradation depended on the direction of the stress shift.

Research by Van den Doel (2006) into prevalent pronunciation errors made by Dutch speakers of English as a second language revealed that erroneous stress placement is judged by native speakers to be the most severe
pronunciation error. However, a strong nonnative accent per se does not automatically imply low intelligibility. Munro & Derwing (1999) used speech fragments produced by advanced speakers of English as a second language with Mandarin Chinese as L1. The fragments were tested for intelligibility (as measured by orthographic transcription), comprehensibility (a subjective judgment on a Likert scale) and accentedness (idem) by native listeners. Unintelligible utterances received lower comprehensibility and higher accentedness scores than intelligible utterances, but many of the highly intelligible utterances did receive a high accentedness score.

Based on a large amount of phonetic research Van Heuven (2008) predicts that stress contributes only modestly to word recognition, but that incorrect stress is detrimental when segmental quality is degraded. Combined with the fact that there is some evidence that suprasegmental errors affect intelligibility of nonnative speech and that segmental errors do not automatically deteriorate intelligibility, it seems reasonable to predict that stress errors influence intelligibility of nonnative speech more than segmental errors, but that a combination is the most harmful.

3. Research questions and approach

What is more harmful to the functional intelligibility of nonnative speakers of Dutch, stress errors or segmental errors? And what are the effects of these errors on the perceived comprehensibility and accentedness of the nonnative speech? To answer these questions, words spoken by learners of Dutch with French or Mandarin Chinese as a first language were presented to Dutch native listeners. The words contained either (1) a stress error and a segmental error, (2) a stress error, (3) a segmental error, or (4) no stress error or segmental error. Three different measurements were taken: functional intelligibility (by means of a transcription task), perceived comprehensibility and degree of foreign accent (both scored on a Likert scale). The first two measurements were combined in one experimental run, while the accentedness judgments were collected in a second experimental session (cf. Munro & Derwing 1999).

It was expected that incorrect stress (Condition 2) is more detrimental to intelligibility than incorrect segments (Condition 3). A combination of suprasegmental and segmental errors (Condition 1) was predicted to be more harmful to intelligibility than stress or segmental errors alone. Furthermore, it was expected that the materials produced by the Chinese DSL-speakers would
be judged as more accented and less comprehensible than the materials from
the French DSL-speakers. Not only because the phonological distance between
unrelated languages may be larger than between related languages, but also
because Dutch speakers supposedly are more accustomed to French accented
speech than to Chinese accented speech.

4. Method

4.1 Stimulus materials

A corpus of read-aloud materials was available from an earlier experiment
(Caspers & Van Santen 2006). Five Chinese, five French and five native speak-
ers of Dutch had read aloud the same text and word list. From these materials
words were selected fitting the conditions mentioned above. Care was taken to
choose different words each time, since they had to be used in an intelligibility
task. Repeated words are probably easier to understand than words that are
presented only once. The word candidates were judged as correct or incorrect
with respect to the realized stress position; words with unclear stress positions
were left out. The words were also inspected for segmental errors, resulting in a
phoneme change in Dutch; words with deleted or inserted phonemes were left
out. Below examples are given of stimulus words in the three error conditions
(the stressed syllable is given in bold face, the segmental errors are underlined):

Condition 1: stress error and segmental error

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
<th>Correct Version</th>
<th>Error Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>balak</td>
<td>‘barracks’</td>
<td>/ˈbɑlak/</td>
<td>/ˈbɑˈrak/</td>
</tr>
<tr>
<td>winnares</td>
<td>‘female winner’</td>
<td>/ˈwɪnərɛs/</td>
<td>/wɪnə'ɾɛs/</td>
</tr>
</tbody>
</table>

Condition 2: stress error

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
<th>Correct Version</th>
<th>Error Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>fluweel</td>
<td>‘velvet’</td>
<td>/ˈflywel/</td>
<td>/fly'wel/</td>
</tr>
<tr>
<td>bacterie</td>
<td>‘microbe’</td>
<td>/ˈbækˈtɛri/</td>
<td>/bæk'teri/</td>
</tr>
</tbody>
</table>

Condition 3: segmental error

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
<th>Correct Version</th>
<th>Error Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>schadeclam</td>
<td>‘insurance claim’</td>
<td>/ˈʃ kadəklam/</td>
<td>/ˈʃ kadəkləm/</td>
</tr>
<tr>
<td>kochen</td>
<td>‘wheat’</td>
<td>/ˈkoχən/</td>
<td>/ˈkoɾən/</td>
</tr>
</tbody>
</table>
Words in Condition 4 did not contain stress errors or phoneme changes, but this of course does not mean that they were pronounced in a completely nativelike manner.

We succeeded in finding 20 words per condition, resulting in 80 nonnative stimulus words (distributed fairly equally over the ten nonnative speakers). Native materials were added as a reference condition, but since the Dutch speakers did not make segmental errors, only words in Conditions 2 and 4 were available. Also, the number of stress errors was understandably low, and as a result there was a small overlap in selected stimulus words: seven of the native words containing a stress error were already present in the nonnative dataset. To prevent problems with intelligibility measurements, these seven words were all presented after those spoken by the nonnative speakers in the first session, in which intelligibility was tested.

4.2 Procedure

25 native speakers listened to the stimuli at a comfortable loudness over Quad ESL-63 electrostatic loudspeakers, while seated in a sound-treated lecture room. They were given answer booklets, which started with a set of questions (sex, age, dyslexia, hearing problems, knowledge of other languages, experience with DSL-speakers, etc.), followed by short instructions. The 100 stimulus words were presented twice, in a different quasi-random order. In the first part of the experiment subjects were asked to listen to each word and to write down what they had heard, and immediately after that to score the comprehensibility of the word on a 10-point scale (1 = utterly incomprehensible to 10 = perfectly comprehensible). In the second part of the experiment the subjects were asked to judge the measure of foreign accent of each word (1 = no foreign accent at all to 10 = a very strong foreign accent). Both parts of the experiment started with a number of practice stimuli. The sound files were presented using a PC and Praat (Boersma & Weenink 2010).

5. Results

In this section the results for the data from the DSL-speakers will be presented. The words produced by the native speakers proved to be highly intelligible and they received high comprehensibility ratings and low accentedness scores. More detailed information on the native data can be found in the appendix.
5.1 Intelligibility

All orthographic transcriptions of the stimulus words were coded as either intelligible (correct word) or unintelligible (no reply or wrong word). Data were aggregated over words but broken down by L1 of speaker (Chinese, French), segmental-prosodic condition and by listener (means of sets of 10 words). An Analysis of Variance on these aggregated data reveals significant main effects of condition ($F_{3,196} = 179.956, p < .001$) and L1 ($F_{1,198} = 116.236, p < .001$), as well as a significant interaction between condition and L1 ($F_{3,192} = 51.984, p < .001$). A post-hoc analysis (Bonferroni) shows that words without (supra)segmental errors (Condition 4) are significantly more intelligible than words containing segmental and/or suprasegmental errors, and that words with both stress and segmental errors (condition 1) are less intelligible than words with either a stress or a segmental error, but there is no significant difference between the Conditions 2 (stress error) and 3 (segmental error). Figure 1 presents the mean percentage of intelligibility per condition for the two groups of DSL-speakers separately.

The mean intelligibility of words with both stress and segmental errors is very low for the Chinese speakers (32%), while the words produced by the
French DSL-speakers remain intelligible in 75% of the cases. For the other error conditions the differences between the two groups of speakers is much smaller (or even absent).

5.2 Comprehensibility

An Analysis of Variance on the aggregated (per word) comprehensibility scores reveals significant main effects of condition ($F_{3,196} = 117.973, p < .001$) and L1 ($F_{1,198} = 28.205, p < .001$), as well as a significant interaction between condition and L1 ($F_{3,192} = 8.229, p < .001$). A post-hoc analysis (Bonferroni) shows that all conditions differ significantly from each other. The mean comprehensibility scores are presented in Figure 2.

The words uttered with a segmental error by the Chinese DSL-speakers receive a significantly lower comprehensibility judgment than words produced with a stress error, which is in contrast with the actual intelligibility scores for these conditions (cf. Figure 1). For the stimuli produced by the French DSL-speakers the judged comprehensibility is more in line with the functional intelligibility.

![Figure 2. Mean comprehensibility score per condition, broken down by L1 of speaker (error bars: 95% confidence interval).](image-url)
5.3 Accentedness

An Analysis of Variance on the aggregated (per word) accentedness scores shows significant main effects of condition ($F_{3,196} = 72.131, p < .001$) and L1 ($F_{1,198} = 42.439, p < .001$), and again a significant interaction between condition and L1 ($F_{3,192} = 4.696, p < .05$). A post-hoc analysis (Bonferroni) shows that the segmental error condition (3) does not differ from either the stress and segmental errors condition (1) or the stress error only condition (2); all other conditions do differ significantly from each other. In Figure 3 the mean accentedness scores per condition can be found, broken down by L1.

For the French data there is no difference in perceived accentedness among the error conditions, but the words without segmental and stress errors clearly sound more nativelike. The words with stress errors produced by the Chinese speakers receive lower accentedness scores than the words containing segmental errors. Furthermore, the judged accentedness of the words without errors is considerably higher than for the French data (see further below).

Figure 3. Mean accentedness score per condition, broken down by L1 of speaker (error bars: 95% confidence interval).
5.4 Combined results

Figure 4 presents the mean scores for all dependent variables for the Chinese and French speakers separately. The intelligibility scores are transposed to a 10-point-scale and the accentedness scores are converted to nativeness scores (1 = very nonnative to 10 = highly native), to facilitate comparison between the three dependent variables.

Comparison of the two panels in Figure 4 reveals that in the no-error condition (4) the results for the Chinese and French stimuli resemble each other quite closely (mean intelligibility: 9.88 versus 9.84 respectively; mean comprehensibility: 8.15 versus 8.08), but the nativeness scores do not: Chinese Condition 4 words sound less native (a mean score of 6.13) than French Condition 4 words (7.64). This indicates that the Chinese stimuli without clear (supra)segmental errors are perceived by native listeners as less nativelike than the French stimuli, as was predicted. Furthermore, the results for the French stimuli in the three error conditions are generally higher (i.e. more positive) than those for the Chinese stimuli and the differences between the error conditions are smaller for the former. Together, these findings suggest that the larger negative impact of the error conditions for the Chinese data can be explained by a higher ‘basic’ accentedness of the DSL-speech.

The correlation between the aggregated intelligibility and comprehensibility scores is relatively high ($r = .685$, $N = 200$, $p < .01$), while the correlation
between intelligibility and accentedness is clearly lower ($r = -.539$, $N = 200$, $p < .01$). The highest correlation is found between the two impressionistic scores: comprehensibility and accentedness ($r = -.739$, $N = 200$, $p < .01$). These results match those obtained by Munro & Derwing (1999).

6. Conclusion

The results of the experiment indicate that Dutch L2 words containing both a suprasegmental and a segmental error have the lowest intelligibility (a mean score of 53%), words without errors are highly intelligible (98%) and those containing either a suprasegmental or a segmental error score between those extremes (83% and 77% respectively). This suggests that a word stress error is approximately as harmful to nonnative word intelligibility as a segmental error, but a combination of the two error types is the most detrimental. The current data provide no support for the view that suprasegmental errors — in this case stress position errors — are generally more harmful to intelligibility than segmental errors, nor for the opposite. Furthermore, stress errors do not seem to stand out more than segmental errors; for the Chinese data rather the reverse seems to be true. The comprehensibility and nativeness scores are clearly higher in the stress error condition than in the conditions containing segmental errors. These results seem to be in contrast with those obtained by Van den Doel (2006). However, it may be the case that stress errors in English are more striking than in Dutch, since the reduction of unstressed syllables is generally much stronger in the former language. As formulated by Cutler & Van Donselaar (2001: 174): “In English, the correspondence between stress and vowel quality is pervasive: vowels in stressed syllables are always full, while vowels in unstressed syllables are nearly always reduced. In Dutch, the correspondence is looser. Many more Dutch words than English words contain unstressed syllables with full vowels.”

As predicted, the French DSL-speakers receive higher comprehensibility and lower accentedness ratings than the Chinese DSL-speakers. Furthermore, the words with both segmental and suprasegmental errors uttered by the Chinese speakers are considerably less intelligible than those produced by the French speakers, indicating that the harmfulness of (supra)segmental errors may depend on the ‘basic’ accentedness of the nonnative speech (cf. Van Heuven 2008).
The mutual effects of segmental and suprasegmental errors should be further investigated, using DSL-speech produced by speakers with a range of different native languages. Also, the influence of various segmental deviations (with respect to type and position in the word, cf. Bent, Bradlow and Smith 2007) on Dutch nonnative speech merits more research. And last but not least, it would be interesting to investigate whether the direction of the stress shift influences intelligibility (cf. Field 2005, Van Heuven 2008), and to use online measurements to establish the impact of (supra)segmental errors on processing time.

Notes

* The experiment was prepared and run in a research seminar under the guidance of the author; Katarzyna Horloza and Olga Kepinska are thanked for their valuable contributions to this seminar.

1. The paper is sparse on information about the construction of the stimulus materials, and therefore it is unclear how the third condition was realized. It seems rather awkward to read aloud a text without producing any pitch accents.

2. It should be noted that it was quite hard to find 20 suitable words for condition 1. This means that the type and position of segmental errors could not be controlled in any systematic way, leading to a mixture of more and less severe consonantal and vocalic errors in different positions in the stimulus words.

References


Appendix: Results for native speaker stimuli

Table 1. Absolute (and relative) intelligibility of the words uttered by the native speakers, broken down by condition (stress error versus no error).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Not intelligible</th>
<th>Intelligible</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>stress error (2)</td>
<td>8 (3.2%)</td>
<td>242 (96.8%)</td>
<td>250</td>
</tr>
<tr>
<td>no error (4)</td>
<td>10 (4.0%)</td>
<td>240 (96.0%)</td>
<td>250</td>
</tr>
<tr>
<td>Total</td>
<td>18 (3.6%)</td>
<td>482 (96.4%)</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 2 Mean comprehensibility and accentedness scores (and standard deviation) for the words uttered by the native speakers, broken down by condition (stress error versus no error), pooled over 20 words and 25 listeners. The difference between the comprehensibility scores in the two conditions is significant ($F_{1,498} = 7.938$, $p = .005$).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Comprehensibility</th>
<th>Accentedness</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>stress error (2)</td>
<td>8.17 (2.209)</td>
<td>1.97 (1.693)</td>
<td>250</td>
</tr>
<tr>
<td>no error (4)</td>
<td>8.71 (2.075)</td>
<td>1.88 (1.936)</td>
<td>250</td>
</tr>
<tr>
<td>Total</td>
<td>8.44 (2.158)</td>
<td>1.93 (1.817)</td>
<td>500</td>
</tr>
</tbody>
</table>
Author’s address

Johanneke Caspers
Leiden University Centre for Linguistics
P.N. van Eyckhof 3
Postbus 9515
2300 RA Leiden
The Netherlands

j.caspers@hum.leidenuniv.nl