On the relationship between perceived accentedness, acoustic similarity, and processing difficulty in foreign-accented speech

Marijt J. Witteman [1,2]*, Andrea Weber* [1,3]*, James M. McQueen* [1,3,4]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] International Max Planck Research School, Radboud University Nijmegen, The Netherlands
[3] Donders Institute for Brain, Cognition and Behaviour,
Radboud University Nijmegen, The Netherlands
[4] Behavioural Science Institute, Radboud University Nijmegen, The Netherlands

`Marijt.Witteman@mpi.nl, Andrea.Weber@mpi.nl, James.McQueen@mpi.nl`

Abstract

Foreign-accented speech is often perceived as more difficult to understand than native speech. What causes this potential difficulty, however, remains unknown. In the present study, we compared acoustic similarity and accent ratings of American-accented Dutch with a cross-modal priming task designed to measure online speech processing. We focused on two Dutch diphthongs: ui and ij. Though both diphthongs deviated from standard Dutch to varying degrees and perceptually varied in accent strength, native Dutch listeners recognized words containing the diphthongs easily. Thus, not all foreign-accented speech hinders comprehension, and acoustic similarity and perceived accentedness are not always predictive of processing difficulties.

Index Terms: speech comprehension, spectral similarity, rating, accentedness, foreign-accented speech

1. Introduction

Listening to native speech seems effortless, but understanding foreign-accented speech can be much more difficult (e.g., [1,2,3]). Foreign accents are common in people who learned a second language (L2) in adulthood, and are mainly caused by language-specific structures from the speaker's native language which influence the production of the L2. Foreign accents can alter the speech signal in a number of ways. For example, phoneme contrasts or allophonic variations can be lacking (e.g., *Italy* pronounced as *Eataly* by Italian speakers), subphonemic as well as suprasegmental cues can be realized differently (e.g., *deFAULT*, pronounced as *DEfault* by Hungarian speakers), and quite regularly L2 speakers replace speech sounds with those that approximate or match native categories but do not occur in the target language (e.g., *Hudson* pronounced as *Hüdson* by Dutch speakers).

While in all the above cases the speech signal deviates from the standard norm of the target language, the amount of deviation likely varies both acoustically and perceptually for different accent markers. That is, foreign-accented speech sounds can vary in how close they are qualitatively and quantitatively to categories of the target language, and at the same time different accent markers can interfere with comprehension to varying degrees. The aim of the current project is to investigate whether acoustic similarity and comprehension go hand in hand, or whether the amount of acoustic deviation is not necessarily predictive of listening performance. Specifically, we are interested in the relationship between acoustic similarity, perceived accentedness, and the degree of processing difficulty. The case will be made with

two specific Dutch diphthongs, and processing will be assessed with an online word recognition task.

Research on cross-language acoustic similarity usually compares the acoustic characteristics of speech sounds in different languages [4]. Based on these comparisons, predictions can then be made about how listeners perceive non-native sounds. Although acoustic similarity indeed often predicts how listeners categorize non-native sounds, acoustic analyses cannot always explain cross-language perceptual difficulties [5]. It is therefore unclear whether acoustic analyses in the present study will be able to fully account for processing difficulties in a comprehension task. (Note that we are not testing L2 listeners, rather we are testing L1 listeners on L2 speech, and our task is an online measure of word recognition rather than phoneme categorization.)

Native listeners are generally very sensitive to foreign accents [6]. That is, even small traces of foreignness can be detected; consequently, even highly experienced L2 speakers are usually rated as more accented than native speakers. There is, however, also evidence that perceived accentedness is not necessarily predictive of comprehension. As Munro and Derwing [6] have shown, for example, it does not take longer for native listeners to assess the truth value of foreign-accented sentences in comparison to L1 sentences. Again, it is therefore unclear whether perceived accentedness in the present study will be in line with processing difficulties

Our knowledge about how variation in speech is being processed comes mainly from research on variation in native speech. An example of this variation is medial /t/ deletion, where 'center' is pronounced as 'senner' or 'sennah' in American English [7]. When participants judged these variants in isolation, they only accepted /t/-deleted variants as words if the phonological environment in which the /t/-deletion occurred was common in production. This effect translated to new variants, but only if the /t/-deletion happened in commonly occurring places. Accepting this type of variation is thus limited to the context in which it occurs.

Sumner and Samuel [8] looked at another type of variation in speech: variation across dialects. They tested how dialectal variants are recognized and stored in the lexicon, and whether the amount of exposure to phonological variants made a difference. They compared processing of words ending in –er. In the New York City (NYC) dialect, the r is dropped (e.g., (bak[ɚ] is pronounced as bak[ə]). They found that speakers of this dialect and speakers of General American (GA) who were very familiar with the NYC dialect correctly interpreted the r-less forms as the intended form. GA-speakers unfamiliar with the NYC-accent, however, did not show priming for these dialectal forms. The authors conclude that listeners who have (passive or active) familiarity with a dialect are apparently

28 – 31 August 2011, Florence, Italy

more flexible in form processing than inexperienced listeners are: experienced listeners can deal with more variation to the standard representation when listening to dialectal speech.

Though foreign-accented speech deviates noticeably from the standard, adaptation can take place quickly. Clarke and Garrett [9], for example, found that native English listeners improved their understanding of a foreign-accented speaker within a couple of sentences. Intelligibility of the speaker also plays an important role. Bradlow and Bent [10] found that when listeners were exposed to speakers of high, medium, and low intelligibility, listeners needed more exposure to improve their understanding of the less intelligible speakers than of the more intelligible speakers.

This study was designed to shed more light on the driving factors behind processing foreign-accented speech. Specifically, we were interested to see what the roles of acoustic similarity and accentedness rating were in online speech processing.

The target language was Dutch, and the foreign accent was an American English one. Because variation in foreign accents often centers on vowels, these were taken as a starting point for item selection. In particular, we focused on the two Dutch diphthongs [œy] and [ɛɪ]. Both diphthongs are difficult for many learners of Dutch and are likely to deviate substantially from their standard pronunciations. Dutch words containing [œy] and [ɛɪ] were compared with Dutch words without any specific known markers of an American accent. These non-specifically accented words contained varying monophthongs (some of which occur in both English and Dutch), and should be relatively easy for American speakers to produce. The Dutch diphthongs [œy] and [ɛɪ], on the other hand, are typically replaced with diphthongs that resemble American [aʊ] and [aɪ] by American learners of Dutch.

In order to investigate the relationship between acoustic similarity, perceived accentedness, and processing difficulty, we first analyzed the spectral quality of the American-accented Dutch diphthongs, then had Dutch listeners rate the strength of foreign accent, and finally presented the American-accented stimuli to native Dutch listeners in a cross-modal priming study.

Forty-eight Dutch mono- or bisyllabic words were selected. Twelve of them contained the diphthong [œy] as in *duif* 'dove', 12 the diphthong [ɛɪ] as in *rijst* 'rice', and the remaining 24 words contained a variety of monophthongs, e.g., [iː] in *diep* 'deep'. The last group of words contained no sounds that are known to be specifically difficult for English learners of Dutch; it was therefore expected that these words would be perceived as the least accented.

The American speaker of the study consistently substituted Dutch [œy] with a sound that resembles American [aʊ] and Dutch [ɛɪ] with sound that resembles American [aɪ]. He was a native speaker of American English, who moved to the Netherlands less than a year ago. His Dutch proficiency was basic. The Dutch words were recorded together with the filler primes from the cross-modal priming experiment in one session. Recordings were made in pseudo-randomized order from correct Dutch spelling. All mispronunciations occurred naturally.

2.1. Acoustic analyses

Vowel durations for the 12 words containing [œy] and for the 12 words containing [ɛɪ] were measured using Praat [11]. Duration was labeled from the release of the constriction of the preceding consonant to the formation of the constriction of the following consonant, with labels being placed at zero crossings. Subsequently, F1 and F2 were measured at the 25

and 75 percent time points for each diphthong. These formant values were compared to the average vowel characteristics of native Dutch speakers of Northern Standard Dutch from Adank et al. [12].

2.2. Rating experiment

The rating experiment contained the 12 American-accented words with [œy], the 12 American-accented words with [ɛɪ], and 12 of the nonspecifically-accented words. In order to add more variation to the materials, two sets of items spoken by a Dutch native speaker and a native speaker of Italian with a very strong accent in Dutch were added to the task. In the rating study, participants heard one word at a time over closed headphones, immediately followed by a rating scale where they could indicate how accented the word was on a scale ranging from 1 to 10 (1: not accented, 10: very strong accent).

2.3. Cross-modal priming experiment

For the priming experiment, the 48 selected Dutch words served as visual targets and were each combined with two auditory primes. Primes were either the American-accented variant of the target (identical primes) or phonologically and semantically unrelated Dutch words (unrelated primes). Unrelated primes matched overall in number of phonemes with their target (e.g., prime *ketting* 'chain' and target *prikkel* 'incentive'), and the overall lexical lemma frequency of unrelated primes was not different from the frequency of targets (log frequency taken from the CELEX database; [13]).

Ninety-six filler items were added to avoid participants using strategic answering methods. Twenty-four of the fillers had a Dutch word as the visual target (18 of them contained [œy] or [ɛɪ] and were preceded by a non-word prime). The remaining 72 fillers had a non-word as their visual target (36 with a word as prime and 36 with a non-word as prime).

Two versions of the experiment were created, so that every participant saw each visual target only once. To control for effects of presentation order, each participant received a different pseudo-randomized list. The first two items of the experiment were always fillers, and there were never two critical items in a row.

Participants were seated in a sound-attenuated cabin and were informed that they would first hear a Dutch word or non-word spoken by an American speaker and then see a Dutch word or non-word on the screen. Their task was to decide as quickly and accurately as possible whether the word presented on the screen was an existing Dutch word or not by pushing one of two buttons on a button box in front of them. Yes-responses were always made with the dominant hand, and reaction times (RTs) were measured from visual target onset.

Auditory primes were presented binaurally over closed headphones at a comfortable listening level. Participants saw the visual targets on a computer screen situated about 50 cm in front of them. Visual targets were presented in white lowercase 24p Tahoma letters on a black background, 500 ms after the acoustic offset of the auditory primes. The visual targets stayed on the screen for 2000 ms, after which the next trial started. The experiment was created in Presentation [14] and used a NESU (Nijmegen Experiment Set-Up) button box.

Participants were 17 native speakers of Dutch (13 females, M age 22.11). None reported a hearing disorder, and all had normal or corrected-to-normal vision. Participants first completed the cross-modal priming task, before they participated in the rating experiment and completed a language history questionnaire.
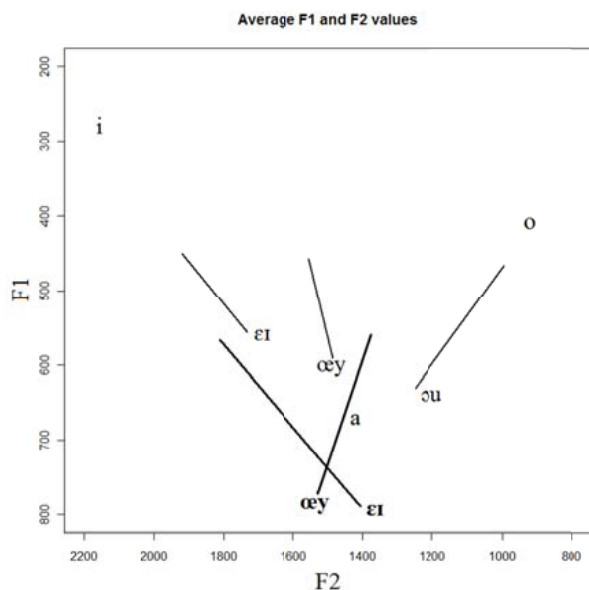
3. Results

3.1. Acoustic analyses

Figure 1: *Average F1 and F2 values for Dutch /a/, /i/, /o/, /œy/, /ɛɪ/, and /ɔu/ for male Dutch speakers in the Adank et al. corpus [12] and in bold, F1 and F2 values for /œy/ and /ɛɪ/ for the American-accented speaker in the present study.*

Figure 1 plots the average F1 and F2 values for the Dutch monophthongs /a/, /i/ and /o/, as well as for the three diphthongs /œy/, /ɛɪ/, and /ɔu/ as spoken by male Dutch native speakers (data from [12]); in bold the F1 and F2 values for the two diphthongs /œy/ and /ɛɪ/ are shown as the American speaker of the study produced them.

As one can see, the diphthongs of the American speaker started in comparison with the Dutch values at a higher F1 with an F2 value typical of Dutch /a/. Thus, the beginning of the diphthongs was lower, and in the case of [ɛɪ] more back than standard Dutch pronunciation would predict. In addition, though the [ɛɪ]-vowel produced by the American speaker has a similar trajectory as its Dutch counterpart, the trajectory of the [œy]-vowel is untypical. The F2 value of the [œy]-vowel at the 75% point is lower than at the 25% point, although standard Dutch pronunciation would predict a higher F2 end point. That is, rather than moving to a near-front vowel, the diphthong moved to a near-back vowel. In doing so, the trajectory of the [œy]-vowel resembles more the trajectory of the Dutch [ɔu].

A comparison of the F1 and F2 values in the present study with the average values for American diphthongs [15] suggests that the diphthongs by our American learner of Dutch unsurprisingly approximated the American diphthongs [aʊ] (for [œy]) and [aɪ] (for [ɛɪ]).

In sum, while both Dutch [ɛɪ] and [œy] were produced non-canonically, [ɛɪ] deviated mostly in the first target of the diphthong, but [œy] deviated in the first and second target. In addition to a more distinct deviation, American-accented [œy] productions are spectrally close to another existing Dutch category, namely [ɔu].

3.2. Rating experiment

The accentedness ratings of the 17 Dutch participants showed that the Dutch native speaker was rated as least accented and the native Italian speaker as most accented. The ratings for the American speaker were further analyzed with paired-samples t-tests. The average ratings can be found in Table 1.

Table 1: *Average ratings for [œy] words, [ɛɪ] words, and words with no specific accent marker produced by American speaker.*

Accent type	Mean	SD
[œy]	7.24	1.41
[ɛɪ]	6.24	1.18
Non-specific	5.38	1.66

The t-tests indicated that the [œy]-words were rated more accented than the [ɛɪ]-words (t(16) = 4.712, *p* < .001) and the non-specifically accented words (t(16) = 6.710, *p* < .001). In addition, the [ɛɪ]-words were rated more accented than the non-specifically accented words (t(16) = 4.378, *p* < .001). The outcome of the rating study is in line with the acoustic analyses in that the accented productions that deviate the most from the standard pronunciation are also perceived as the most accented.

3.3. Cross-modal priming experiment

Three non-specifically accented primes were discarded from the analysis because they had very high error rates (more than 10%). A further 1.2% trials on which participants made errors, as well as trials with RTs that deviated more than 2.5 SD from the condition's mean overall were discarded (together < 4% of all trials).

The remaining cross-modal priming data was analyzed with a Repeated Measure General Linear Model (GLM) using a 3 (*type of accent* - strong, medium, weak;) x 2 (*relatedness* - identical vs. unrelated) design. All factors were within-participants. We analyzed the results separately by participants (*F1*) and items (*F2*). The calculated priming effects, i.e. the difference in RTs to targets following unrelated and related primes, are shown in Figure 2.
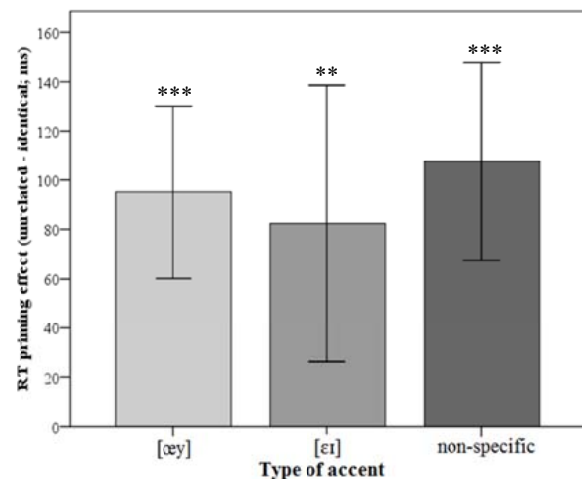

Figure 2: *Average priming effects (and SE) in ms by accent type (** p < .01; *** p < .001)*

Dutch listeners showed a significant main effect of *relatedness* (*F1* (1,16) = 83.643, *p* < .001; *F2* (1,11) = 41.122, *p* <.001) and *type* (*F1* (2,32) = 10.800, *p* < .001; *F2* (2,22) = 4.417, *p* = .024) and no interaction. That is, they responded faster to targets following identical primes compared to targets following unrelated trials, but the size of the priming effect differed between the three word types, with the effect being

the largest for words with non-specific accent markers. Priming effects for [œy] and [ɛɪ] words were comparable.

Thus, even though American-accented [œy]-words deviated spectrally the most from their standard Dutch pronunciation and were also perceived as the strongest accented, they were recognized quite easily by native Dutch listeners. In fact, there was no difference in ease of interpretation between the [œy]-words and the [ɛɪ]-words, even though the latter words were rated as more weakly accented and deviated less acoustically. Further analyses showed that Dutch listeners did not learn how to interpret American-accented words during the course of the experiment: the priming effects were already present in the first half of the experiment.

4. Discussion

The purpose of this study was to investigate some of the factors that could cause processing difficulty in the recognition of foreign-accented speech. Native speakers of Dutch performed a cross-modal priming task in which they heard three types of Dutch words spoken by an American speaker: words containing [œy], words containing [ɛɪ], and non-specifically accented words. Moreover, the degree of perceived accentedness of these words and the acoustic similarity of the American-accented diphthongs to the intended Dutch vowels were measured.

The acoustic analyses showed that the American speaker indeed produced diphthongs that clearly deviated from their standard Dutch pronunciations. The speaker's productions resembled existing American diphthongs, and deviated from standard Dutch more substantially in the case of [œy] than in the case of [ɛɪ]. The rating study confirmed this pattern of results with the accent being perceived more strongly in words with the larger acoustic deviation (i.e., words containing [œy]). Although this difference in ratings can also be influenced by other uncontrolled accent markers within the lexical items, vowel quality is very likely an important factor.

Even though the accent markers were thus clearly noticed (also in comparison to L1 Dutch), and furthermore differed in their degree of deviation, the priming study showed that Dutch listeners could interpret words with both accent markers quite well, and with no significant difference in the ease of recognition between them. Thus, the amount of deviation did not measurably affect processing. This is good news for native listeners (and indeed non-native speakers), since acoustic variation and perceived accentedness thus do not necessarily imply difficulties for online comprehension.

There are a number of possible explanations why American-accented words were recognized so easily and why there were no processing differences between variant forms. American-accented words could have been recognized easily because Dutch listeners are familiar with the tested variant forms (either from hearing American-accented Dutch or from other accents that pronounce the Dutch diphthongs similarly). The fact that facilitatory priming was found from the start of the experiment further supports this explanation. Dutch listeners apparently did not need to learn how to interpret American-accented words but could do so right from the start (for converging evidence on familiarity effects, see [3]). Alternatively, American-accented variant forms could simply not have been accented enough to interfere with processing. The accents tested in [3] were indeed perceived as even more accented than the accent tested here. Processing difficulties may thus arise only for listeners unfamiliar with the accent and/or for more extreme accents.

The lack of a difference in processing between variant forms either could reflect insufficient sensitivity of the paradigm (though differences have been found before using cross-modal priming [3]), or indeed could reflect a dissociation between acoustic similarity, perceived accentedness and processing. In the latter case, it is for example possible that the American-accented variant forms of Dutch [œy] and [ɛɪ] assimilate to the Dutch categories despite their deviations and therefore processing is not hindered. A categorization task with Dutch listeners categorizing the American-accented vowels into Dutch categories could help to clarify this point.

Further research will be needed to further tease apart the characteristics of foreign-accented speech and the subsequent consequences for its processing by native listeners. The current study shows that relying on acoustic measurements or rating tasks might not be sufficient to fully explain all aspects of speech processing.

5. Acknowledgements

6. References

[1] Adank, P. and McQueen, J. M., "The effect of an unfamiliar regional accent on spoken-word comprehension", In J. Trouvain, and W. J. Barry [Eds.], Proceedings of the 16th International Congress of Phonetic Sciences, 1925-1928, 2007.

[2] Munro, M. J. and Derwing, T. M., "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech", Language and Speech, 38(3): 289-306, 1995.

[3] Witteman, M. J., Weber, A. and McQueen, J. M., "Rapid and long-lasting adaptation to foreign-accented speech" [Abstract], Journal of the Acoustical Society of America, 12: 2486, 2010.

[4] Strange, W., Levy, E. S. and Law II, F. F., "Cross-language categorization of French and German vowels by naïve American listeners", Journal of the Acoustical Society of America, 126(3): 1461-1476, 2009.

[5] Flege, J. E., "Second language speech learning: theory, findings, and problems", in W. Strange [Ed], Speech Perception and Linguistic Experience: Issues in Cross-language Research, 233–277, 1995.

[6] Munro, M. J. and Derwing, T. M., "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners", Language Learning, 45(1): 73-97, 1995.

[7] Pitt, M. A., "How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words", Journal of Memory and Language, 61: 19-36, 2009.

[8] Sumner, M. and Samuel, A. G., "The effect of experience on the perception and representation of dialect variants", Journal of Memory and Language, 60: 487-501, 2009.

[9] Clarke, C. M. and Garrett, M. F., "Rapid adaptation to foreign-accented English", Journal of the Acoustical Society of America, 116: 3647-3658, 2004.

[10] Bradlow, A. and Bent, T., "Perceptual adaptation to non-native speech", Cognition, 106: 707-729, 2007.

[11] Boersma, P., "PRAAT, a system for doing phonetics by computer", Glot International, 5: 341-345, 2001.

[12] Adank, P., van Hout, R. and Smits, R., "An acoustic description of the vowels of Northern and Southern Standard Dutch", Journal of the Acoustical Society of America, 116(3): 1729-1738, 2004.

[13] Baayen, H., Piepenbrock, R. and Gulikers, L. "The CELEX lexical database (CD-ROM)", Philadelphia: Linguistic Data Consortium, 1995.

[14] Presentation version 13, Neurobehavioural Systems Inc.

[15] IPA Handbook "Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet", Cambridge: Cambridge Univ Press, 1999.