THE INFLUENCE OF MUSICAL ABILITIES ON THE PROCESSING OF L2 FOCUS PROSODY: AN EYE-TRACKING STUDY

Nelleke Jansen[1,2], Hanneke Loerts[1], Eleanor E. Harding[1,2], Deniz Başkent[1,2], Wander Lowie[1]

[1]University of Groningen; [2]University Medical Center Groningen
n.g.jansen@rug.nl

ABSTRACT

The perception of speech prosody in a second language (L2) remains challenging for proficient L2 users. Previous eye-tracking evidence indicates that Dutch listeners show difficulty in the processing of pitch accents signalling contrastive focus in English, whereas native English listeners use this cue in perception to anticipate upcoming information. We investigated whether musical abilities influenced the processing of contrastive focus accents by 40 Dutch adult L2 English users. In a visual-world eye-tracking paradigm, participants listened to sentences with the focus particle *only* while viewing images of the objects and characters mentioned. We measured participants' anticipatory fixations on the image showing the alternative of the contrast. Participants also completed a music perception test. Initial analyses indicate that individuals with higher music perception scores show more anticipatory fixations in L2 listening. This suggests that having stronger perceptual resources underlying both music and speech perception may benefit the processing of L2 focus prosody.

Keywords: L2, prosody, focus, musical abilities, visual world paradigm

1. INTRODUCTION

1.1. Contrastive focus accents in Dutch and English

Focus is an aspect of information structure which indicates new, contrastive, or otherwise important information in an utterance. In many languages, including Dutch and English, focus is realized by means of a pitch accent on the focal word [1, 2]. Users of English as a second language (L2) need to process this prosodic cue during sentence comprehension, which is challenging even to advanced L2 users [3, 4]. In this study, we investigated the interpretation of L2 contrastive narrow focus accents, which can play a role in the semantic interpretation of a sentence. In sentences that contain the focus particle *only*, *only* associates with a word or word group that carries focus [5]. Because focus is signalled by a pitch accent, this type of sentence needs an accent to resolve ambiguity. For example, the sentence *I only gave a spoon to the girl* has several possible readings indicated by different patterns of accentuation (focus indicated by square brackets, accent indicated by capitalisation):

(1) *I only gave [a SPOON]$_F$ to the girl*
'I gave a spoon to the girl and I didn't give anything else to the girl'

(2) *I only gave a spoon [to the GIRL]$_F$*
'I gave a spoon to the girl and I didn't give a spoon to anyone else'

While both Dutch and English use pitch accents to signal focus, and focus accents have similar phonetic realisations in Dutch and English [6], these languages differ somewhat in the way *only* is used in sentences. In Dutch, *alleen* ('only') preferably directly precedes the focal word (group) [7], which means that the word order of a sentence with *alleen* usually changes with a different focus. In English, *only* preferably precedes the verb, rather than the focal word [8, 9]. This difference is illustrated by comparing Dutch example (3) to (2).

(3) *Ik heb alleen [aan het MEISJE]$_F$ een lepel gegeven*
I have only to the GIRL a spoon given

In Dutch sentences with *alleen*, word order is thus an important focus cue besides prosody, whereas in English sentences with *only*, prosody has a higher relative importance as a focus cue.

The online processing of focus requires the integration of prosodic cues into syntactic, semantic, and discourse domains [10]. The integration of prosodic information happens fast in L1 English processing of sentences with *only* [4, 11] and L1 Dutch processing of sentences with *alleen* [12], where L1 listeners are able to use focus prosody to anticipate upcoming information [4, 12]. However, the integration of information from different linguistic (e.g., syntax) and extra-linguistic domains (e.g., discourse) is notoriously difficult for L2 users, forming an obstacle in the L2 attainment of proficient users [13]. Furthermore, the differences between languages regarding focus cues may hinder the processing of focus in a second language. For L1 Dutch users of English as an L2, difficulties could especially arise when the focus particle *only* does not directly precede the word(group) in focus, which means that listeners have to rely on accentuation alone to get to the correct interpretation. A previous eye-tracking study by Ge et al. [4] has shown that proficient L1 Dutch L2 English listeners do not show the anticipatory eye-movements demonstrated by L1 English listeners. They found that when listening to a sentence like *The dinosaur is only CARRYING the bucket, not throwing the bucket*, L1 English listeners showed anticipatory fixations on picture of the focus alternative (the dinosaur throwing the bucket) upon hearing the word *not*. However, Dutch L2 English users shifted their gaze later, upon hearing the verb *throwing*. These findings suggest that despite their high proficiency and language similarities, L1 Dutch advanced L2 English users have difficulty with the integration of prosodic cues to update their interpretation of sentences with contrastive focus in online processing.

1.2. Musical abilities and prosody perception

Music training and music perception abilities have been related to a more refined perception of prosody in foreign languages [14]. It has been proposed that such a connection may be due to neural plasticity in overlapping neural networks for music and speech [15, 16]. However, previous research has mainly considered the perception of unfamiliar foreign languages, rather than perception by proficient L2 users [14]. Following previous research supporting a link between musical abilities and speech prosody perception, we expected musical abilities to benefit the integration of prosody in determining the location of the focus when processing L2 speech. Findings from a previous ERP study [17] on the processing of L2 focus prosody suggested that refined music perception abilities were related to a reduced processing effort and better online adjustment of expectations for focus accentuation.

In the current study, we used eye-tracking to test if Dutch adults' musical abilities influenced their online focus interpretations in English sentences with *only*. We hypothesised that listeners with higher music perception scores would show more anticipatory eye-movements than listeners with lower music perception scores, reflecting more target-like processing of focus prosody in the L2.

2. METHOD

2.1. Participants

So far, we have included 40 L1 Dutch adults (ages 19-37) who were advanced L2 English users. Participants reported normal hearing and (corrected-to-)normal eyesight. We used a C-test [18] as a general measure of English proficiency. Participants had an average score of 85.90% (SD = 8.42), which indicates high proficiency levels (comparable to CEFR levels B2 to C2).

2.2. Materials and design

2.2.1. Musical abilities

We used the Short Profile of Music Perception Skills (Short-PROMS) test battery [19] to measure music perception abilities. The Short-PROMS consists of the subtests Melody, Rhythm, Rhythm-to-Melody, Tuning, Accent (i.e. intensity), Timbre, Tempo, and Pitch. In each trial, the participant listens to two musical excerpts and indicates whether they are the same or different. The PROMS gives a score for each subtest, which we used in the current analysis. Because contrastive focus is signalled by a pitch accent, we expected Pitch scores to be most strongly related to the processing of focus in speech.

2.2.2. Visual world paradigm

We used the visual world paradigm to investigate sentence processing. Participants listened to spoken sentences while looking at pictures on a screen. Auditory stimuli consisted of English sentences with the focus particle *only* in two conditions: direct object focus (DO condition) and indirect object focus (IO condition). An example of a stimulus with DO focus is given in (4) and Fig. 1.

(4) *I only gave a SPOON to the girl. [pause] I didn't give a FORK to the girl.*

Before the target sentences were presented, the speaker introduced two possible indirect objects (a girl and a boy) and two possible direct objects (spoons and forks), giving potential focus alternatives. In the first clause of the target stimulus, the focus accent indicated the location of focus, which disambiguated the meaning of the sentence. The second clause gave a lexical disambiguation, which could be anticipated based on the focus

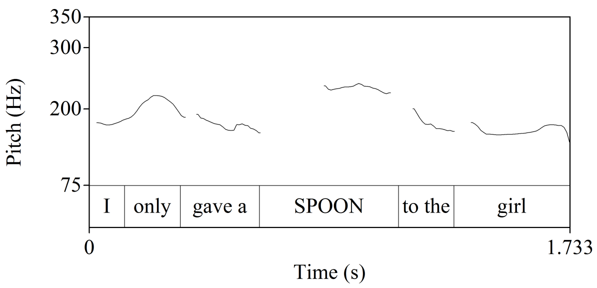

Figure 1: Pitch contour of the first clause in a DO focus trial.

accent. While listening to the introductory stories and target sentences, participants viewed pictures [20, 21] on a screen in a 2x2 grid showing the potential direct objects and indirect objects in four combinations (Fig. 2).

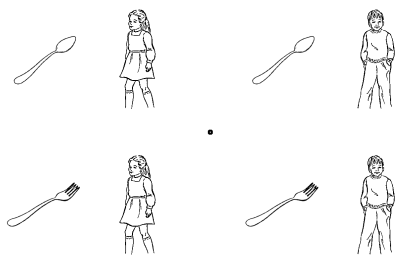

Figure 2: Example of a visual display.

The experiment consisted of 48 experimental trials and 48 fillers. The fillers contained different sentence constructions. Stimuli were recorded by a female L1 speaker of American English.

2.3. Procedure

Participants were tested in the lab. They first completed the C-test using pen and paper. Next, they took part in the eye-tracking task. Trials were presented in eight blocks. Participants listened to the stimuli using headphones. They were instructed to look at the pictures while listening to the speaker, without any active task. For 25% of the trials, participants had to answer a question about what happened in the story, to ensure they kept paying attention. Fixations were recorded with the SR Research EyeLink Portable Duo eye tracker. Finally, participants completed the Short-PROMS test.

2.4. Analysis

We measured proportions of fixation time (%FT) on four interest areas: target 1, the target given in the first clause; target 2, which is the focus alternative (given in the second clause); the competitor, which would be the focus alternative in the other focus condition; and a distractor. Taking example (4), participants were expected to look at the girl with the spoon (target 1) upon hearing the first clause. We assessed to what extent participants anticipated the alternative of the focus by fixating on the girl with the fork (target 2) before hearing the disambiguating words *FORK to the girl*. Higher %FT on target 2 compared to the competitor indicated correct interpretations of the focus accent. For visualisation, stimuli were divided into small time windows (see Fig. 3), each starting 200 ms after the onset of the relevant word, taking into account the timing of saccades in response to language [22]. To answer our research question, we measured %FT in a larger time window where anticipation may occur, i.e. *[I didn't give a]*. Length in ms differed between items.

We analysed the %FT on target 2 and the competitor with a zero-inflated beta regression (ZIBR) model, using the package glmmTMB [23] in R [24]. Beta regression is suitable for the analysis of proportion data which follows a beta distribution, and a zero-inflated model was required because %FT was often zero for both target 2 and competitor, because participants kept looking at target 1. We included a random-effects structure of `Item` nested within `Participant`. The influence of proficiency and musical abilities on anticipatory fixations was assessed by modelling interactions between the dummy variable `Competitor` (0 = target 2, 1 = competitor) and `Proficiency`, and between `Competitor` and the Short-PROMS subtest scores. We included the subtests that significantly improved the model in model comparisons, namely `Pitch` and `Tempo`. We expected higher proficiency and musical abilities would lead to higher %FT (and fewer %FTs of zero) on the target, and lower %FT (and more zeros) on the competitor.

3. RESULTS

Averaged %FT on the interest areas are given in Fig. 3. Fixations on target 2 start to deviate from the competitor before participants have heard the direct object in the second clause. This suggests that to some extent, participants anticipate the focus alternative based on the pitch accent in the first clause, while %FT on target 1 remains high.

Model outcomes are given in Table 1. The beta model, which models the %FT, shows a negative main effect of `Competitor`. This indicates %FT is higher for target 2 than for the competitor.

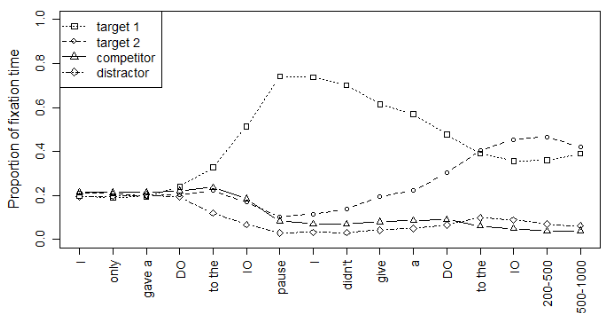

Figure 3: Averaged %FT (y-axis) on the four interest areas per time window (x-axis).

The zero-inflation model estimates the probability to get extra %FTs of zero. The main effect of `Competitor` indicates this probability is higher for the competitor than for target 2. The interaction `Competitor*Proficiency` indicates proficiency is related to fewer zeros on target 2 (negative slope on the probability of zeros), and more zeros on the competitor (positive slope). Similarly, the interaction `Competitor*Pitch` shows musical pitch perception is related to fewer zeros on target 2 and more on the competitor, in line with our expectations. The interaction `Competitor*Tempo` shows an effect of musical tempo perception in the opposite direction. However, while the effect of `Tempo` differs between target and competitor, the positive slope for target 2 is not significant, and `Tempo` does not affect the probability of zeros on the competitor (flat slope).

Parameter	Beta model Estimate (SE)	Zero-infl. model Estimate (SE)
Intercept	0.320 (0.087)*	1.189 (0.158)*
Competitor	-0.194 (0.094)*	0.838 (0.089)*
Proficiency	0.016 (0.010)	-0.027 (0.019)
Pitch	0.077 (0.055)	-0.351 (0.111)*
Tempo	-0.056 (0.076)	0.228 (0.136)
Competitor:Proficiency	-0.016 (0.010)	0.042 (0.010)*
Competitor:Pitch	-0.104 (0.065)	0.486 (0.062)*
Competitor:Tempo	0.030 (0.086)	-0.232 (0.082)*

* significant at α-level of 0.05

Table 1: ZIBR model (coefficients are log-odds).

4. DISCUSSION

In this eye-tracking study, we investigated the influence of musical abilities on the processing of contrastive focus prosody in sentences with *only* by L1 Dutch proficient L2 English users. We examined to what extent participants showed anticipatory fixations on the alternative of the contrast (i.e., the second target), versus the competitor (reflecting a different contrast). Anticipatory fixations indicated the participant had come to the correct focus interpretation by processing the contrastive pitch accent in the first clause of the stimulus.

Results showed that participants (to some degree) anticipated the correct focus alternative, as fixations proportions were higher on the second target than on the competitor. However, participants also continued looking at the target of the first clause, in line with previous findings [4]. Continued fixations on the first target may partly reflect participants' attention towards *what happened*, rather than the alternative situation (*what did not happen*).

Based on previous research indicating that individuals with stronger music perception abilities showed a more fine-grained prosody perception [14], we hypothesised that these individuals would show more anticipation based on the contrastive pitch accent. Although results showed no effect on the fixation time proportions, musical pitch perception ability increased the probability to fixate on the target (i.e., lower probability of no fixation on the target) and decreased the probability to fixate on the competitor. This increased anticipation indicates that musical abilities facilitate the integration of prosodic information in language processing, which may be challenging for L2 users [3, 13]. Our study thus extends previous findings on the perceptual advantage related to musical abilities in foreign language perception to an advantage in prosody-to-meaning mapping in proficient L2 perception.

A beneficial influence of musical abilities on the processing of L2 prosody may be explained on the basis of transfer between music and speech in overlapping neural networks [15, 16]. Having strong perceptual resources underlying both music and speech perception may thus also benefit the processing of L2 focus prosody. Such strong perceptual resources could be the result of genetic predispositions, music training, or both [25]. Because we currently only used music perception scores as our measure of musical abilities, further analyses are planned to elucidate the potential role of participants' music training.

English proficiency also had a positive effect on anticipation. This indicates participants with higher (written) proficiency show more efficient L2 prosodic processing, perhaps because they can use processing resources that would otherwise be dedicated to lexical or syntactic processing.

Finally, this study will be extended to include a control group of L1 English participants. This will allow us to compare performance of the L2 listeners to L1 listeners, and investigate whether musical abilities also play a role in L1 processing.

5. ACKNOWLEDGEMENTS

We thank Victor Bernal from the Data Science team of the Center for Information Technology at the University of Groningen for advice on the statistical analysis.

6. REFERENCES

[1] C. Gussenhoven, *On the Grammar and Semantics of Sentence Accents*. Foris Publications, 1984.
[2] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 1996.
[3] E. Akker and A. Cutler, "Prosodic cues to semantic structure in native and nonnative listening," *Biling.: Lang. Cogn.*, vol. 6, no. 2, pp. 81–96, 2003.
[4] H. Ge, I. Mulders, X. Kang, A. Chen, and V. Yip, "Processing focus in native and non-native speakers of English: An eye-tracking study in the visual world paradigm," *Appl. Psycholinguist.*, vol. 42, no. 4, pp. 1057–1088, 2021.
[5] M. Rooth, "A theory of focus interpretation," *Nat. Lang. Semant.*, vol. 1, no. 1, pp. 75–116, 1992.
[6] A. M. Sluijter and V. J. Van Heuven, "Acoustic correlates of linguistic stress and accent in Dutch and American english," in *Proc. 4th International Conference on Spoken Language Processing*. IEEE, 1996, pp. 630–633.
[7] A. Foolen, R. van Gerrevink, L. Hogeweg, and P. Prawiro-Atmodjo, "The placement of focus particles in dutch," *Linguistics in the Netherlands*, vol. 26, no. 1, pp. 51–63, 2009.
[8] G. Bouma, P. Hendriks, and J. Hoeksema, "Focus particles inside prepositional phrases: A comparison of Dutch, English, and German," *J. Comp. Ger. Linguist.*, vol. 10, no. 1, pp. 1–24, 2007.
[9] M. M. Vann, "Is it only about only? A study on focus particle placement acceptability in L2 English learners and English monolinguals," *Annali di Ca' Foscari. Serie occidentale*, vol. 56, 2022.
[10] K. Lambrecht, *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, 1994.
[11] S. Gennari, L. Meroni, and S. Crain, "Rapid relief of stress in dealing with ambiguity," in *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*, J. C. Trueswell, M. K. Tanenhaus, and M. K. Tanenhaus, Eds. MIT Press, 2005, pp. 245–259.
[12] I. Mulders and K. Szendrői, "Early association of prosodic focus with alleen 'only': Evidence from eye movements in the visual-world paradigm," *Front. Psychol.*, vol. 7, p. 150, 2016.
[13] A. Sorace, "Pinning down the concept of "interface" in bilingualism," *Linguist. Approaches Biling.*, vol. 1, no. 1, pp. 1–33, 2011.
[14] N. Jansen, E. Harding, H. Loerts, D. Başkent, and W. Lowie, "The relation between musical ability and sentence-level intonation perception: A meta-analysis comparing L1 and non-native listening," in *Proc. Speech Prosody*, 2022, pp. 713–717.
[15] A. D. Patel, "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," *Front. Psychol.*, vol. 2, p. 142, 2011.
[16] A. Tierney and N. Kraus, "Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH)," *Front. Hum. Neurosci.*, vol. 8, p. 949, 2014.
[17] N. Jansen, R. Ganga, M. Struiksma, and A. Chen, "Musicality influences the processing of accentuation in sentences with *only* by L2 English learners," in *Hanyang International Symposium on Phonetics and Cognitive Sciences of Language*, 2019, pp. 53–54.
[18] M. Daller, A. Müller, and Y. Wang-Taylor, "The C-test as predictor of the academic success of international students," *Int. J. Biling. Educ. Biling.*, vol. 24, no. 10, pp. 1502–1511, 2021.
[19] M. Zentner and H. Strauss, "Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS," *Ann. N. Y. Acad. Sci.*, vol. 1400, no. 1, pp. 33–45, 2017.
[20] E. Bates, S. D'Amico, T. Jacobsen, A. Székely, E. Andonova, A. Devescovi, D. Herron, C. Ching Lu, T. Pechmann, C. Pléh *et al.*, "Timed picture naming in seven languages," *Psychon. Bull. Rev.*, vol. 10, no. 2, pp. 344–380, 2003.
[21] J. G. Snodgrass and M. Vanderwart, "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity." *J. Exp. Psychol.: Learn. Mem. Cogn.*, vol. 6, no. 2, p. 174, 1980.
[22] G. T. Altmann and Y. Kamide, "Now you see it, now you don't: Mediating the mapping between language and the visual world," in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, J. Henderson and F. Ferreira, Eds. Psychology Press, 2004, pp. 347–386.
[23] M. E. Brooks, K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker, "glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling," *The R Journal*, vol. 9, no. 2, pp. 378–400, 2017.
[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/
[25] E. G. Schellenberg, "Music training and speech perception: A gene–environment interaction," *Ann. N. Y. Acad. Sci.*, vol. 1337, no. 1, pp. 170–177, 2015.