Computers & Education 114 (2017) 206–221

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Training predictive L2 processing with a digital game: Prototype promotes acquisition of anticipatory use of tone-suffix associations

Andrea Schremm*, Anna Hed, Merle Horne, Mikael Roll

Department of Linguistics and Phonetics, Lund University, Box 201, 22100 Lund, Sweden

A R T I C L E I N F O

Article history:
Received 2 December 2016
Received in revised form 12 July 2017
Accepted 13 July 2017
Available online 14 July 2017

Keywords:
Adult learning
Applications in subject areas
Teaching/learning strategies
Evaluation of CAL systems

A B S T R A C T

The present article introduces the concept of an educational game application aimed at providing training in predictive second language (L2) processing. The prototype of the game, focusing on Swedish tone-suffix associations, was tested during a two-week-period, with L2 learners whose native language lacked the targeted anticipatory linguistic cue. Results indicated that the game successfully promoted the learning of a novel L2 predictive strategy, as reflected in a general increase in accuracy throughout the test period and a gradually faster performance of the predictive task. More time spent on the highest level of the game was associated with greater accuracy gains. Furthermore, results suggest that perceptual training provided by the prototype even leads to improved production of the tonal cue. Implementation of the presented game concept in the form of a platform game is also discussed.

1. Introduction

Active and continuous predictive processing have been suggested to constitute a core function of the human brain (Bar, 2009; Clark, 2013), which has been linked to learning in general and language acquisition in particular (e.g. Dell & Chang, 2014; Phillips & Ehrenhofer, 2015). There is indeed considerable empirical evidence that native speakers (NSs) are able to make use of a wide range of cues to generate expectations about upcoming information at several different levels of representation during language processing (e.g. Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; Dikker, Rabagliati, Farmer, & Pylkkänen, 2010; Kamide, Altmann, & Haywood, 2003; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004; see Kuperberg & Jaeger, 2016 for a review), and even second language (L2) learners in general seem to be capable of making linguistic predictions while processing their L2 (Kaan, 2014). Intuitively, predictive processing skills could benefit L2 learners in a variety of ways, being powerful mechanisms assumed to aid in the comprehension of noisy and ambiguous input (Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007), rapid speech processing (Lau, Holcomb, & Kuperberg, 2013), as well as potentially facilitating L2 acquisition itself due to their proposed essential role in learning. Nevertheless, learners appear to be more restricted in their anticipatory processing in the L2 relative to NSs, especially if they are at lower proficiency levels, the predictive cue is absent in the native language (L1), or the predictive

Abbreviations: L1, native language; L2, second language; NS, native speaker; RT, response time.

* Corresponding author. Department of Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Box 201, 22100 Lund, Sweden.
E-mail address: andrea.schremm@ling.lu.se (A. Schremm).

http://dx.doi.org/10.1016/j.compedu.2017.07.006

information is more complex (Dussias, Valdes Kroff, Guzzardo Tamargo, & Gerfen, 2013; Mitsugi & MacWhinney, 2015). Focused training in specific L2 predictive processing skills might therefore be advantageous for a wide range of L2 learners and it could contribute to more native-like processing of the L2. The present paper introduces a game application currently under development, which aims at providing such training for L2 learners. It also reports on testing the prototype of the game.

The game mechanics have been designed to facilitate learning of specific predictive language processing strategies, by requiring the player to form expectations about upcoming speech input based on a cue and providing immediate feedback on the accuracy of this prediction in the form of actually experienced input. The key features of the game mechanics were developed by drawing on results from previous research on NS and L2 predictive language processing. In its current implementation, the game focuses on Swedish tone patterns realized on word stems, which constitute predictive cues due to their associations with specific immediately following grammatical suffixes. The game trains L2 learners to use such stem tone patterns to anticipate upcoming suffixes during online speech comprehension, i.e. to apply a predictive strategy similar to that observed in NSs of Swedish (Roll, 2015; Roll, Horne, & Lindgren, 2010; Roll, Söderström, & Horne, 2013; Roll et al., 2015; Söderström, Horne, & Roll, 2016; Söderström, Roll, & Horne, 2012).

In an experiment, we investigated if playing a prototype implementing the discussed game mechanics would promote acquisition of the trained L2 predictive strategy in low-proficient learners whose native language lacked the specific language feature (tone) with predictive value. Therefore, the investigation aimed at determining whether L2 learners showed improvements in carrying out the task constituting the essence of the game after a two-week-period of play, as reflected in their speed and accuracy of performance. Furthermore, the study tested whether such perceptual training would lead to gains even in L2 speech production, manifested in more native-like production of the language feature constituting the predictive cue. The experiment and its results are presented below. A more complete version of the tested prototype, implementing the game mechanics in the context of a platform game, will also be introduced.

1.1. Predictive processing

In cognitive neuroscience, there is an increasingly growing emphasis on predictive processing as an essential computational mechanism of the brain, shaping and supporting the perceptual system and even underlying action (Bubic, von Cramon & Schubotz, 2010; Clark, 2013). Within this framework, sensory processing is often seen as being realized in a hierarchically organized system, with higher, more abstract levels constantly generating predictions about the probability of activities at lower levels, based on hypotheses drawn from previous knowledge as well as the current model of the context. Subsequent matching of the prediction to the actually experienced input may result in the detection of discrepancies, constituting a prediction error. Information from such an error signal can then be used at higher levels to adjust the functioning of the system in order to minimize the discrepancy with subsequent predictions (e.g. Clark, 2013; Friston, 2005). The result is the modification of processing mechanisms and/or behavior and as such, this prediction-feedback cycle has been suggested to constitute an essential learning mechanism (e.g. Schultz & Dickinson, 2000). Ongoing predictive functioning of the brain would provide the further cognitive benefit of freeing up resources from processing what is predictable, which would enable the individual to concentrate on the unexpected, discovering new things to learn. Also, in cases where prediction of stimuli facilitates task completion, this mechanism could guide directed allocation of attentional resources (Bar, 2009).

Along these lines, several models and accounts of language processing involve prediction or expectation as a core mechanism (e.g. Chang, Dell, & Bock, 2006; Levy, 2008; Pickering & Garrod, 2007), and there is now compelling evidence that NSs anticipate, at least to some degree, upcoming information prior to encountering the actual input (Kuperberg & Jaeger, 2016). For instance, contextual visual information in combination with the semantics of verbs and nouns has been found to facilitate anticipatory selection of likely upcoming arguments (e.g. Altmann & Kamide, 1999; Kamide et al., 2003). Comprehenders were also reported to anticipate certain specific properties of expected input such as gender class of upcoming nouns, based on inflections on preceding adjectives (Van Berkum et al., 2005) or gender-marked articles (Wicha et al., 2004). Speech comprehension might even involve the generation of more concrete sensory predictions concerning, for instance, the phonological form of expected nouns (DeLong et al., 2005) and orthographic forms cued by syntactic structure of the preceding context (Dikker et al., 2010). NSs have also been observed to use prosodic cues to predict syntactic structure (Roll & Horne, 2011), upcoming argument (Weber, Grice, & Crocker, 2006), to generate pragmatic inferences (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014) and to anticipate upcoming inflections (Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström et al., 2012, 2016).

In addition to the nature of the predictive cue and anticipated information, available time also appears to be an important factor modulating anticipatory behavior in NSs (Wlotko & Federmeier, 2015). Temporal constraints seem to affect the degree to which more complex cues can be used for generating predictions (Phillips & Ehrenhofer, 2015), as well as the extent to which information even at lower levels of representation is pre-activated (cf. Kuperberg & Jaeger, 2016).

1.2. Predictive processing in an L2

Applying predictive strategies during the comprehension of an L2 could be thought to provide various benefits for the learner. For instance, generating expectations about upcoming information and linguistic structures based on what has just been processed, and then subsequently evaluating the correctness of the prediction by matching it against the incoming input provides a way of testing one's knowledge about the L2 and learning about various dependencies in the language (Phillips &

Ehrenhofer, 2015). The way predictive processing can have a scaffolding function during the comprehension of noisy or otherwise impoverished input might be especially advantageous for learners, who have greater difficulties with processing speech in noise in their L2 than in their L1 (Shi, 2010).

Non-native anticipatory processing has been extensively studied in languages such as Spanish by comparing L2 learners' and NSs' use of morphologically marked gender-agreement on determiners as anticipatory cues to the following noun that the determiner agrees with. For instance, in a looking-while-listening test, L2 learners of Spanish were found to rely on the predictive value of gender information in the article in a native-like way only at a high proficiency level. Learners with low proficiency displayed no or only limited anticipatory behavior depending on the absence vs. presence of the corresponding feature in the L1 (Dussias et al., 2013). There is, however, some indication that even highly proficient learners use gender information less consistently than NSs. Grüter, Lew-Williams, and Fernald (2012) observed that gender cues had a reduced facilitating effect on L2 learners' processing of familiar nouns relative to the NS controls. This has been argued to be related to relatively weak associations between lexical items and gender markers, as a result of limited reliance on distributional information in typical adult L2 acquisition. Similarly, Hopp (2013) found that the degree to which advanced L2 learners of German evidenced native-like gender information processing was related to their mastery of the gender system of the language as a whole, and was presumably dependent on the presence of well-established links between nouns and their gender classes supporting online comprehension. Interestingly, the learners in Grüter et al.'s (2012) study displayed native-like predictive processing with novel words acquired during teaching trials where learning conditions simulated characteristics of native language acquisition.

Studies on syntactic processing also highlight the fact that learners might be limited in their ability to make use of complex predictive cues, even if, in principle, they seem to be capable of acquiring strategies for anticipating upcoming syntactic structures which do not have any correspondence in their L1 (Lee, Lu, & Garnsey, 2013). Thus, Mitsugi and MacWhinney (2015) reported that L2 learners of Japanese did not appear to make use of their attested knowledge of case markers in order to anticipate upcoming arguments in a native-like way. The high complexity of the predictive cue, requiring the integration of lexical and grammatical information from more than one constituent, and the short time-window for generating the prediction, the span of a single word, were suggested to have contributed to the lack of predictive processing.

1.3. Implementing training in L2 predictive processing

As discussed above, empirical evidence suggests that predictive language processing as such is available to L2 learners. At the same time, learners might approximate the degree to which NSs use anticipatory cues only at a higher proficiency level, and even then they might be more limited in their ability to exploit predictive features of the L2. In certain cases, as the findings of Grüter et al. (2012) indicate, adult learners might not even be able to make use of the full potential of cues to facilitate language comprehension without learning the relevant co-variations under specially designed conditions. One reason might be the absence of strong enough associations between cues and the associated upcoming information, the target, in the mental representations of learners (Grüter et al., 2012; Hopp, 2013). Therefore, training should ideally involve fostering the underlying links between the predictive cue and its target. Learning of such co-variations has often been associated with implicit learning mechanisms, which are assumed to underlie people's apparent ability to unconsciously pick up statistical regularities present in the input they are exposed to (e.g. Perruchet, 2008; Reber, 1967). Simple passive exposure to cue-target dependencies might, however, not be sufficient as more recent results on domain-general implicit learning mechanisms suggest that attention to the input material might be necessary for acquisition (Perruchet & Pacton, 2006), and covariation-learning mechanisms might only be engaged once the predictive cue has been attended to (Hoffmann & Sebald, 2005). Task-relevance of the predictive cue could effectively direct attention to it, indicating that carrying out language tasks for which co-occurrence relations between cue and target are crucial could be thought to promote the emergence of associations underling anticipatory processing. Such a task might be one that explicitly requires the learner to perform the actual predictions based on the cue. In addition, as incomplete overall knowledge of the relevant dependencies might also interfere with effective predictive processing (e.g. Hopp, 2013), exposure to varied, carefully designed material is also essential.

Furthermore, as previous research has indicated, the lack of corresponding features with predictive significance in the L1 seems to modulate the success or speed with which anticipatory mechanisms are acquired (Dussias et al., 2013). This is not surprising considering the fact that learners in these cases need to acquire a new online language processing strategy. In training, the emergence of such a new strategy could be thought to be promoted by repeated practice of the skills involved, such as recognition of the predictive cue and generation of the associated expectations.

Learners might also not have the necessary processing speed or automaticity for integrating a range of information constituting more complex predictive cues. Available time for generating useful predictions has been seen to be an important variable even in native language processing (Phillips & Ehrenhofer, 2015; Wlotko & Federmeier, 2015), and considering the presumably relatively slower processing speed of inexperienced L2 learners (e.g. McDonald, 2000; McElree, Jia, & Litvak, 2000), it would not be surprising if temporal constraints had an even more profound effect. As gradual speed-up and eventual automatization of L2 skills have been argued to be expected as a result of a great deal of practice (DeKeyser & Criado, 2012), effective training should preferably involve extensive performance of the skills underlying the predictive mechanism. One possible strategy that could be thought to be consistent with promoting the rapid generation of expectations is the introduction of time constraints on the predictive task. Making speed a central aspect of the task and rewarding rapid responses might be assumed to motivate learner efforts to make faster responses.

1.4. Digital games and training in L2 predictive processing

Digital games can be argued to provide an especially suitable medium for implementing training in predictive processing. To begin with, video gaming is pervasive in society, even among adults (Entertainment Software Association (ESA), 2016), suggesting that digital games have the potential to appeal to a large learner group. Also, the inherent ability of well-designed games to sustain motivation and engagement (Plass, Homer, & Kinzer, 2015; Whitton, 2011) could be crucial considering the large number of repetitions presumably needed to establish strong enough cue-target associations as well as to speed up and eventually automatize anticipatory processing. Well-designed games have been argued to have characteristics that contribute to maintaining motivation and engagement in an educational context, including the ways the goal-structure, feedback and challenges in the game are constructed. First, the goals within the game should be clear and appropriately organized so that they allow for gradual progression (Garris, Ahlers, & Driskell, 2002). Immediate feedback is also important (e.g. Kiili, 2005), in a form that is relevant to achieving the goals of the game. Well-designed games also maintain an optimal level of challenge throughout the gameplay (Malone & Lepper, 1987). Challenges that are dynamically adopted to the player's current skill level in a way that they are perceived neither too simple nor impossible to achieve have been associated with the experience of “flow”, resulting in an intrinsically rewarding state of optimal performance during complete immersion in an activity (Csikszentmihalyi, 1990; Hamari et al., 2016). Furthermore, games constitute their own virtual world, allowing the user to take risks, experiment and potentially fail without having to face real-world consequences (Plass et al., 2015). This creates a learning environment that is very much different from the one experienced in traditional classroom settings and might be especially important for learners to gain confidence in their L2 skills.

In what follows, the outline of the game mechanics will be introduced. The game has been constructed so that it promotes the emergence of predictive processing, the core of the educational tool presented in this paper. In general terms, the game is made up of cycles of stimulus presentation, followed by response from the player and subsequent automatic feedback. First, the player is presented with a sentence context leading up to a predictive cue. In the current form of the game, the sentence context is provided auditorily, while the corresponding text is also shown to the player. If the cue is auditory in nature, such as a segmental or prosodic feature, the audio recording of the sentence is played up until after the cue, the disambiguation point. The sentence context is presented in written form as soon as the audio starts. For instance, the alternatives could be represented by different platforms, i.e. suspended surfaces on which the targets are shown as text, and the player would indicate his or her choice by selecting one of them. In the present form of the game, the choice is always binary and only one of the continuations results in an utterance that would be judged by NSs as well-formed including features involved in the cue-target relation. The player is instructed to make a choice as quickly as possible once the disambiguation point has been reached, i.e. the player has arrived at the point where the next step would be to move on to one of the possible continuations, for instance by clicking on one of the alternative choices. Decision latency is calculated from the disambiguation point up until the player's selection of one of the displayed alternatives. As soon as the selection of a continuation has been made, feedback is given on the correctness of the choice. Feedback is partly visual, showing the correct complete sentence, with the visual representation form depending on whether the choice was correct or not; and partly auditory as the recording of the correct continuation is played up. After selecting a continuation and receiving feedback, the player can choose to replay the complete and correct auditory recording of the last item. Players then receive scores, which are calculated by taking into account, not only the accuracy of the response, but also the speed of the decision.

1.5. The Swedish word accent-suffix association

The game mechanics are implemented in an educational software currently under development, which promotes the acquisition of predictive processing of Swedish word stem tones. In Swedish, each prosodic word is pronounced with one of the two word accents of the language, either accent 1 or accent 2. As the tonal pattern (accent 1 or accent 2) associated with the same stem may vary depending on the grammatical suffix attached to it (Bruce, 1977; Riad, 2014), Swedish word accents can function as cues to the upcoming suffix. For example, the word stem *bil* ‘car’ is pronounced with a low tone (accent 1) if it is followed by the singular definite article –*en* attached to the stem as in $bil_{accent1}$-*en* ‘car-sg’. The same stem, however, receives a high tone (accent 2) if the word ends in the plural suffix –*ar*, i.e. $bil_{accent2}$-*ar* ‘car-pl’. NSs have been shown to rely on stem tones to anticipate how words end (Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström et al., 2012, 2016) and neuroimaging data suggests that the recognition of the stem tone is immediately followed by the pre-activation of associated suffixes (Roll et al., 2015). The fact that previous research has consistently demonstrated native speakers' tendency to anticipate upcoming suffixes based on the tone might be taken to suggest that such a strategy effectively supports speech processing. Tones are important features of the Swedish language that increase the predictability of upcoming speech material, thereby playing a role in reducing processing effort and freeing up resources to focus on other aspects of language and communication. Indeed, in the context of non-compound nouns and verbs, word accents on stems are highly predictive, greatly restricting the possible set of upcoming suffixes. For instance, accent 1 on a word stem such as *bil* ‘car’ excludes the possibility of any continuation other than the singular definite suffix -*en*. Also, the word accent on the verb stem distinguishes between upcoming simple present and past tense suffixes in the conjugational class of verbs that take regular -*er*/-*te* endings ($rök_{accent1}$-*er* ‘smoke-pres.’/$rök_{accent2}$-*te* ‘smoke-past’). Results obtained with stimulus material designed to test listeners' ability to predict an upcoming ending relying solely on the word stem tone support this assumption. In a study where the suffix on pseudoword stems was completely masked by a cough, and the only cue to the singular versus plural ending was the stem tone (e.g. *tvuk*$_{accent1}$*-COUGH*), Swedish native speakers were still able to recover the correct ending (*tvuk*$_{1}$*-en*) with above-chance accuracy, 87.8% for accent 1 and 72% for accent 2 (Söderström et al., 2016).

L2 learners of Swedish often make mistakes in their use of word accents (Bannert, 2004), suggesting that it is an especially difficult feature of the L2 to acquire. Considering the fact that every prosodic word is pronounced with either accent 1 or accent 2, such difficulties can significantly interfere with communication with Swedish native speakers. On the one hand, if L2 learners fail to use word accents, native speakers cannot rely on the same pervasive cues facilitating speech comprehension as they are accustomed to. On the other hand, L2 learner speech might be characterized by frequent incorrect tone-suffix combinations, leading native speakers to generate tone-based predictions that frequently turn out to be wrong, hampering efficient speech processing and, in turn, mutual comprehension. Indeed, when suffixes were preceded by invalid tones in experimental stimuli, native speakers displayed not only significantly slower response times for deciding on suffix meaning, but also a P600 in their electrophysiological brain response, reflecting increased efforts to reprocess the perceived word forms (Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström et al., 2012, 2016). Native speakers also rated such words with incorrect tone-suffix combinations significantly less acceptable than words in which endings were preceded by their correctly associated tones (Roll et al., 2010). Nevertheless, despite the importance of Swedish word accents for communication in the L2, the association between tones and suffixes is a feature that is not commonly taught or practiced in Swedish L2 courses (Schremm, Söderström, Horne, & Roll, 2016).

Training focusing on Swedish tone-suffix connections provides an interesting testing ground for the proposed game mechanics. First, word accents are a pervasive property of the language, constituting therefore a great potential for facilitating speech comprehension in L2 learners. At the same time, the predictive connection between the word accent and the upcoming suffix is a feature that is largely unique to Swedish, and related Norwegian. Also, the predictive cue itself, the word accent, is a prosodic property that is absent from the L1 of most L2 learners of Swedish. As a result, learning effects due to training would indicate the acquisition of a novel L2 predictive strategy, instead of simply the transfer of L1 speech processing mechanisms. Furthermore, previous results suggest that it is in principle possible for L2 learners to acquire tone-suffix associations, but apparently only at a higher proficiency level and after extensive exposure to the L2: whereas beginner learners of Swedish showed no indication of predictive use of word accents (Gosselke Berthelsen, Horne, Shtyrov, Brännström, & Roll, submitted), learners at an upper-intermediate level displayed facilitated processing of suffixes correctly cued by the preceding stem tone (Schremm et al., 2016). Nevertheless, these more proficient learners apparently made use of the anticipatory value of word accents to a more limited extent than NSs, as reflected in a relatively smaller disruption in suffix processing caused by incorrect tonal cues in this group. Possibly, learners still had not developed associations between the mental representation of cues and targets which were as strong as those of NSs, something which could have led to weaker pre-activations of predicted material, and therefore less inhibition of alternative suffix continuations. With increasing L2 exposure, however, word accent processing in the L2 group showed signs of becoming more native-like. These results thus point to the value of providing training in Swedish tone-suffix associations, especially for L2 learners at a lower proficiency level. Finally, since perceptual training of segmental and suprasegmental features of an L2 has previously been seen to lead to improvements in the production of the relevant features (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Huensch & Tremblay, 2015; Wang, Jongman, & Sereno, 2003), playing the game application might prove to be a way to promote the emergence of more native-like productive use of word accents by L2 learners of Swedish.

In addition, L2 research has shown that certain language features such as the English present tense third person singular *-s* inflection on the verb are notoriously difficult to acquire despite extensive exposure to relevant L2 input (R. Ellis, 2006). It has been argued that factors such as complex form-meaning mappings as well as the redundancy of these features in interpreting meaning might contribute to this difficulty (DeKeyser, 2005; N. C.; Ellis, 2006). In several respects, Swedish word accents are characterized by similar properties: the same tone is associated with a variety of upcoming inflections with different morphological functions; thus, there is no straightforward one-to-one connection between a specific tonal cue and its grammatical target. Also, it is possible to extract the meaning of suffixes independently of the preceding word accent. Therefore, successful acquisition of Swedish tone-suffix connections with the help of the proposed game mechanics would have important implications for a training strategy that might be effective for teaching other anticipatory relations across different L2s that have proved to be especially challenging. For instance, one might consider training the English present tense *-s* inflection as a feature reflecting a predictive association between third person singular subjects and the verb ending.

2. Current study

For the present study, a prototype of the game was constructed in order to test the proposed game mechanics. For that reason, the prototype implemented the basic game mechanics but lacked any visual context and animations characteristic of digital games. Training focused on a selected subset of Swedish word accent-suffix connections, involving only stem tone variations related to the appearance of singular versus plural endings on nouns as well as present tense versus past tense endings on verbs. Participants played the prototype 15–60 min each day, for ten days during a period of two weeks. The game involved responding to sentence items such as *Kungen bygger aldrig* 'The king never builds'. The task was to listen to the first

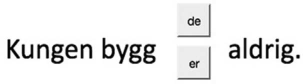

Fig. 1. Example sentence item as presented in the game prototype. Two buttons represented the alternative suffix continuations.

part of the sentence *Kungen bygg-* 'The king never build-' and then choose one of two suffix alternatives (present tense suffix *-er* or past tense suffix *-de*) as quickly as possible, based on how the player thinks the sentence will continue (see Fig. 1). For instance, if the verb stem *bygg-* 'build' was pronounced with accent 1, the correct continuation would be *−er,* due to the association between accent 1 and the present tense suffix *-er.* Alternatively, if accent 2 had appeared on the verb stem, the correct choice would be the past tense suffix *−de*, the inflection associated with accent 2. The time it took for participants to decide on an ending after the presentation of the word stem tone was measured, and the accuracy of each choice as well as the amount of time participants spent playing the game were logged. The material to be presented was organized into three successive main levels. The first two main levels were further broken down into three sublevels each, which were assumed to consist of language material of increasing difficulty (see section 3.3 for more details). Main level 3 contained a mixture of material from all previous levels. Directly before and after the gaming period, participants were recorded while reading a set of Swedish sentences. This was done in order to be able to investigate if perception training would result in more native-like production of word accents.

As a result of training with the game, the hypothesis was that participants would begin to acquire the ability to use word accents in order to predict the upcoming suffix. Several sub-hypotheses can be derived from this main hypothesis (main hypothesis 1). First, learning effects would be seen in a general increase of accuracy by the end of the gaming period, relative to an initial accuracy close to chance level (hypothesis 1.1). As participants progressed through several levels of complexity throughout the game, accuracy was not expected to display a steadily increasing tendency, since at the beginning of each new level, participants had to get experience in applying the skill under training in a somewhat different context. Nevertheless, since the same general predictive strategy was required to perform the task throughout the whole duration of the game, improvement in predictive processing was expected to show up in increased final versus initial accuracy. Second, extensive training was also expected to result in a more rapid performance of the suffix prediction, which would be reflected in a general reduction of response times at later stages of the game compared to the beginning (hypothesis 1.2). From the point of view of the acquisition of tone-suffix associations, the relevant response time measure is the speed with which learners make correct suffix choices, which is expected to generally decrease from the initial to the final levels of the game, even if, as it has been argued for the accuracy measurement, a completely monotonous decrease is not expected. Observing reduced final versus initial response times for correct responses would be consistent with the assumption that using the game leads to a speeding up in performing the predictive L2 skill, not just to a general performance improvement in mechanistically making a button choice. Third, if training with the game promotes acquisition of tone-suffix associations, more time spent with the game might be expected to result in higher accuracy and faster response times in the final measurements relative to participants' initial performance (hypothesis 1.3).

There are also a number of game-external factors that can be thought to influence the outcome of L2 training (main hypothesis 2). One possible factor is the length of residence in the target language country (hypothesis 2.1). Those learners who have spent more time in the target language environment might possibly benefit more from using the game due to greater previous exposure to L2 input, which could provide a more solid foundation to build on when developing the trained L2 skill. Positive associations between learners' length of stay in Sweden and response time or accuracy improvement scores in the game would suggest such a relationship. Alternatively, those learners who have spent a shorter time in Sweden might show relatively greater improvements, due to having very little initial knowledge of the trained L2 features. For such learners, the present experiment might capture the complete developmental trajectory from zero to functional knowledge of Swedish word accents and their association with suffixes, manifested as a relatively greater improvement from the beginning to the end of the training period. Secondly, formal Swedish instruction might also play a role in learners' performance on the game, given that they have received an adequate amount of instruction relevant to the trained L2 skill (hypothesis 2.2). Failing to find a positive relationship between length of Swedish instruction and response time or accuracy improvement scores would be consistent with the assumption that the game provides a form of L2 training not commonly implemented in Swedish L2 teaching.

Perceptual training with the game might be hypothesized to influence learners' production of the targeted L2 features (main hypothesis 3). Based on previous studies reporting more native-like speech production following training in segmental or suprasegmental features of an L2 (Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2003), participants were expected to use word accents in a more native-like way in their L2 speech, as would be reflected in increased accuracy in the production data post-training relative to pre-training (hypothesis 3.1). Previous research has often failed to find a direct relationship between learners' improvement on perceptual and production measures, suggesting that development of perceptual skills might not directly translate into increased L2 production accuracy (Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2003). Therefore, production accuracy improvement scores might not be associated with a corresponding increase in accuracy or decrease of response speed in the game (hypothesis 3.2).

3. Method

3.1. Participants

Nineteen L2 learners of Swedish (13 females) participated in the testing of the prototype, with an average age of 23 years ($SD = 1.8$). All of them were NSs of non-tone languages (see Table 1 in the Supplementary Material for participant background information). Their self-reported level of Swedish proficiency corresponded to A1 to B2 levels of the Common European Framework of Reference. On average, they had received instruction in Swedish for 22.2 months ($SD = 24.6$), abroad and/or in Sweden, and had lived in Sweden on an average of 5.8 months ($SD = 4.7$).

3.2. The prototype

In the prototype, sentence items were shown as text in dark font color, in the middle of a white presentation area. Each item was presented auditorily as well, and the start of the visual and auditory presentations were synchronized. Sentences were visually divided into three parts as shown in Fig. 1. The first part consisted of the beginning of the sentence up until the target word that was associated with the prosodic predictive cue. Next, the two possible alternative continuations of the target word were shown on two buttons with text representing the choices. The choice was always between two suffixes, one of which was cued by the word accent on the preceding target word stem. After the two buttons, the part of the sentence that followed the target word plus suffix choice was shown as text. The buttons were only enabled when the cue had been presented in the audio playback. When the player then clicked on one of the suffix choices, the audio of the rest of the sentence was played up, with the correct continuation regardless of the player choice, and feedback was displayed. The whole sentence was shown at that point as a single stretch of text, incorporating the appropriate continuation to the target word (Fig. 2). When the choice of suffix was correct, the text color switched to green; otherwise it remained in the default, dark text color. By associating a salient visual signal only to correct choices, emphasis was on providing positive feedback. The players also had the possibility of replaying the audio of the whole sentence after responding, as many times as they wished.

Sentences were presented in game rounds, consisting of 18–20 items. At the end of a round, statistics were shown summarizing performance at the last round as well as the previous rounds of the current level (except for level 3, where all the rounds played on the same day were displayed). At this point, the player could choose to stop playing or to continue with the next round. Returning after a break from the game, players could continue where they left off.

The prototype also implemented a simple, adaptive mechanism that advanced the player to the next level as soon as performance indicated learning the task of the current level. This performance limit was defined as 80% accuracy calculated over a whole round. The minimum amount of play required at a given level was two rounds. The prototype had three main levels, with three sublevels at both main level 1 and 2 (for a description of the levels, see section 3.3). In order to make sure that all participants got the opportunity to play at each level, players who had spent too much time on a single level, in the sense that they would not have been able to complete the whole game before the end of the experiment continuing at the same pace, were automatically advanced to the next level (a next sublevel under a main level or a new main level), even if they had not reached 80% accuracy. In order to determine whether such progression was necessary, the actual time that a participant had already spent playing was evaluated against expected playtime values per main level and sublevel set by the authors. Expected playtime was three complete days for main level 1 (one day for each sublevel) and six days for main level 2 (two days for each sublevel). In this way, a guaranteed one day of practice was left for the final main level 3, considering the requirement that participants were expected to play the game for at least ten days during a two-week-period.

As mentioned above, one form of feedback the player received was always immediate, following choice of a suffix continuation after the predictive cue. Furthermore, summary statistics displayed after each round of play showed the percentage of correct choices, the average speed of responding, and response speed considering only correct choices (see Fig. 1 in the Supplementary Material for an example). Points were calculated taking into account the proportion of correct responses and the speed of these. The point system was formulated in a way that very fast but apparently random choices (around chance level) were not rewarded. This summary after each round served to give players feedback on their progression towards the goals within the game (cf. Kiili, 2005). By selecting a menu option on the website of the prototype, players could also view a similarly structured summary for every round they had played, presented in a chronological order, organized according to the levels of the game. Here, they could assess their progression towards reaching high enough accuracy to move to the next level. They could also compare their current performance level, both in terms of speed and accuracy, to those at earliest stages of the game. This gave players the opportunity to formulate their own future goals and evaluate their progression towards these.

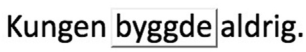

Fig. 2. Display of the correct suffix continuation in the game prototype after the player made a choice of suffix. After a correct response, the sentence was shown in green instead.

Instruction texts about each main level were made available for players on the game website, both in English and in the target L2, Swedish. By providing detailed explanations and examples, these texts also served as a form of scaffolding. Key concepts such as accent 1 and accent 2, as well as focused and unfocused sentence positions, conditioning the way the intonational contour of word accents are realized (cf. Section 3.3), were defined and explained. The fact that word accents are associated with different suffixes was also mentioned, and associations to the singular suffix *-en* and the plural suffix *-ar* were exemplified. To facilitate the recognition and the discrimination of the predictive tonal cue, visual illustration of the tone curves of word accents were provided along with corresponding audio examples (see Fig. 2 in the Supplementary Material for the illustrations). Players could return to the instruction text at any point while using the game.

3.2.1. Technical description of the prototype

The prototype was imbedded in a website, developed using the ASP.NET MVC framework. The website ran on a server at Lund University and participants could access the game after authentication in the browser with pre-assigned credentials. The interactive part of the game was implemented using HTML5 and JQuery, and the public SoundJS API was used to manage audio playback. The website was optimized for viewing in Firefox web browser, and participants were requested to use that specific browser. In order to decrease variations related to playing the prototype using a wide range of devices, the game was designed so that it did not load on mobile devices.

At the beginning of each game round, the client-side requested the game material from the server for the complete next round to be presented for the specific authenticated user. The server-side application compiled the appropriate game material based on the current status of the user in the game. The game material as well as player-associated data was stored in a local database on the server. The presentation of the sentence items and the collection of response data were then managed by the client-side script. At the end of each round, response data was sent to the server, where it was saved in the local database, where scores as well as summary statistics were calculated taking into account the results of the most recent rounds.

The website provided functionality for each player to view accuracy and response time statistics per round. They could also track the amount of time they had spent playing the game and play time distribution throughout the test period. Participants had access only to data concerning their own results, whereas users with administrative rights, the experiment leaders, could monitor the activity of all players.

3.3. Materials

In the game prototype, each presented item was a single sentence. The target word after which the choice alternatives were displayed was either a monosyllabic verb or noun stem. The alternative continuations were always inflectional suffixes attaching to the target word stem. The noun sentences were constructed by selecting the 20 most frequent nouns from the Stockholm-Umeå corpus (Stockholm-Umeå Corpus SUC 1.0, 1996), placing each in 5 different carrier sentences. For the verb sentences, the 45 most frequent verbs in SUC were chosen to appear in 8 different carrier sentences. Not every verb could meaningfully occur in each carrier sentence due to differences in argument and thematic structure (transitive, intransitive and reflexive verbs). Therefore, there was a certain variation as regards the number of different sentence frames a verb appeared in. A set of 100 verb sentences was constructed in this way. Within the same carrier sentence, each target word stem appeared once with an accent 1 suffix and once with an accent 2 suffix. For the noun stems, the accent 1 version was a definite singular noun form, taking the inflectional suffix *-en*, and the accent 2 version constituted of the indefinite plural form, with one of the suffixes *-ar* and *-or*. For verb stems, the accent 1 version was in the present tense form, as indicated by the inflection *-er*, and the accent 2 version was in the past tense (preterite) form, taking one of the suffixes *-te*, *-de*, or *-e*.

The game was divided into three main levels (1, 2 and 3), and main levels 1 and 2 were further divided into three sublevels each. Sublevels are referred to as level 1.1, level 1.2 etc., where the first number identifies the main level and the second number indicates the sublevel under the main level. The difference between sublevels of the same main level concerned the type of target word. At the first two sublevels of both main level 1 and 2 (levels 1.1 and 1.2 as well as 2.1 and 2.2), all target words were of the same type, nouns and verbs, respectively. At the third sublevel (levels 1.3 and 2.3), both nouns and verbs were presented in a mixed manner. Level 3 had no further subdivision, corresponding therefore to a single sublevel, due to the fact that sentences with both noun and verb targets were randomly mixed throughout the entire level. Originally, each of the levels 1.1, 1.2 as well as 2.1 and 2.2 contained 200 sentences (100 sentence pairs, appearing with both accent 1 and accent 2 on the target word), but due to artifacts in the recordings, one sentence pair was removed from level 1.1, and two sentence pairs were removed from level 2.1, resulting in 198 sentences for level 1.1, 200 sentences for level 1.2, 196 sentences for level 2.1 and 200 sentences for level 2.2. The mixed levels 1.3 and 2.3 consisted of all the sentences that appeared on the previous two levels. Therefore, the material for level 1.3 was the combination of the 398 (198 noun + 200 verb) sentences associated with levels 1.1 and 1.2, whereas level 2.3 contained 396 items (196 noun + 200 verb sentences) from levels 2.1 and 2.2. The total number of unique sentences was thus 794. At main level 1, the target word always appeared in focused sentence position, i.e. it was pronounced with the prosodic characteristics of the most prominent phrase, constituting the most informative element in the sentence. At main level 2, targets were invariably in unfocused sentence position. Both accent 1 and accent 2 are characterized by different acoustic patterns depending on whether they are realized on words in focused or unfocused sentence positions (Bruce, 1977) (see Fig. 2 in the Supplementary Material), and due to the less prominent intonational contours of word accents in unfocused positions, the discrimination of word accent types was thought to be initially more challenging here than at the previous main level. For the final stage, level 3, presented sentences were drawn from the complete stimulus set (794 sentences), randomly mixing noun and verb targets as well as focused and unfocused variants.

The stimulus sentences were recorded in an anechoic chamber by a female speaker of Central Swedish. The sentences were read as responses to context questions to ensure the placement of focus on the appropriate word. Each sentence was read twice. After the recordings, phonetically trained NSs of Swedish screened the stimulus material and excluded items with artifacts. Three sentence pairs were excluded in this way (from levels 1.1 and 2.1). Using the speech editing software Praat (Boersma & Weenink, 2014), sentences were subsequently divided into two parts, cutting after the target word stem syllable, at zero crossing points.

Stimulus selection and sequencing in the prototype game was designed so as to promote participants' exposure to a varied and more complete range of stimulus material. When a player started a new level, a stimulus-sequence was generated for the specific player and divided into rounds so that all available sentences in the given level appeared exactly once. For instance, for level 1.1 where there were 198 stimulus sentences available, a sequence of 10 rounds were generated, with 9 rounds consisting of 20 sentence items and 1 round of 18 items. Sentences were randomly sequenced, while ensuring that each round had an equal number of accent 1 and accent 2 sentences. The sequencing algorithm also ensured that the same sentence item was not presented in both an accent 1 and accent 2 version within the same round. At levels of mixed target types, sentences with verbs and nouns were distributed equally within the round, as long as there were enough not-yet-presented sentences available for both target types. If the participant interrupted playing during a level, the game continued with the next round in the pre-generated sequence following the last round completed before the interruption. For the 3rd, final level, sentences were randomly selected from the complete stimulus material before each round, ensuring 50-50% presentation of accent 1 and accent 2 sentences.

Participants played the prototype for a period of 14 days, during which time they were requested to play at least on 10 different days, for a minimum of 15 min and a maximum of 60 min each day. Response time and accuracy of response for each presented sentence item as well as playtime measures were collected and stored in a database on the website's server, associated with specific players. Response times were measured from the offset of the auditory presentation of the predictive cue, i.e. from the end of the target word stem, until the moment the player pressed one of the buttons representing an alternative continuation. Tracking of playtime was based on logging the exact date and time of day when the player started as well as finished playing each round.

3.4. Production test

During the production test before and after the training period, participants read aloud a list of 20 sentences. Participants read the same set of sentences both pre- and post-test. Learners were instructed to read the sentences as responses to context questions provided next to each item, in order to control for the placement of sentence focus. Similar to the materials in the game, each sentence had a target word, a noun or a verb, which in its correct pronunciation was associated with either accent 1 or accent 2 depending on the suffix appearing on the stem. For example, the test sentence *Anna såg bilen på TV* 'Anna saw the car on TV' was read as a response to the context question *Var såg Anna bilen?* 'Where did Anna see the car?'. The most prominent phrase in the response carrying new information elicited by the question, and hence the constituent expected to receive intonational sentence focus, was the sentence-final *på TV* 'on TV'. In native speaker speech, the target word (*bilen* 'car') preceding the final, focused, constituent would therefore always have been pronounced with the word accent pattern associated with unfocused sentence position. Ten of the sentences had noun targets, with either the singular definite suffix *-en* (associated with accent 1 on the stem) or with the plural indefinite suffix *-ar* (accent 2). Ten sentences had verb targets, appearing either with the present tense suffix *-er* (accent 1) or the past tense suffix '*-te*' or '*-de*' (accent 2). There was one carrier sentence for nouns and one for verbs. In order to see whether knowledge gained during playing the prototype would generalize to novel stimuli, the target words were selected so that half of them appeared in the game and half of them did not.

All participants were recorded using a portable recorder (TASCAM DR-07) in the same experiment room. Subsequent analysis was conducted in Praat, during which a NS of Central Swedish, a trained phonetician, performed audiovisual evaluation of the fundamental frequency (F_0) curve of the target words. These were classified into five pre-defined categories, coded as (1) focused accent 1, (2) focused accent 2, (3) unfocused accent 1, (4) unfocused accent 2 or (5) no resemblance to either Central Swedish accent. One additional class, that of South Swedish accents, was introduced during the analysis since the F_0 curves of some of the target words were found to resemble accent patterns characteristic of the dialect spoken in the area where the participants lived in Sweden during the test period.

3.5. Data analysis

Initial accuracy and initial response time averages were calculated based on the first two rounds played (40 items), which was at level 1.1 for all players. Final accuracy and final response time values were calculated based on the average of the last 10 rounds (200 items), which were played on the final level by all participants (level 3). A smaller number of rounds were used for the initial as compared to the final performance measurements, due to the fact that some players progressed to level 1.2 already by the third round played. Also, a significantly larger number of rounds were played at level 3 relative to level 1.1, 167 vs. 9 rounds on average.

Table 1
Average accuracy (percentage of correct responses) per level.

Level	Accuracy (%)
Initial	59.9
1.2	66.2
1.3	66.4
2.1	73.5
2.2	76.4
2.3	75.1
Final	76.0

Note. Initial level value: the average of the first 2 rounds played at level 1.1. Final level value: the average of the last 10 rounds played at level 3.

For all calculations, items with response times (RTs) above 5000 ms were excluded (1.28% of all items) and RTs were calculated by including only correct responses. Data were collapsed over word accents, by considering both accent 1 and accent 2 target words for all accuracy and RT analyses. The time participants spent on a specific level or playing the game in total was measured by summing the time span difference between the beginning and the end of each round, on a specific level or for the whole gaming period. All correlational tests were two-tailed. In pairwise comparisons on accuracy and RT measures per level, p-values were adjusted for multiple comparisons using Bonferroni-correction.

On the production data, a repeated measures ANOVA was carried out with the within-subjects factors Pre-post (levels: pre-test, post-test), Accent (levels: accent 1, accent 2) and Game-presence (levels: present, absent). The factor Game-presence distinguished between target words that appeared in the game material and those that did not.

4. Results

4.1. Game data

Level and accuracy showed a correlation ($r = 0.923$, $p = 0.003$) as accuracy increased throughout the levels of the game (see Table 1 and Fig. 3). Average accuracy for the first two rounds was 59.9% and 76.0% for the last ten rounds (hypothesis 1.1). Pairwise comparisons of accuracy values at each level also indicated a significant accuracy increase from the initial two to the final ten rounds ($p = 0.002$). A significant accuracy increase was observed even from the first mixed level, level 1.3 to the final

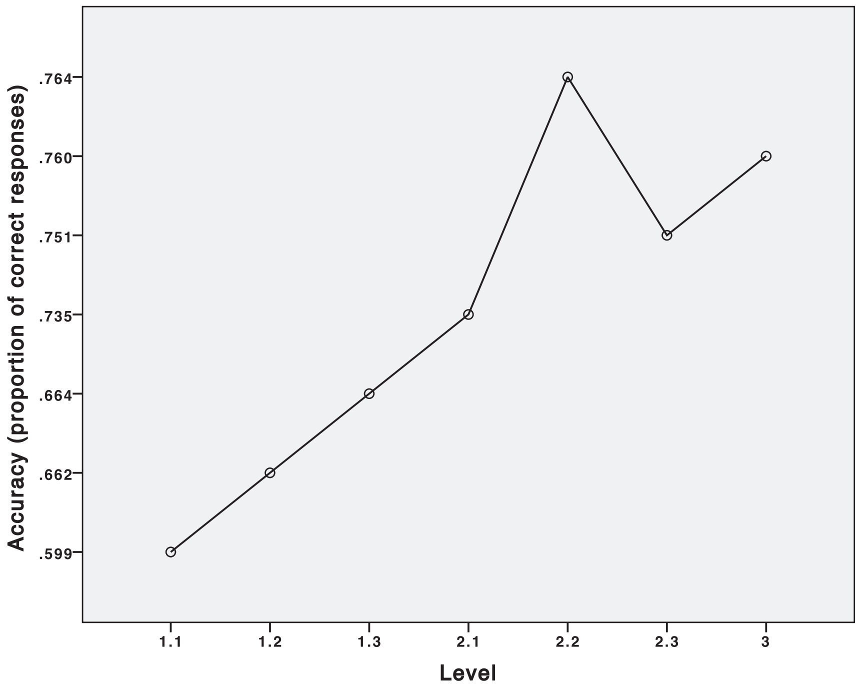

Fig. 3. Average response accuracy increased throughout the levels in the game.

Table 2
Average RTs for correct responses per level.

Level	RT (ms)
Initial	1787
1.2	1253
1.3	1357
2.1	1102
2.2	1002
2.3	1233
Final	748

Note. Initial level value: the average of the first 2 rounds played at level 1.1. Final level value: the average of the last 10 rounds played at level 3.

ten rounds ($p = 0.013$): at level 1.3 some learning was already expected to take place due to training; nevertheless, the sentence material was more similar to the last level due to mixing noun and verb target word types, providing a more homogenous comparison in that respect. For results of the pairwise comparisons, see Table 2 in the Supplementary Material.

There was also a negative correlation between level and RTs of correct responses averaged for level ($r = .\text{-}840$, $p = 0.018$) indicating that RTs generally decreased throughout the game (see Table 2 and Fig. 4). Average RT was 1787 ms for the first two rounds of the game and 748 ms for the last ten rounds (hypothesis 1.2). Pairwise comparisons of RT values at each level showed a significant RT reduction from the initial two to the final ten rounds ($p < 0.000$). Such significant RT reduction could be observed even from each mixed level, level 1.3 ($p = 0.007$) as well as level 2.3 ($p = 0.009$), to the last ten rounds. For results of the pairwise comparisons, see Table 3 in the Supplementary Material.

On average, players spent 9.10 ($SD = 8.42$) rounds (block of 20 sentences) on level 1.1 and 166.79 ($SD = 109.77$) on the last level (level 3). Levels in between took an average of 5.56 rounds to complete (level 1.2: $M = 6.84$, $SD = 6.81$; level 1.3: $M = 6.63$, $SD = 5.83$; level 2.1: $M = 5.84$, $SD = 7.54$; level 2.2: $M = 4.26$, $SD = 6.04$; level 2.3: $M = 4.21$, $SD = 3.79$). There was no correlation between total playtime and initial to final RT improvement ($r = 0.081$, $p = 0.743$) or initial to final accuracy increase ($r = 0.240$, $p = 0.322$) (hypothesis 1.3). Due to the proportionally greater amount of time players spent on level 3 training the complete game material, correlational analyses were conducted considering play time on that level. Results showed that the total time in hours players spent on the final level 3 correlated with the difference between final and initial accuracy ($r = 0.573$, $p = 0.010$), indicating that final accuracy increase relative to the initial performance was greater the more time spent on the last level (see Fig. 5). There was no correlation between initial to final RT improvement and time spent on level 3 ($r = 0.141$, $p = 0.564$).

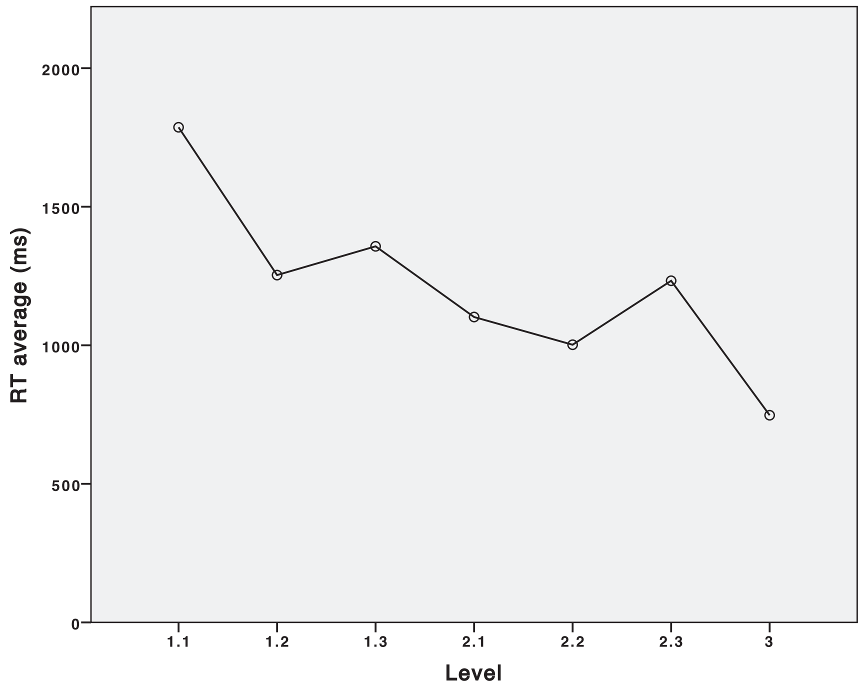

Fig. 4. There was a negative correlation between average RTs for correct responses and levels in the game.

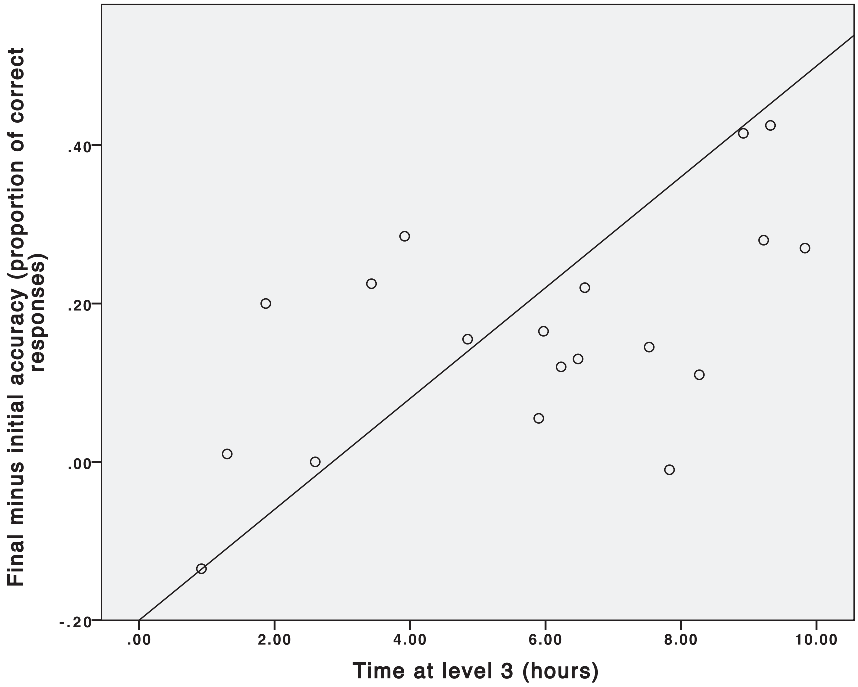

Fig. 5. Increase in accuracy from the initial to the final level correlated with the time individual participants spent playing at the highest level of the game.

The time (in months) participants reported to have had spent in Sweden showed a negative correlation with the difference between initial RT minus final RT of correct responses ($r = .\text{-}467$, $p = 0.044$). Thus, less time spent in Sweden was associated with greater relative RT decrease (hypothesis 2.1). No correlation was found between accuracy improvement and reported length of stay in Sweden ($r = .\text{-}028$, $p = 0.908$). There were no significant correlations with the reported total length of Swedish instruction (initial RT minus final RT of correct responses, $r = 0.071$, $p = 0.772$; final minus initial accuracy: $r = 0.235$, $p = 0.333$) (hypothesis 2.2).

4.2. Production data

The accuracy with which learners produced both word accents increased significantly from the pre-test to the post-test, as indicated by a main effect of Pre-post ($F(1,18) = 11.47$, $p = 0.003$) (hypothesis 3.1). There was no interaction between the factors Pre-post, Accent and Game-presence ($F(1,18) = 0.329$, $p = 0.573$). Accent 1 was generally associated with higher production accuracy relative to accent 2, reflected in a main effect for Accent ($F(1,18) = 22.86$, $p < 0.001$). The accuracy advantage for accent 1 remained stable across the test occasions based on the lack of interactions with the factor Accent. Before training, mean accuracy was 68.75% ($SD = 34.79$) for accent 1 and 16.12% ($SD = 27.15$) for accent 2, which increased to 73.18% ($SD = 30.13$) and 26.01% ($SD = 30.36$) respectively, by post-training. There were no interactions involving the factor Game-presence either, suggesting that training-effects transferred even to novel words not included in the game material. For production accuracy values per participant and condition, see Table 4 in the Supplementary Material. Accuracy increase in production by post-test did not correlate with initial versus final accuracy increase ($r = 0.035$, $p = 0.885$) or decrease in RTs of correct responses ($r = 0.212$, $p = 0.383$) in the game (hypothesis 3.2).

5. Discussion

In the present study, we tested the prototype of an L2 learning game which has been developed to facilitate the acquisition of predictive language processing in the L2. The present implementation of the game mechanics focused on Swedish and trained L2 learners to use tone patterns on word stems as anticipatory cues to immediately upcoming suffixes. Learners with low to intermediate proficiency in Swedish used the game for a period of two weeks, while their activity and progress were continuously logged. Results of the experiment indicate that the prototype effectively promoted the learning of the trained language feature, which was seen in a gradually more accurate and faster performance of the predictive task in the game as well as in a correlation between accuracy improvement and time spent on playing at the last level. Furthermore, perceptual training with the prototype seems to have led to increased accuracy even in the production of the L2 tonal cue.

While accuracy in predicting upcoming language material based on the tonal cue was generally close to chance level (59.9%) at the initial stage of playing the game, the proportion of correct responses on average increased to 76.0% by the final rounds of the test period. Taken together with the observed general decrease of response times throughout the game, the obtained results indicate that playing the prototype effectively promoted the acquisition of the trained tone-suffix co-variations as well as processing the anticipatory function of tones. Furthermore, more extensive play seemed to have led to relatively greater improvements in performing the task, as longer total playtime on the last level (level 3) was associated with greater differences between individual players' final and initial accuracy. Findings related to play time at the final level can be considered especially informative about the effectiveness of the game, due to the fact that this was the level where all participants spent a significant amount of time, after a relatively rapid progression until this point in several cases. As no correlations were obtained between play time and relative RT decrease from the initial to the final values, it seems that more extensive practice did not straightforwardly lead to a continuous speeding up of performing the suffix prediction. Possibly, significant increase in speed might not be an issue of a simple quantitative change but might take place as a result of a more qualitative reorganization as the skill becomes automatized.

The tendency of gradually decreasing response latencies with increasing level throughout the game is only broken by level 1.3 and 2.3, where RTs temporally increased (Fig. 4 and Table 2). At these levels, the word accent cue appeared either on a verb or a noun in a randomly mixed fashion, while level 1.1 and 1.2 as well as level 2.1 and 2.2 focused on a single target word class. Mixing target word types thus seems to increase task difficulty. Arguably, learners needed to acquire not only the pre-activation of language features associated with a cue, but also the selective inhibition of predictions that are not relevant in the current context. For example, following accent 1, strong pre-activation of the singular definite suffix *−en* attaching to nouns is beneficial as long as the context of the word accent cue is a noun. However, pre-activation of the same feature might even slow down processing if the context is instead a verb, due to the emerging competition with the correct present tense inflection *-er*. The need to acquire the ability to appropriately inhibit temporally non-relevant predictions could have created additional difficulty at levels 1.3 and 2.3. Nevertheless, during the relatively long time most players spent at level 3 (similarly mixing noun and verb targets), results indicate they eventually seem to have absorbed such context-based modulation of pre-activations, leading to decreased final RT values.

Appropriate performance of the task in the game was crucially dependent on correct identification of word accent patterns constituting the predictive cue. The fact that participants generally became more accurate at producing word accents by the end of the gaming period suggests that this form of predictive speech processing training leads to gains even in the production of the relevant L2 cue. Increased accuracy was observed even in the pronunciation of novel target words not included in the training material. The finding that game performance data did not show a correlation with specific post-test production improvements is in accordance with the results of several previous studies that similarly failed to observe a direct one-to-one correspondence between increased production accuracy, following training in the identification of L2 segmental or supra-segmental features, and corresponding perceptual tests (e.g. Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2003). In general, these results are in line with proposals in the field of segmental phonology suggesting that the degree to which learners form accurate perceptual L2 categories limits the extent to which they approximate NSs in the production of the same sounds (Flege, 1995). While modifications in perceptual categories might be a prerequisite for attaining more accurate production, it is not assumed that any change in perception would automatically be reflected in learners' speech production. For instance, learners might experience difficulties in acquiring the required articulatory commands to produce the L2 sound or they might be reluctant to sound more native-like due to psychosocial reasons (Flege, 1999). Furthermore, dissociation between production and perception improvements is consistent with the assumption that different neural networks are involved in transforming acoustic input into articulatory motor representations (dorsal stream) as opposed to mapping speech signals to their meaning during comprehension (ventral stream) (Hickok & Poeppel, 2007).

A negative correlation was obtained between RT decrease from the initial to the final stage in the game and the length of residence in Sweden. In more proficient learners, exposure to Swedish L2 in the target language country was previously found to be associated with gradual implicit acquisition of predictive tone processing (Schremm et al., 2016). It is therefore possible that those participants who had been in Sweden for a longer time had some basic implicit knowledge of the predictive significance of tones already at the outset of the study. From this perspective, the obtained correlation with length of residence in the current study would suggest that the game had greater benefit for those who presumably had less initial knowledge of the tone-suffix associations. Thus, it seems that training provided by the game is especially effective for speeding up the acquisition of predictive processing at early stages of L2 acquisition. Moreover, no association was observed between game performance and the length of instruction the participants received in the L2, which indicates that the game might be able to develop skills not targeted by traditional L2 Swedish instruction.

5.1. Platform game implementation of the prototype

Results of the study indicated that the game mechanics work as intended and that the game is an effective tool for promoting the acquisition of L2 predictive processing. Therefore, the prototype is currently being developed into a full-featured engaging platform game, the first version of which is presented through screen shots in Figs. 6,7. Implementation as a platform game was motivated by a number of considerations. First, alternative linguistic targets following a predictive cue can be appropriately represented by two or more platforms, one of which the player is required to choose in order to be able to move on. Such visual presentation in the form of platforms makes it easy for the player to quickly recognize an upcoming decision point and get an overview of the possible choices. Also, the speed and fluency with which the player moves across the platforms can be important contributors to the game experience in this genre, which can be thought to promote attempts

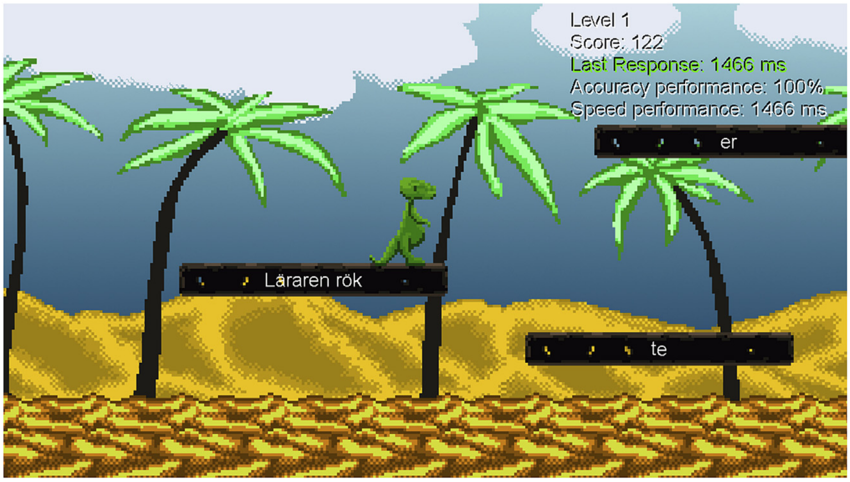

Fig. 6. Screenshot from the platform game implementation. Alternative choices following the auditory predictive cue are represented by different platforms, here displaying the present tense versus past tense suffixes *-er* and *-te* after the sentence fragment *Läraren rök-* 'The teacher smoke-'.

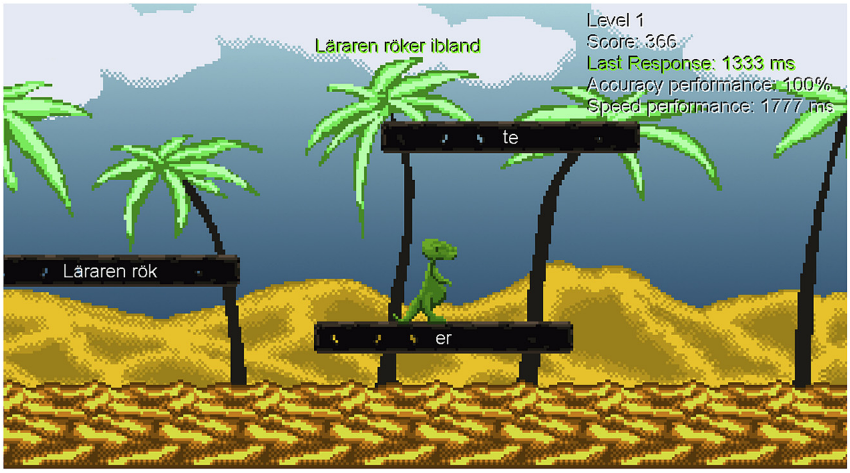

Fig. 7. Screenshot from the platform game implementation, showing the moment when the player has selected a continuation to the target word stem by having the dinosaur figure jump on the *-er* '-s' present tense suffix platform after processing the sentence fragment *Läraren rök-* 'The teacher smoke-'.

at performing rapid choices between the target alternatives following a cue. Furthermore, linguistic co-variations underlying the predictive strategy to be learned can be made inherently task-relevant in this context, by ensuring that correct platform choices and subsequent rewards are dependent on the correct knowledge of the relevant co-variation relations. This might be thought to support the development of strong cue-target associations, the presence of which might be essential for the emergence of effective predictive processing strategies (Grüter et al., 2012; Hopp, 2013). Furthermore, the generally straightforward nature of the game-mechanics and rules ensures that players quickly develop an understanding of the game itself and are then able to focus on the learning material. Finally, platform games offer the possibility of relatively shorter periods of casual play that still create a sense of achievement and learning: it is inherent in the genre that players can frequently experience small achievements, such as each time when they have chosen a correct platform.

6. Conclusions

In conclusion, after only a two-week period of play with the game prototype presented in this paper, participants' performance results showed learning effects concerning predictive L2 processing. These are assumed to involve the gradual acquisition of the underlying tonal cue-suffix associations in the Swedish L2 as well as the development of faster performance of the targeted anticipatory speech processing skill. Furthermore, purely perceptual training led to more accurate production of the tonal cue. The game, therefore, seems to be an appropriate tool for developing an essential predictive feature of Swedish in L2 learners' speech, at the same time as it enables learners to acquire anticipatory strategies that can underlie more effective processing of the target L2, promoting efficient and successful communication. As the game idea introduced simulates mechanisms assumed to be related to general predictive linguistic processing, it has a broad range of possible implementations for different types of predictive co-variations in a variety of L2s. Importantly, such L2 skills might be especially difficult to train without specially designed educational tools, highlighting the value of harnessing the motivational power of a digital game that provides learners with extensive exposure to relevant co-variations, while also creating opportunities for the actual performance of the developing L2 predictive skills.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (grant number 2014.0139) and Marcus and Amalia Wallenberg Foundation (grant number 2014.0039).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.compedu.2017.07.006.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.
Bannert, R. (2004). *På väg mot svenskt uttal*. Lund: Studentlitteratur.
Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B, 364*, 1235–1243.
Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer* [Computer software]. Retrieved from: (Version 5.4.08). http://www.praat.org.
Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify/r/and/l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.
Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerups.
Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4*, 25.
Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 233–253.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper and Row.
DeKeyser, R. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning, 55*(Suppl. 1), 1–25.
DeKeyser, R. M., & Criado, R. (2012). Automatization, skill acquisition, and practice in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 323–331). Oxford, UK: Wiley-Blackwell.
Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B, 369*, 20120394.
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121.
Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21*(5), 629–634.
Dussias, P. E., Valdes Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35*, 353–438.
Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*, 164–194.
Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*, 83–107.
Entertainment Software Association (ESA). (2016, April). *2016 Essential facts about the computer and video game industry*. Retrieved from: http://essentialfacts.theesa.com/Essential-Facts-2016.pdf. (Accessed 1 December 2016).
Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). Timonium, MD: York Press.
Flege, J. E. (1999). The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granveille, & A. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences* (pp. 1273–1276).
Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836.
Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.
Gosselke Berthelsen, S., Horne, M., Shtyrov, Y., Brännström, J., & Roll, M. (Submitted). Early second language learners' processing of pre-activating prosodic cues: An electrophysiological study of Swedish word accents.
Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215.
Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.
Hoffmann, J., & Sebald, A. (2005). When obvious covariations are not even learned implicitly. *European Journal of Cognitive Psychology, 17*, 449–480.
Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*, 33–56.
Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics, 52*, 105–120.
Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism, 4*, 257–282.
Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.
Kiili, K. (2005). Digital game-based Learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13–24.
Kuperberg, G., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience, 31*(1), 32–59.
Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133*, 335–342.
Kutas, M., DeLong, K., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain* (pp. 190–207). Oxford: Oxford University Press.
Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484–502.

Lee, E.-K., Lu, D. H.-Y., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition, 16*(4), 761–775.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.
Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow, & M. J. Farr (Eds.), *Aptitude, learning and instruction III: Conative and affective process analyses* (pp. 223–253). Hillsdale, NJ: Erlbaum.
McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics, 21*(3), 395–423.
McElree, B., Jia, G., & Litvak, A. (2000). The time course of conceptual processing in three bilingual populations. *Journal of Memory and Language, 42*(2), 229–254.
Mitsugi, S., & MacWhinney, B. (2015). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*, 19–35.
Perruchet, P. (2008). Implicit learning. In J. Byrne (Series Ed.) & H. L. Roediger, III (Vol. Ed.), Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory (pp. 597–621), Oxford: Elsevier.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238.
Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism, 5*(4), 409–453.
Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105–110.
Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 6*, 855–863.
Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.
Roll, M. (2015). A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics, 38*, 149–162.
Roll, M., & Horne, M. (2011). Interaction of right- and left-edge prosodic boundaries in syntactic parsing. *Brain Research, 1402*, 93–100.
Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology - ERPs of Swedish word processing. *Brain Research, 1330*, 114–123.
Roll, M., Söderström, P., & Horne, M. (2013). Word-stem tones cue suffixes in the brain. *Brain Research, 1520*, 116–120.
Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., et al. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain & Language, 150*, 14–21.
Schremm, A., Söderström, P., Horne, M., & Roll, M. (2016). Implicit acquisition of tone-suffix connections in L2 learners of Swedish. *The Mental Lexicon, 11*(1), 55–75.
Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23*, 473–500.
Shi, L.-F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech Language and Hearing Research, 53*(4), 821–835.
Söderström, P., Horne, M., & Roll, M. (2016). Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research*. http://dx.doi.org/10.1007/s10936-016-9434-2. Advance online publication.
Söderström, P., Roll, M., & Horne, M. (2012). Processing morphologically conditioned word accents. *The Mental Lexicon, 7*(1), 77–89.
Stockholm-Umeå Corpus SUC 1.0. (1996). Stockholm University: Department of Linguistics and University of Umeå: Department of linguistics.
Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 443–467.
Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America, 113*, 1033–1043.
Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*, B63–B72.
Whitton, N. (2011). Game engagement theory and adult learning. *Simulation & Gaming, 42*(5), 596–609.
Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288.
Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex, 68*, 20–32.