French Listeners Can Use Stress to Segment Words in an Artificial Language

Michael D. Tyler

MARCS Auditory Laboratories
University of Western Sydney, Australia
m.tyler@uws.edu.au

Abstract

This study is concerned with whether nonnative listeners are able to use information not present in their native language to recognise new words in an unknown foreign language. One potential source of information, lexical stress, is lacking in French, and French listeners have poor discrimination of nonwords differing only in lexical stress. Vroomen, Tuomainen and de Gelder (1998) showed also that Finnish and Dutch listeners, whose languages have lexical stress, but not French listeners, were able to use a pitch accent to aid segmentation of an artificial language. The aim of the present study was to extend that study by using an artificial language that was a step closer to the complexity observed in natural speech. In Experiment 1, French and Dutch participants listened to an artificial language consisting of 6 three-syllable words and 3 four-syllable words during a 10-min exposure. For half of the participants the only information available for word segmentation was the statistical grouping of syllables into words, but the other half of participants were also provided with the Vroomen et al. pitch accent. Contrary to the results of Vroomen et al., both the French and Dutch listeners benefited from the stress cue. To test whether the increased complexity of the artificial language helped French listeners to recognise the pitch accent, in Experiment 2 French and Dutch listeners were exposed to the exact stimuli used in Vroomen et al.'s artificial language experiment. Again, both French and Dutch listeners benefited from the stress cue, indicating that the increased complexity was not responsible for the performance improvement in Experiment 1. The discrepancy between the results of Experiment 2 and those of Vroomen et al. may be due to procedural or cohort differences, but the important finding here is that listeners are not restricted to native language word segmentation strategies when confronted with an unknown foreign language.

1. Introduction

When listening to our native language we hear distinct words. However, unlike words on a page, which are separated by spaces in English orthography, words in speech are not separated by acoustic gaps. How, then, do humans recognise words in continuous speech?

There are a number of models of spoken word recognition that seek to explain how adults recognise words that they already know (Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). However, these models require an already-developed vocabulary; they cannot explain how words are first recognised in the speech flow. Although infants are not the focus of this paper, it is informative to consider how they might solve this bootstrapping problem.

It is widely assumed that the first step in this bootstrapping process is *word segmentation*. That is, infants must first learn to recognise word forms in continuous speech before learning their referential properties. The question of how infants learn to segment words from a continuous speech stream without acoustic gaps between words is referred to here as *the word segmentation problem*.

In a widely cited study, Saffran, Aslin, and Newport (1996) proposed that infants solve the word segmentation problem by keeping track of transitional probabilities between adjacent syllables. Given a particular syllable, the transitional probability would be high if the next syllable often followed that specific syllable and low if the next syllable rarely followed that specific syllable. As such, transitional probabilities tend to be high for within-word sequences and low for between-word sequences. Thiessen and Saffran (2003) argued that it is these dips in transitional probability that might first cue infants to the presence of word boundaries in continuous speech.

Saffran et al. (1996) tested infants' sensitivity to distributional information by using an artificial language, consisting of four synthesised nonsense words, strung together in random order, with no acoustic gaps between them. They argued that the only way that infants would be able to segment the words of this new language would be via the tracking of transitional probabilities. After an exposure phase, infants' listening preferences were obtained for two types of test items: words from the language, or *partwords*, which were sequences of syllables from the exposure stream with low transitional probability (e.g., the last syllable of one word and the first two

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

syllables of another). Infants exhibited a novelty preference for partwords, suggesting that they had previously segmented, and habituated to, the words of the language during the exposure phase.

As a result of this initial language-general solution to the word segmentation problem, infants are thought to pick up language-specific regularities that provide additional cues to word boundaries (for a review, see Saffran, Werker, & Werner, 2006). One such strategy that appears to be beneficial for adult speakers of stress-timed languages is the *metrical segmentation strategy* (MSS, Cutler & Norris, 1988). Most words in English begin with a strong syllable, and Cutler and Norris showed that English listeners do treat strong syllables as potential word onsets. There is also evidence for this in infant populations. Nine-month-old English-learning infants prefer to listen to trochaically versus iambically stressed English words (Jusczyk, Cutler, & Redanz, 1993), and 7.5 month-olds missegment the sequence *taris* from the phrase *guitar is* (Jusczyk, Houston, & Newsome, 1999). This latter result suggests that young English-learning infants also treat stressed syllables as potential word onsets.

The purpose of this study is to consider the word segmentation problem for the case of an adult listening to a foreign language for the first time. Everyday experience attests to the fact that we do not hear distinct words in an unknown foreign language. How do adult second language learners solve the word segmentation problem? Obviously, the tracking of statistical probabilities is as possible a solution for adults as it is for infants, but native-language strategies, such as the MSS, may aid (or disrupt) initial word segmentation.

Vroomen, Tuomainen, & de Gelder (1998) conducted an experiment examining whether listeners can use properties of their native languages, vowel harmony and stress, to segment words in an unknown artificial language similar to that used by Saffran et al. (1996). Words in the harmonic language were composed of all front vowels or all back vowels, whereas in the disharmonic language front and back vowels were mixed. In the stressed language words contained a pitch accent, consisting of a fundamental frequency (F0) rise over the first syllable of the words and a fall over the next two syllables (all words were three syllables long). This was compared to a control condition in which the only cues to word segmentation were transitional probabilities between syllables (as in Saffran et al., 1996).

As Finnish has both front/back vowel harmony and stress, it was predicted that Finnish listeners would benefit from both sources of information. Dutch does not have vowel harmony, but it is stress timed and listeners do appear to use the MSS (Vroomen, van Zon, & de Gelder, 1996). As such, Dutch listeners were expected to benefit from the stress cue but not from the vowel harmony. Finally, French does not have vowel harmony, nor is it stress timed, so French listeners were not expected to benefit from either cue. The results conformed to these hypotheses. Therefore, it appears that listeners are restricted to their native-language word-segmentation strategies when confronted with an unknown foreign language, and that they do not learn to use new word boundary information after a short exposure to the language.

That French listeners have difficulty perceiving stress is also supported by a study examining Spanish and French listeners' ability to match nonsense words varying only in syllable stress as it occurs in Spanish (e.g., bópelo vs. bopélo or bopélo vs. bopeló, Dupoux, Pallier, Sebastian, & Mehler, 1997). French listeners had high error rates and slow reaction times, compared to Spanish listeners, regardless of whether the contrast was 1st versus 2nd syllable or 2nd versus 3rd syllable stress. In contrast, in a task focussed purely on phonological differences, French listeners, but not Spanish listeners, were able to ignore irrelevant stress variation.

Those experiments used an ABX paradigm, in which two different items are first presented and then the participant must decide whether a third item is the same as the first or the second item. To assess whether French listeners' stress difficulties were perceptual or at a higher level, Dupoux et al. (1997) conducted an additional experiment using a same-different (AX) task, which has lower memory requirements than the ABX task. In that task, French participants successfully discriminated tokens that differed only in stress placement. They performed equally well regardless of whether the interstimulus interval was 200 ms or 2200 ms, suggesting that their poor performance in the ABX task was not due to memory retention. Dupoux et al. concluded that differences in performance were due to the complexity of the task.

The findings of Dupoux et al. (1997) suggest that, while French listeners may be able to perceive stress, they may not have benefited from the stress cue in the artificial language of Vroomen et al. (1998) due to the complexity of the task. Given that natural language is much more complex than the artificial language used in that study, it is possible that the benefit of the MSS was underestimated. That is, if the stimulus had contained the complexity and variation of natural language, the Dutch and Finnish listeners would have outperformed the French listeners by an even greater margin. If this were the case, it would suggest that the second-language learner's native language background confers a significant advantage when the target language shares key structural characteristics with the native language. The purpose of this study, therefore, is to extend the Vroomen et al. (1998) findings by using an artificial language that is one step closer to the complexity and variability of natural speech.

2. Experiment 1: French vs. Dutch use of stress in an artificial language

The purpose of this experiment was to test whether the findings of Vroomen et al. (1998) would generalise to a more complex and varying artificial language. The artificial language used in Experiment 1 contained nine words, instead of six, words of varying length (three- and four-syllable words) instead of three-syllable words only, and the syllables of the language were constructed using six consonants instead of only four. Given that vowel harmony is not of central importance here, only French and Dutch listeners were tested in this experiment.

If, following the results of Dupoux et al. (1997), French listeners are even less likely to benefit from stress cues with a more complex artificial language, then they should perform even more poorly, relative to Dutch listeners, than in Vroomen et al. (1998). However, there is another possibility. Given that training Japanese listeners to perceive the difference between /r/ and /l/ is more effective if there is variability in stimulus tokens (Logan, Lively, & Pisoni, 1991), it is possible that the increased variation and complexity might help focus French listeners on the crucial constant of word-initial stress. If this is the case, then French performance should improve in the more complex language.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

2.1. Method

2.1.1. Design

The experiment had a factorial design with two between-participant independent variables. *Language background* had two levels, French and Dutch, and *language type* also had two levels, no stress and stress. Participants within each language group were randomly allocated to one of the two language type conditions.

2.1.2. Participants

There were 16 participants in each Language Background x Language Type group (N = 64). French participants were Psychology students at the Université de Bourgogne, Dijon, France. All participants in the no stress group (14 females, 2 males; M_{age} = 18.88 yr, SE = 0.37), and the stress group (15 females, 1 male; M_{age} = 19.38 yr, SE = 0.50) had learned some English at school (note: the experiment was conducted in French). Around half of the participants had also learned Spanish. Dutch participants were from the participant panel at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. All participants in the no stress group (14 females and 2 males; M_{age} = 21.81 yr, SE = 0.71), and the stress group (15 females and 1 male; M_{age} = 22.88 years, SE = 1.72) had learned English. The majority had also studied French and/or German at school.

2.1.3. Stimuli and apparatus

The artificial languages consisted of the concatenation of nine words. In artificial language studies the “words” usually consist of an equal number of syllables (e.g., three syllables; Saffran et al., 1996; Vroomen et al., 1998). To increase the complexity of the language, here the language consisted of words of varying length. Six of the words were trisyllabic and the other three words had four syllables each. The pool of 30 syllables used to construct the words of the languages were formed from the exhaustive combination of six consonants /p, b, m, f, s, k/ and five vowels /a, i, ɛ, ɔ, u/. These phonemes were chosen to be as phonetically similar as possible between French and Dutch.

To overcome potential native-language word boundary biases (see Reber & Perruchet, 2003), 16 different artificial languages were created; that is, a different random combination of syllables was used to create the words of each language. For each language, a stressed and unstressed version was created.

The languages were generated using the MBROLA diphone synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). Each consonant and vowel was assigned a length of 116 ms, resulting in a syllable length of 232 ms (following Peña, Bonatti, Nespor, & Mehler, 2002). To guard against the effect of phonetic differences between French and Dutch, and the effect of native language experience, half of the participants in each language group heard a language synthesized with a male French voice (MBROLA’s fr1 diphone database) and the other half a male Dutch voice (the nl2 diphone database).

Prosodic characteristics of the languages were the same as those used in Vroomen et al. (1998). In the unstressed version of the language, the fundamental frequency (F0) was set to a monotonous 120 Hz. For the stressed version, each word contained a pitch accent, consisting of a linear increase from 120 Hz to 170 Hz over the first syllable, followed by a linear decrease back to the 120Hz baseline over the following two syllables.

As in Vroomen et al. (1998), the exposure phase lasted for a total of 10 min, and was broken up into five blocks of 2 min. A 5-s fade-in and fade-out was applied to each 2-min block so that participants would not have access to word boundary cues from the beginning and end of the sequence. The words were presented in random order with the constraint that a given word could not follow itself.

There were 27 pairs of items in the test phase. One member of each pair was a word from the language and the other was a partword; that is, a sequence of syllables that occurred in the stream but crossed a word boundary. For half of the participants the partwords were formed from the last two (or three) syllables of a word and the first syllable of another, and for the other half they were formed from the last syllable of a word and the first two (or three) syllables of another. All nine words of the language were used in the test phase, along with nine partwords. Each word was paired with three partwords and each partword was paired with three words to counteract learning during the test phase. Following Vroomen et al. (1998) both words and partwords had the same pitch accent in the stress condition. The words and partwords were separated by an interstimulus interval of 500 ms. The order of words and partwords in the item pairs was counterbalanced, and the item pairs were presented in random order.

The five exposure blocks and 27 pairs of test items were saved as sound files and presented over headphones using a computer which also collected responses. Vroomen et al. (1998) used a loudspeaker to present their stimuli and collected responses via an answer sheet. A computer was used here to automate the presentation of the 32 different languages (16 x 2 stress conditions), and collection of responses using the computer keyboard. The use of headphones reduced the possibility of distraction from other participants or adjacent testing rooms.

2.1.4. Procedure

Participants were tested in groups of one, two, or three. Each participant was seated in front of a different computer. Verbal instructions were given by the experimenter, who interacted with the French participants in French and with the Dutch participants in English. To ensure that all of the directions were understood, written instructions were also provided in the participant’s native language.

The instructions were to listen to an artificial language, consisting of a sequence of syllables with no pauses. Participants were asked to simply pay attention to the exposure stream without reflecting too much on what they were listening to or trying to second-guess the purpose of the experiment. If they felt their attention start to wander they were asked to try to focus again on the task. Participants were made aware that there would be a test phase after the exposure phase, but that more precise information would be given at that stage.

Before the test phase, the participants were informed that the artificial language consisted of nonsense words that the experimenter had designated to be the “words” of the language. The purpose of the test phase was to find out if they

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

had learned anything about the words of the artificial language during the exposure phase. It was stressed that the words would not be real words in the participant's native language, and that any resemblance to known words would be coincidental. Participants listened to each word-partword pair and indicated by pressing the keys "1" or "2" whether the word of the language was the first or second member of the pair, respectively. They were told to guess if they were unsure.

2.2. Results

Percent correct scores were averaged for each participant, and then each group's mean score was derived from these values. An alpha rate of .05 was used for all statistical tests reported in this paper.

The mean percent correct responses for the French listeners in the no stress condition was 58.56% (*SE* = 2.56%). A one-sample *t*-test revealed that this was above chance (50%; $t(15) = 3.35$). The French listeners in the stress condition scored 82.18% (*SE* = 2.58%), also significantly above chance ($t(15) = 12.47$). For the Dutch listeners, the no stress group scored 59.49% (*SE* = 3.47%), above chance ($t(15) = 2.73$), and the stress group scored 72.45% (*SE* = 3.31%), also above chance ($t(15) = 6.78$). These results are presented in Figure 1.

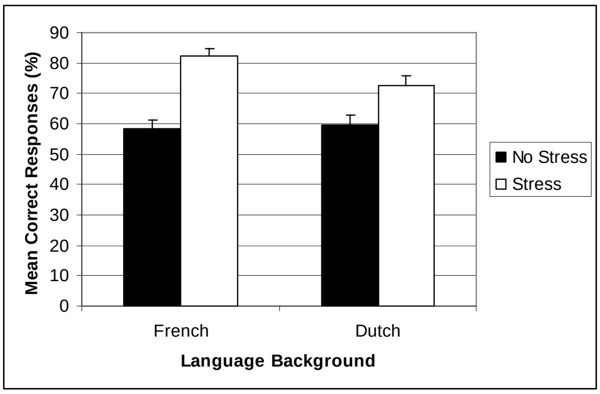

Figure 1: Mean percent correct responses in Experiment 1. Error bars represent standard errors of the mean.

To test for an effect of stress, the results were analysed using a 2 (language background) x 2 (language type) analysis of variance (ANOVA). Scores in the stress condition were significantly higher than in the no stress condition ($F(1,60) = 36.93$, $\eta_p^2 = .38$). There was no main effect of language background ($F(1,60) = 2.14$) and, importantly, there was no interaction between language background and language type ($F(1,60) = 3.13$) – both French and Dutch participants benefited from stress equally.

Recall that half of the participants in each language background group heard stimuli synthesized in either a French or Dutch voice. This *accent* factor was not included in the first ANOVA because it would result in a low cell size ($n = 8$). Nevertheless, it is important to test for any systematic effect of accent on the results, so an additional Language Background x Language Type x Accent ANOVA was conducted. There was no main effect of accent ($F(1,56) = 0.28$), and it did not interact with either language background or language type.

2.3. Discussion

In contrast to the results of Vroomen et al. (1998), French listeners in this study were able to benefit from a pitch accent to segment the words of an artificial language to the same extent as Dutch listeners. Did the increased complexity and variability of the language used here help French listeners to attend to the constant stress cue, or can other procedural factors account for these different results? For example, while both studies were conducted in group sessions, Vroomen et al. tested in groups of two to eight, whereas the groups had no more than three participants in Experiment 1. Furthermore, Vroomen et al. used a loudspeaker to present stimuli, whereas headphones were used here. These factors may have led to greater distraction in the testing area.

To test whether the increased complexity and variability of the language was responsible for the improvement in French listeners performance, Experiment 2 was a replication of the Vroomen et al. study using their own materials. To rule out those procedural differences outlined above, participants were again tested in groups of no more than three, and stimuli were presented over headphones, as in Experiment 1.

3. Experiment 2: A replication using Vroomen et al.'s (1998) materials

The original stimulus materials from Vroomen et al. (1998) were obtained from the authors. Given that vowel harmony was not of interest to this study, only the harmonious stimuli from Vroomen et al. were used here.

3.1. Method

3.1.1. *Design*

The design was the same as in Experiment 1.

3.1.2. *Participants*

Again, there were 16 participants in each Language Background x Language Type group ($N = 64$), drawn from the same populations as in Experiment 1. All but one of the French participants in the no stress group (13 females, 3 males; $M_{age} = 20.94$ yr, $SE = 1.22$), and all in the stress group (16 females; $M_{age} = 20.56$ yr, $SE = 0.77$) had learned some English at school. Around half had also learned Spanish, and a quarter German. All but one of the Dutch participants in the no stress group (14 females and 2 males; $M_{age} = 21.00$ yr, $SE = 0.61$), and all in the stress group (14 females and 2 males; $M_{age} = 21.69$ years, $SE = 0.76$) had learned English, and the majority had also studied French and/or German at school.

3.1.3. *Stimuli and apparatus*

The exact stimuli used by Vroomen et al. (1998), in the harmonious condition of their Experiment 3, were obtained from the authors for use in this experiment. The artificial language consisted of six trisyllabic words: /vomuvu/, /tokuvo/, /motamu/, /mymɛty/, /vykɛvɛ/, and /tykɛty/. The stimuli were prepared similarly to those in Experiment 1, with the exception that Vroomen et al. used the Dutch Spengi text-to-speech synthesizer. As such, in contrast to Experiment 1, all stimuli were synthesized with a Dutch accent.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after full paper review

Items for the test phase consisted of words and two types of foils. The *nonwords* /vutato/, /kutavo/, and /vytymɛ/ consisted of syllables from the language in an order that would not have occurred in the exposure stream. The *partwords* /vomuto/, /kɛmɛty/, and /vykɛmy/, on the other hand, contained either the first two or the last two syllables of one of the words of the language. Note that these partwords were not constructed in the same way as those for Experiment 1. Each word was paired exhaustively with each foil to obtain 36 pairs of items, with a counterbalanced order of words and foils. Four practice trials were also included. In this experiment the test items were concatenated into a long sound file. Each stimulus pair was followed by 4 s of silence, a warning tone of 500 ms, and another 1 s of silence.

The soundfiles were played over headphones via computer from an audio playback program. Following Vroomen et al., the responses were obtained via an answer sheet.

3.1.4. Procedure

The instructions were the same as in Experiment 1, with the exception that participants were instructed to circle either “1” or “2” on the answer sheet, rather than pressing a key on the keyboard, to indicate whether the first or second member of the pair was the word of the language.

3.2. Results

The mean percent correct responses for the French listeners in the no stress condition was 59.90% (SE = 3.85%), which was significantly above chance ($t(15) = 2.57$), and those in the stress condition scored 72.22% (SE = 3.48%), also above chance ($t(15) = 6.39$). For the Dutch listeners, the no stress group scored 61.11% (SE = 2.98%), above chance ($t(15) = 3.73$), and the stress group scored 69.10% (SE = 2.88%), also above chance ($t(15) = 6.63$). The mean percent correct responses are presented in Figure 2.

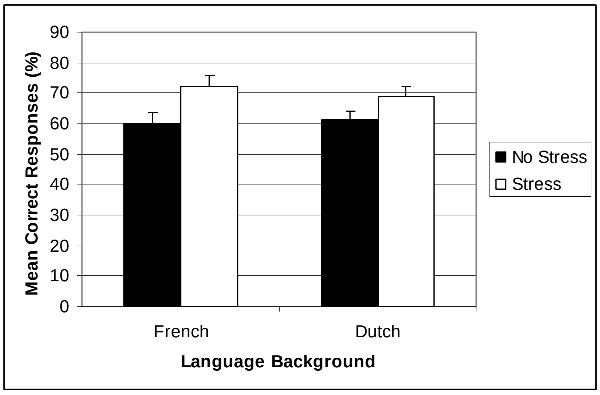

Figure 2: Mean percent correct responses in Experiment 2. Error bars represent standard errors of the mean.

To test for an effect of stress, the results were analysed using a 2 (language background) x 2 (language type) ANOVA. As in Experiment 1, the effect of language type was significant ($F(1,60) = 9.36$, $\eta_p^2 = .14$), there was no main effect of language background ($F(1,60) = 0.08$), and no interaction between language background and language type ($F(1,60) = 0.43$). Again, both French and Dutch participants benefited from stress equally.

3.3. Discussion

The results of Experiment 2 mirror those of Experiment 1. Therefore, the complexity and variability of the language in Experiment 1 was not responsible for the apparent improvement in performance of French listeners relative to Vroomen et al. (1998).

What might explain the differing results between Experiment 2 and Vroomen et al.’s (1998) experiment? Procedural differences are one possibility. As outlined above, Vroomen et al. conducted their experiment in larger groups and with a single loudspeaker instead of headphones. It is possible that these conditions provided greater distraction for the participants and that is why a null result was obtained for the French participants. However, these factors should apply equally to the French and Dutch participants, so they are unlikely to provide a full explanation for the discrepancy.

Another possibility is that the smaller sample size used in Vroomen et al.’s study (no stress n = 10; stress n = 11) lacked sufficient power to detect a significant difference. However, in that study both groups obtained 58% correct, so there was not even a tendency towards superior performance for the stress group.

Vroomen et al.’s (1998) study was conducted 10 years before the present study. It is possible that the students participating in the present study had more experience with English than those in the Vroomen et al. study, and that they were able to use this knowledge to aid their performance in the task. Indeed, the nearly all of the participants in Experiments 1 and 2 had studied English at school, but Vroomen et al. did not report the language experience of their participants so this possible explanation must remain highly speculative. Other cohort differences, such as local dialect (Paris vs. Burgundy), may also have played a role.

4. General discussion

It is clear from the results of this study that, at least under certain conditions, both French and Dutch listeners can benefit from the Vroomen et al. (1998) pitch accent to aid segmentation of an artificial language. This contrasts with French listeners’ poor performance on a task requiring discrimination of words differing in stress placement (Dupoux et al., 1997).

It was argued in Dupoux et al. (1997) that French listeners were less likely to detect stress differences in more complex tasks. For this reason it was suggested that the increased complexity of the artificial language in Experiment 1, relative to that of Vroomen et al. (1998), might have led to even poorer performance of French listeners, relative to Dutch listeners, who can use their native MSS to aid performance. The fact that French listeners benefited from the stress cue in both Experiments 1 and 2 suggests that language complexity did not play a role in performance here. Nevertheless, their good performance could lead to the prediction that French listeners would also benefit from stress when learning a stress-timed natural language, such as English. However, it should also be noted that stress in English and Dutch is based on syllable weight (i.e., strong vs. weak syllables), which was controlled, rather than manipulated, in these artificial language studies. To allow more valid generalisations from artificial

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after full paper review

language studies to second language learning, future studies on lexical stress should investigate French listener's sensitivity to syllable-weight variation.

An assumption made by Vroomen et al. (1998), and continued in the experiments presented here, is that the pitch accent cue would be perceived as stress by the listeners. The pitch accent was argued to be appropriate for Finnish and Dutch, but it was not based directly on either language (see Suomi, Toivanen, & Ylitalo, 2003, for comments that Finnish word stress is not always signalled by an F0 rise). Perhaps the rise-fall pattern across all three syllables of each word in Vroomen et al. (and in Experiment 2), and across all three or three of the four syllables of each word in Experiment 1 provided a simple perceptual, that is, non-linguistic grouping. We are currently testing whether a simple visual cue, such as a tapping finger, that is visually aligned with a word onset will also improve listeners' word segmentation abilities in an artificial language. Alternatively, the grouping provided by the pitch accent may have been linguistic. That is, participants may have heard each word as its own intonational phrase, which would correspond somewhat to listening to a list of words spoken out loud.

Questions over the cohort, language complexity, and the appropriateness of the stress cue make it difficult to make specific conclusions about the effect of native-language speech-segmentation strategies on word segmentation in an unknown foreign language. Certainly, Dutch participants unambiguously benefited from stress here, and in Vroomen et al. (1998). This is consistent with the idea that they were able to use their MSS to aid segmentation. While it is unclear exactly by which mechanism French listeners were able to benefit from the pitch accent, one thing is clear – French listeners were not restricted to their native-language word-segmentation strategies when confronted with the artificial languages used here. Therefore, our perceptual systems may be more malleable than suggested by Dupoux et al. (1997). The extent of this malleability could be tested in future research by comparing monolingual French, monolingual English, and French-English bilingual participants on their segmentation of an artificial language exhibiting English stress.

5. Acknowledgements

The author thanks his colleagues Anne Cutler and Pierre Perruchet for their input and support for this project and to Cathi Best, Anne Cutler, and an anonymous reviewer, for helpful comments on an earlier draft of this paper. Michael Tyler was supported by a postdoctoral fellowship from the Conseil Régional de Bourgogne, a visiting fellowship from the Max Planck Institute for Psycholinguistics, a Spinoza Grant awarded by the NWO to A. Cutler, and NIHCD grant DC00403 (PI: C. Best). Thanks to Jyrki Tuomainen, Jean Vroomen, and Beatrice de Gelder for kindly providing their stimuli for Experiment 2 of this study.

6. References

Cutler, A., & Norris, D. (1988). The role of the strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 113-121.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language, 36*, 406-421.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In H. T. Bunell & W. Isardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393-1396).

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604-607.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 56A*, 97-115.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development* (6th ed., pp. 58-108). New York: Wiley.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics, 31*, 113-138.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706-716.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition, 24*, 744-755.