Author(s): Dawson, Caitlin; Aalto, Daniel; Šimko, Juraj; Putkinen, Vesa; Tervaniemi, Mari; Vainio, Martti

Title: Quantity language speakers show enhanced subcortical processing

Year: 2016

Version:

Please cite the original version:

Dawson, C., Aalto, D., Šimko, J., Putkinen, V., Tervaniemi, M., & Vainio, M. (2016). Quantity language speakers show enhanced subcortical processing. Biological Psychology, 118(July), 169-175. https://doi.org/10.1016/j.biopsycho.2016.06.003

Contents lists available at ScienceDirect

Biological Psychology

BIOLOGICAL
PSYCHOLOGY

Quantity language speakers show enhanced subcortical processing

Caitlin Dawson [a,b,*], Daniel Aalto [d,e], Juraj Šimko [a], Vesa Putkinen [b,c], Mari Tervaniemi [b,f], Martti Vainio [a]

[a] Phonetics and Speech Synthesis Research Group, Institute of Behavioural Sciences, University of Helsinki, Finland
[b] Cognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, Finland
[c] Department of Music, University of Jyväskylä, Finland
[d] Institute for Reconstructive Sciences in Medicine (iRSM), Canada
[e] Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada
[f] Cicero Learning, University of Helsinki, Finland

ABSTRACT

The complex auditory brainstem response (cABR) can reflect language-based plasticity in subcortical stages of auditory processing. It is sensitive to differences between language groups as well as stimulus properties, e.g. intensity or frequency. It is also sensitive to the synchronicity of the neural population stimulated by sound, which results in increased amplitude of wave V.

Finnish is a full-fledged quantity language, in which word meaning is dependent upon duration of the vowels and consonants. Previous studies have shown that Finnish speakers have enhanced behavioural sound duration discrimination ability and larger cortical mismatch negativity (MMN) to duration change compared to German and French speakers.

The next step is to find out whether these enhanced duration discrimination abilities of quantity language speakers originate at the brainstem level. Since German has a complementary quantity contrast which restricts the possible patterns of short and long vowels and consonants, the current experiment compared cABR between nonmusician Finnish and German native speakers using seven short complex stimuli. Finnish speakers had a larger cABR peak amplitude than German speakers, while the peak onset latency was only affected by stimulus intensity and spectral band. The results suggest that early cABR responses are better synchronised for Finns, which could underpin the enhanced duration sensitivity of quantity language speakers.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In the past ten years, it has become apparent that the auditory brainstem, previously thought to be a byway tasked with preserving and transmitting temporal information rather than transforming it, is also plastic to environmental and perceptual experiences. Beginning with studies into the enhancement of pitch representations in people who speak tonal languages (Krishnan, Xu, Gandour, & Cariani, 2005) and moving to enhancements in phase-locking and pitch representation in both musical and speech sounds from musical training (Lee, Skoe, Kraus, & Ashley, 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007), it is clear that the auditory brainstem plays a crucial role in the pre-processing of incoming auditory

information. The mechanisms of these plastic changes are not completely understood, but are thought to represent synaptic plasticity in the form of long-term potentiation (and depression) and intrinsic plasticity as changes in neuronal excitability (Tzounopoulos & Kraus, 2009). The complex process of timing regulation in the auditory brainstem arises from the interaction of excitatory and inhibitory input as well as the physiological arrangement of neurons of different cell types and responses that may be regulated in a top-down process from experience-dependent changes in the cortex (Winer & Schreiner, 2005).

The complex auditory brainstem response (cABR) is an early onset response that is sensitive to the properties of auditory stimuli, particularly intensity and frequency: decreasing the intensity or frequency delays the onset peak (Neely, Norton, Gorga, & Jesteadt, 1988). It is also sensitive to the synchronicity of the neural population stimulated by sound; more temporally precise firing of this population results in more aligned peaks and thus, increased peak amplitude of wave V (Rønne, Dau, Harte, & Elberling, 2012).

* Corresponding author at: Cognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, Finland.
E-mail address: caitlin.dawson@helsinki.fi (C. Dawson).

Scalp-recorded cABRs putatively originate from the cochlear nuclei, lateral lemniscata, and inferior colliculi (Chandrasekaran & Kraus, 2009) where there is a convergence of multiple inhibitory and excitatory inputs.

Language background has been shown to influence subcortical responses as enhanced frequency following response (FFR) to Mandarin tones for Mandarin speakers compared to English speakers, evidence for language-based plasticity for pitch at the brainstem level (Krishnan et al., 2005). While Mandarin is a tonal language that encodes phonological contrast by pitch changes, Finnish is a fully-fledged quantity language that uses segmental duration contrastively. An example is the opposition of *tuli*, 'fire', *tuuli,* 'wind', and *tulli*, 'customs,' where the phonological length (duration) of the vowel/u/or the consonant/l/determines the meaning of the word. In spoken Finnish, the difference is also partially marked by phonetic adjustments to durations of neighboring segments and by pitch contours (Lehtonen, 1970; O'Dell, 2003; Vainio, Järvikivi, Aalto, & Suni, 2010). German utilizes a quantity system restricted to a subset of vowels and contrast is generally encoded by vowel quality (tense and lax vowels) (Wiese, 1996).

It has been shown that Finnish speakers have an enhanced behavioural sound duration discrimination ability as well as larger cortical MMN (mismatch negativity) to duration change compared to German speakers (Tervaniemi et al., 2006), where German represents a language without a full quantity system. Finnish speakers also show a larger cortical MMN to duration deviants compared to French nonmusicians (Marie, Kujala, & Besson, 2012). Thus, the impact of language on duration processing happens at an early, pre-attentive level and generalizes to nonspeech sounds.

The goal of the present work is to further investigate the neural underpinnings of language based adaptations. The next step, then, is to show these language-based adaptations of Finnish speakers at the brainstem level. Since studies of Mandarin speakers have traced language-based enhancements in pitch processing from behaviour, through cortical EEG, to differences visible in the auditory brainstem, it is reasonable to expect that, similarly, the observed behavioural and cortical enhancements in duration processing shown by Finnish speakers also originate in the auditory brainstem. Specifically, Finnish speakers' cABR responses would show a higher amplitude that reflects enhanced temporal precision in the brainstem.

2. Methods

2.1. Participants

30 participants participated in the experiment and were recruited in two groups: 15 German speakers (8 females), and 15 Finnish speakers (9 females). Data from 21 participants were analyzed: 10 native German speakers (mean age 30.2 years, 7 females) and 11 native Finnish speakers (mean age 25.7 years, 7 females), all right-handed. Native speakers were defined as using only the native language at home for the first 15 years of life and self-reported not being bilingual. Data from four participants were excluded from analysis at the beginning due to mild to moderate hearing loss discovered during audiometry screening. Data from five more participants were excluded during preprocessing due to retaining less than 35% usable data (see Analysis for exclusion criteria).

One German-speaking participant was excluded in post hoc analyses after the discovery of extra music training in the questionnaire. Results are reported excluding this participant.

Participants were recruited by student email lists within the University of Helsinki, from local Facebook groups for German expats or researchers, and word of mouth. They had no more than 2 years' formal music training and no musical hobbies, and used only the native language in the first 15 years of everyday life. They were screened for normal hearing ($\leq$20 dB). The experiment was conducted according to the ethical guidelines of the Declaration of Helsinki and the study protocol was approved by the Committee for ethical review in the humanities and social and behavioural sciences at the University of Helsinki. Participants gave written informed consent before the experiment and were compensated for their time.

2.2. Stimuli

A set of seven complex stimuli was used together with their polarized counterparts. The signals were created so that different pitch patterns (static, rise, fall) would have as equal loudness as possible. This was done in three steps. First, a sawtooth wave of desired pitch contour was created: static 200 Hz (stimulus 1), 150 Hz (stimuli 2, 6, and 7), 112.5 Hz (stimulus 3), an exponential rise from 150 Hz to 200 Hz (stimulus 4), and an exponential fall from 150 Hz to 112.5 Hz (stimulus 5). Second, the sawtooth signals were narrow band filtered using a fourth order polynomial gammatone filter of centre frequency 3141.56 Hz (stimuli 1–5, and 7) and 1570.78 Hz (stimulus 6; see Cooke 1993). Finally, the average intensities of the stimuli were normalized and stimulus 7 was scaled 10 dB weaker. All the stimuli had equal duration of 80 ms. The stimuli were generated in Matlab with 16 kHz sampling frequency and presented in four blocks in random order with a fixed 69.56 ms silent interval between the stimuli. These stimuli were chosen in order to obtain good quality data that could be compared to previously observed stimulus effects as well as to rule out stimulus as a confounding cause of differences between language groups. These frequencies were chosen to fall within the fundamental frequency range of typical conversational speech and to resemble the narrowband, amplitude modulated sounds that are important for speech processing (Rees & Langner, 2005).

There is an ongoing investigation and debate in the field of auditory sciences regarding the use of synthetic stimuli. The earliest synthesized sounds used in auditory experiments were simple sine waves which could not adequately show how complex and linguistic sounds are processed. The linguistic content and naturalness of sounds may be abstracted in higher cortical areas in order to assess subcortical input according to behavioural goals (Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006). Some recent studies suggest that the subcortical sound encoding can even be modulated on-line based on conscious switching of task goals (Coffey, Colagrosso, Lehmann, Schönwiesner, & Zatorre, 2016). Clearly, there are many questions to be asked about the extent and quality of top-down modulation of auditory input. Moreover, clicks and tones have different effects on the auditory brainstem responses (Picton, 2011) and the structure of the response depends on the stimulus. However, complex synthetic sounds work well in the present context of non-linguistic, basic peripheral sound processing plasticity that applies to many kinds of sounds. The current experimental questions relate to long term linguistically induced plasticity rather than to the linguistic content or subjective naturalness of the stimuli themselves.

Here, the decision to use artificial speech-like stimuli rather than speech sounds serves the purpose of allowing strict control over the stimulus properties: first, in order to normalize the loudness; second, to create narrowband sounds that would stimulate a specific portion of the auditory pathway neurons; and last, to avoid a language bias inherent in natural speech sounds. Evidence from animal research shows that synthesized amplitude modulated sounds resemble natural sounds in the auditory brainstem (Joris, Schreiner, & Rees, 2004). These sounds are part of a larger ongoing project investigating sound feature processing in different languages; for a thorough explanation of the stimuli, see Aalto et al.

(2013). These synthesized complex sounds represent a pioneering step toward full control of natural-like synthesized sounds.

2.3. Procedure

Participants took a short audiometry test, around 10 min long, to ascertain hearing thresholds for each ear at frequencies ranging from 125 Hz to 8 kHz. The range from 1 kHz–4 kHz was of particular interest, and those participants who responded above 20 dB in one or both ears inside this range were excluded from analysis.

Participants watched a self-chosen film silently with subtitles while wearing headphones for passive presentation of the stimuli. They were instructed not to pay attention to the auditory stimuli but to attend to the film, since relaxing tasks such as reading and film watching are thought not to affect the fast ABR responses (Skoe & Kraus, 2010) but serve to regulate participants' long-term arousal, keeping them relaxed but alert, and preventing sleeping, movements, and alpha waves which disrupt recording quality. The recording sessions lasted around 70 min consisting of 5 blocks of 14 min each. For the cABR recording, the arrangement was a 10-electrode vertical montage along the midline channels (Skoe & Kraus, 2010) with 2 mastoid references and 4 electrooculography (EOG) electrodes, attached with conductive gel and surgical tape. The recording was captured with the Biosemi ActiveTwo system at a sampling rate of 16384 Hz in an electrically shielded room with stimuli presented binaurally using shielded circumaural Sennheiser HD250 linear II headphones.

Participants also answered a 2 page survey of their linguistic and musical background, including all other languages that they have studied and spoken, how often they listen to music, and whether they like music.

2.4. Analysis

The cABR recordings were offline bandpass filtered at 80–4000 Hz with a baseline level 50 ms before stimulus onset and re-referenced to the average of the mastoids. With 7 stimuli and 28,000 total trials per participant, there were 4000 epochs of 50 ms prestimulus baseline and 180 ms post stimulus in order to capture the full length of the stimulus, inter-stimulus silence, and the onset of the next stimulus. Epochs with amplitudes greater than $\pm 30\,\mu V$ in any of the 10 channels were rejected and participants with greater than 65% of epochs rejected were excluded from analysis. Otherwise, the 10 channels were averaged together; source analysis is not needed since in ABR research, the source is assumed to be the auditory brainstem. Epochs were averaged separately for each stimulus type, with polarities combined, across all 10 channels, giving 7 different cABRs per participant. The average contours were visually inspected and all the responses had the same overall shape. Peak sizes and peak locations aligned according to stimuli were extracted by a custom thresholding algorithm to detect amplitude and latency of onset peaks as a percentage of total peak size.

2.5. Statistical analyses

Statistical analyses used linear mixed effects models with fixed effects of stimulus, language, epoch number, and peak order, and with participant as a random effect (Bates, Maechler, & Bolker, 2012). Normality of the measured data and the residual errors of the models were checked using quantile plots. Amplitude was logarithmically transformed to partially solve the problem of non-normality.

The primary analysis is done on wave V, which represents the first major peak related to auditory processing. A secondary analysis was performed on the successive peaks A, C, and D comparing

between language groups and corrected for multiple comparisons; however, interpretation of these later peaks is unclear and depend in part on the stimulus features. A 30 ms pre-stimulus period was also analyzed beforehand in order to ensure that the language group differences were only found within the stimulus window. There was no effect of language on the amplitude of the pre-stimulus period (t = 0.55, df = 18.04, p = 0.59).

3. Results

3.1. Effects of stimuli

The data were first analyzed using a linear mixed effects model with stimulus and language as fixed effects and participant as a random effect. Extended waveforms for individual channels appear in Fig. 1, grand-averaged over subjects and stimuli, showing the language group difference. The main focus for amplitude comparisons was wave V because it is the first prominent peak and has been typically used for groups-based comparisons of peripheral auditory processes (Skoe & Kraus, 2010; Johnson, Nicol, Zecker, & Kraus, 2008; Musacchia, Strait, & Kraus, 2008; Wible, Nicol, & Kraus, 2005).

For wave V peak in the mixed effects model, there was a main effect of stimulus on latency. The weaker stimulus and the stimulus with 1571 Hz centre frequency resulted in a delayed response (Table 1).

There was a significant effect of peak order on both amplitude and latency, with the first peak coming stronger (t = 9.95, df = 1690.80, p < 2 × 10⁻¹⁶) and earlier (t = −7.38, df = 1776, p = 2.38 × 10⁻¹³). There was an interaction of peak order and stimulus for amplitude (Table 2) but not for latency. This pattern of high initial peaks varying by frequency of the stimulus likely reflects the response of a population of onset-sensitive neurons.

3.2. Effects of language group

There was a main effect of language on peak amplitude of wave V response (Fig. 2) with Finnish speakers showing larger amplitude (t = −2.60, df = 17, p = 0.019). This language difference was present for all stimuli (Fig. 3). Finnish speakers also had larger peak amplitude across the whole signal (t = −2.60, df = 17, p = 0.019) and for the first peak only (t = −2.60, df = 17, p = 0.019).

There were no significant interactions of peak order and stimulus for latency. There were no significant interactions between stimulus and language for either amplitude of the signal, nor for latency or amplitude of wave V.

The number of epochs collected has an important impact on the signal-to-noise ratio and can distort the effects. The standard deviation of the averaged signals depends proportionally on the number of epochs. In this experiment, the German speakers had a significantly higher number of usable epochs (t = −4.48, df = 112.32, p = 1.79e-05). In order to determine the contribution of this difference, the effect of epoch number on peak amplitude was modeled by testing the algorithm on the 5 ms pre-stimulus silent region. When this term was added to the main mixed effects model as a factor, it showed that the number of epochs collected did not significantly contribute to the amplitude difference between language groups (t = 0.36, df = 18.10, p = 0.72).

Post-hoc analyses were run both including and excluding the German-speaking participant with more musical experience. In all cases, excluding this participant either did not change the results or strengthened the language group difference, which is in line with the prediction that language background and musical training change the auditory brainstem in similar ways.
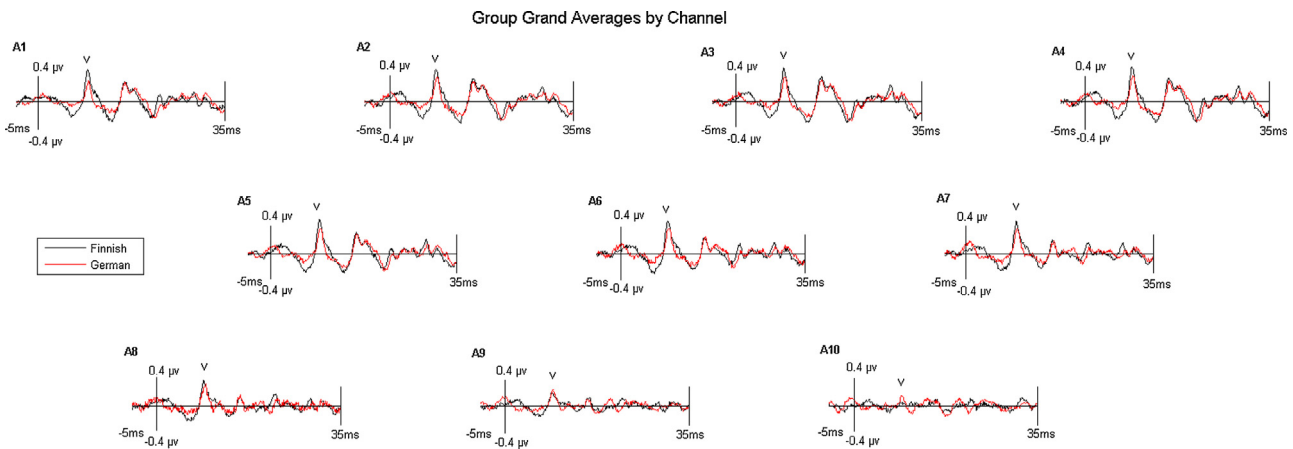
Fig. 1. Grand averages over stimuli for each language group, per channel. Channels were arranged in a vertical montage along the midline, with A1 at the forehead and A10 above the inion. The black (dark) line represents the Finnish group and the red (light) line represents the German group. Peaks of wave V are labelled with a "V". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Stimulus effect on latency in ms.

Fixed effect	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.052×10^{-3}	1.56×10^{-3}	18.16	5.18	6.19×10^{-5}***
language	8.033×10^{-5}	9.97×10^{-5}	18.01	0.81	0.43
stimc	9.79×10^{-4}	1.02×10^{-4}	120.0	9.61	2.22×10^{-16}***
stimF	2.96×10^{-6}	1.02×10^{-4}	120.0	0.03	0.98
stimH	-6.68×10^{-5}	1.02×10^{-4}	120.0	-0.66	0.51
stimL	4.050×10^{-5}	1.02×10^{-4}	120.0	0.40	0.69
stimR	-4.66×10^{-5}	1.02×10^{-4}	120.0	-0.46	0.65
stimw	5.78×10^{-4}	1.02×10^{-4}	120.0	5.68	9.79×10^{-8}***
epoch	7.02×10^{-5}	1.94×10^{-4}	18.10	0.36	0.72

Linear mixed effects model with fixed effects of language, stimulus, and epoch and random effect of participant, for latency (in ms) of wave V. Abbreviations: stimc, stimulus with different filter centre; stimF, falling; stimH, high; stimL, low; stimR, rising; stimw, weak (lower intensity).

Table 2
Interaction of stimulus and peak order on amplitude.

Fixed Effect	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-8.90×10^{-1}	19.63	18.1	-0.45	0.66
language	-3.38×10^{-1}	0.13	17	-2.60	0.019*
stimc	-7.55×10^{-1}	0.038	168.5	-19.68	$<2 \times 10^{-16}$***
stimF	1.086×10^{-1}	0.039	168.5	2.77	0.0057**
stimH	-2.74×10^{-1}	0.036	168.5	-7.70	2.35×10^{-14}***
stimL	3.44×10^{-1}	0.042	168.5	8.24	4.44×10^{-16}***
stimR	-1.01×10^{-1}	0.037	168.5	-2.74	0.0063**
stimw	-3.59×10^{-1}	0.038	168.5	-9.37	$<2 \times 10^{-16}$***
order	1.63×10^{-1}	0.094	168.5	1.74	0.082
epoch	-3.73×10^{-2}	0.25	168.5	-0.15	0.88
stimc × order	6.24×10^{-1}	0.13	168.5	4.71	2.71×10^{-6}***
stimF × order	-8.08×10^{-3}	0.13	168.5	-0.061	0.95
stimH × order	3.11×10^{-1}	0.13	168.5	2.36	0.019*
stimL × order	-1.30×10^{-1}	0.13	168.5	-0.98	0.33
stimR × order	2.78×10^{-1}	0.13	168.5	2.099	0.036*
stimw × order	2.67×10^{-1}	0.13	168.5	2.010	0.045*

Linear mixed effects model with fixed effects of language, epoch, interaction of stimulus and peak order, and random effect of participant, for amplitude (in microvolts) of the first peak. Abbreviations: stimc, stimulus with different filter centre; stimF, falling; stimH, high; stimL, low; stimR, rising;

3.3. Secondary analysis

A separate, secondary analysis was performed on the successive negative peaks A, C, and D. For this analysis, the absolute values of the negative amplitudes were log-transformed and p-values corrected for multiple comparisons. Differences between language groups generally did not reach significance. The language-related difference in amplitude was not significant for waves A and C (t = −0.57, df = 18.02, p = 1; t = −1.077; df = 18.03; p = 1; respectively). For wave D, the amplitude difference was significant only for the low stimulus (t = −3.89 df = 120, p = 0.0024). The difference in latency of waves A and D was not significant (t = 0.48, df = 18.01, p = 1; t = −1.21, df = 18.01, p > 1, respectively), but the latency difference of wave C was significant for the low stimulus (t = 3.533, df = 120.6, p = 0.008162).

Interactions of peak order and stimulus for both amplitude and latency did not reach significance for the later negative peaks, and the interaction between peak order and stimulus for latency only reached significance for the low stimulus (t = −3.28, df = 1780, p = 0.015). There was an effect of peak order for the interaction
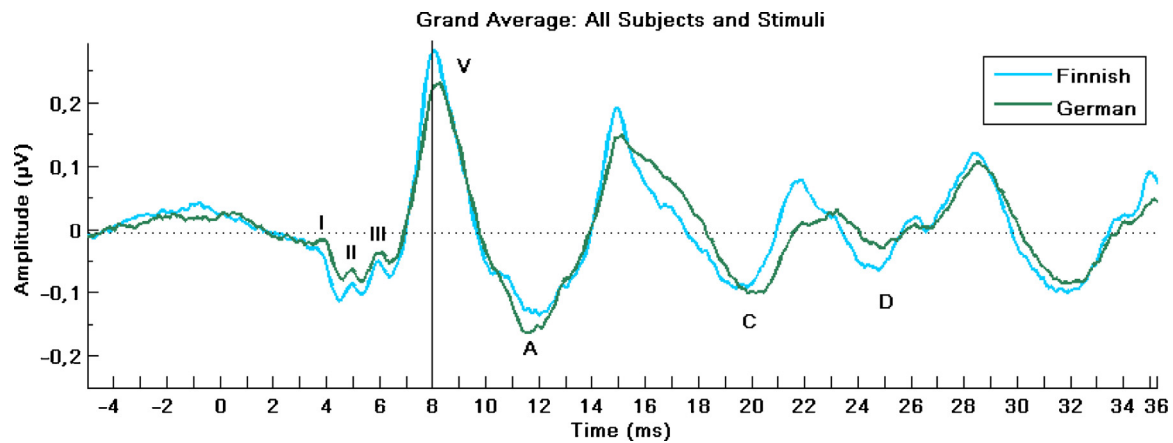
Fig. 2. Grand averages across stimuli, by language group, with wave peak labels. Blue (light) line represents Finnish speakers' waveform; green (dark) line represents German speakers' waveform. The vertical line at around 8 ms shows the peak amplitude. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
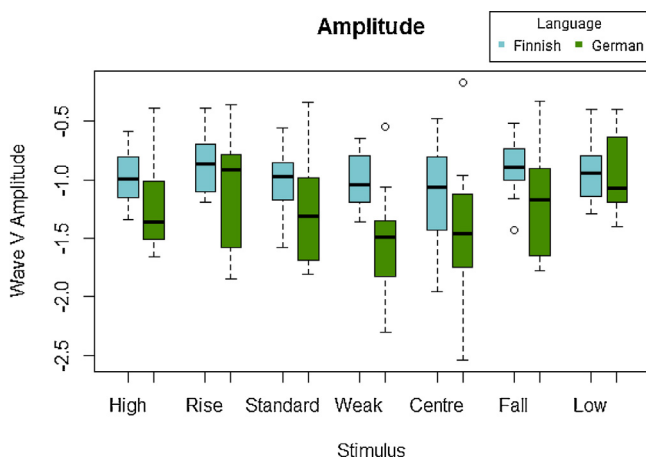

Fig. 3. Logarithm of wave V amplitude (in microvolts) showing higher means across stimuli for Finnish group (blue/light) compared to German group (green/dark). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between stimulus and language for latency (t = −3.79, df = 1780, p = 0.0022). For amplitude, there was an interaction between stimulus and language for the weak stimulus (t = −3.8, df = 1772, p = 0.0021) and for peak order (t = −6.42, df = 1772, p = 2.52 × 10⁻⁹).

Waves I–III, while visible in the grand-averaged waveform, are extremely low-amplitude and cannot be detected for individual subjects and stimuli with the current number of epochs due to a low signal-to-noise ratio. Additionally, the vertical electrode montage (as opposed to a horizontal montage) was chosen specifically because it has been shown to increase visibility of wave V and diminish the amplitudes of the earlier waves. In clinical settings focusing on diagnostic criteria using latencies of the earlier waves and equal representation of waves, a horizontal or ipsilateral electrode montage is recommended, which enhances visibility of these earlier waves but consequentially reduces wave V (Katbamna, Metz, Bennett, & Dokler, 1996; King & Sininger, 1992; Stuart, Yang, & Botea, 1996). Wave IV is often subsumed by the much more prominent wave V and is rarely studied independently. The recording techniques used in this experiment did not give data that clearly show these earlier waves that could be used in meaningful analysis; therefore, waves I–IV are not analyzed here.

4. Discussion

The aim of this study was to determine whether the enhanced duration processing of Finnish speakers could be observed at the brainstem level, which is shown by a greater amplitude in the cABR compared to German speakers' responses. Indeed, this was the case: Finnish speakers showed a significantly larger amplitude cABR to all 7 stimuli in both wave V and across the signal, which reflects the enhanced response synchrony of subcortical neural populations.

The design of the current study with a range of stimuli allows analysis of the effects of different stimulus properties (frequency, intensity, spectral band, and static/dynamic), as well as the effects of language background, on wave V response latency and amplitude. Stimuli with a lower intensity level or carrier frequency centre delayed the response, which is similar to the results found for click responses and tone bursts (Eggermont & Don, 1980; Neely et al., 1988).

Evidence from click-response literature also suggests that the greater amplitude for Finnish speakers in the wave V response and across the signal could reflect a higher degree of precision in alignment of the onset response latencies due to populations of neurons firing in better synchrony. The additive effect of aligned peaks represents better encoding of timing in the inferior colliculus (Elberling, Callo, & Don, 2010). An alternative explanation comes from voice onset time (VOT) literature, which suggests that the difference in processing between Finnish and German speakers could be due to differences in segmental distribution of voiced and unvoiced stops and the more variable voice onset time in German language (Suomi, 1980; Braunschweiler, 1997). This provides an interesting hypothesis that could be tested by comparing a Finnish speaking group to a language group with similar voice onset time but without quantity, such as Russian. However, this distinction is outside the focus of the current research, which focuses on addressing the hypothesis from Tervaniemi et al. (2006), whether the cortical and behavioural differences they observed between Finnish and German language groups could originate in the brainstem.

The number of epochs collected influences average peak amplitude: averaging more epochs attenuates noise and makes it less likely that the peaks are caused by noise rather than signal. Because there were fewer epochs collected in the Finnish compared to the German participants, it was necessary to create a model of the effect of epoch on noise level in order to rule out a differing signal-to-noise ratio as a cause of the difference in amplitude between language groups. The epoch model was created by running the same peak-

picking algorithm on the 5 ms prior to stimulus onset, where it should detect no periodic peaks as seen in the portion of the signal during stimulus presentation. This gives the effect of epoch number on noise during silence, which was then added as a fixed effect term into the mixed effects model, showing that the difference in epoch number did not contribute to the main effect of amplitude difference between language groups.

The language effect for peak amplitude was strengthened after post hoc removal of a German-speaking participant who was found to have had three years of musical training and semi-regular casual music practice through childhood and adolescence. Musical training is known to affect the cABR: musicians show a faster onset peak and a higher amplitude to periodic stimuli (Kraus & Chandrasekaran, 2010; Musacchia et al., 2008; Strait & Kraus, 2014) that is modulated by extent of musical training in general but also appears in adults with childhood musical experience (Skoe & Kraus, 2012). Since post-hoc analyses showed that models excluding this participant improved the effect sizes and significance levels, it is reasonable to conclude that the current results stand for strict non-musicians only. The effects of different kinds and intensities of musical experiences on these neural enhancements are not completely understood, and more research is necessary to disentangle possible compounding effects from both language and music.

There were few strong effects in the later negative peaks (A, C, D), and the interpretation of the secondary analysis is less clear. The later peaks are the result of subsequent periodic pulses in the stimuli and may interact with each other. Therefore, it is less effective to analyze them individually and the main analysis here focuses on wave V.

Overall, these results suggest that Finnish speakers' language-based enhancements in duration processing are similar to brainstem enhancements for pitch processing in Mandarin speakers. Both pitch and duration processing differences can be observed in behavioural and cortical responses as well as in the auditory brainstem, which suggests that the auditory brainstem is plastic to several different properties of auditory experience. It also suggests that it is not only tone, but other acoustic properties of native languages that play a crucial role in shaping the auditory system.

5. Conclusions

Quantity language speakers use duration cues to differentiate between word meanings. This, in turn, enhances the duration discrimination of the speakers of these languages. Thus, during language acquisition, the brainstem (in parallel to other brain structures e.g., at the cortical level) adapts to the specific needs imposed by the particular social and sound environments.

Language background of the participants had a significant effect on peak amplitude both for wave V and also across the entire signal, but not on latency, which suggests that the source of the effect lies in differences in alignment precision of the brainstem response: the Finnish participants have more precise responses to the stimuli, which contributes to a larger peak amplitude without affecting peak latency. Stimulus properties (frequency, intensity, spectral band of filter) modulated peak latency but not amplitude, and the difference in number of epochs collected did not explain the group difference.

Having a precise representation of both frequency and duration is crucial for discriminating phonemes in Finnish. The language group differences shown here provide evidence that the auditory brainstem is plastic to native language environment not only for pitch but also duration.

References

Aalto, D., Simko, J., & Vainio, M. (2013). Language background affects the strength of the pitch bias in a duration discrimination task. *Proceedings of the interspeech 2013: 14th annual conference of the international speech communication association.*

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: linear mixed-effects models using S4 classes. *R Package Version 0*, 999999–0.

Braunschweiler, N. (1997). Integrated cues of voicing and vowel length in German: a production study. *Language and Speech, 40*(4), 353–376.

Chandrasekaran, B., & Kraus, N. (2009). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology, 47*(2), 236–246. http://dx.doi.org/10.1111/j.1469-8986.2009.00928.x

Coffey, E. B., Colagrosso, E. M., Lehmann, A., Schönwiesner, M., & Zatorre, R. J. (2016). Individual differences in the frequency-following response: relation to pitch perception. *PloS One, 11*(3) http://dx.doi.org/10.1371/journal.pone.0152374, e0152374

Cooke, M. (1993). *Modeling auditory processing and organization.* New York, NY: Cambridge University Press.

Eggermont, J., & Don, M. (1980). Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking: II. Effect of click intensity. *Journal of the Acoustical Society of America, 68*, 1671–1675.

Elberling, C., Callo, J., & Don, M. (2010). Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation. *Journal of the Acoustical Society of America, 128*(1), 215. http://dx.doi.org/10.1121/1.3397640

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *The Journal of Neuroscience, 28*(15), 4000–4007.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*(2), 541–577.

Katbamna, B., Metz, D. A., Bennett, S. L., & Dokler, P. A. (1996). Effects of electrode montage on the spectral composition of the infant auditory brainstem response. *Journal of the American Academy of Audiology, 7*, 269–273.

King, A. J., & Sininger, Y. S. (1992). Electrode configuration for auditory brainstem response audiometry. *American Journal of Audiology, 1*(2), 63–67.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*(8), 599–605.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161–168. http://dx.doi.org/10.1016/j.cogbrainres.2005.05.004

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience, 29*(18), 5832–5840.

Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish. Studia Philological Jyväskyläensia VI.* University of Jyväskylä.

Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex, 48*(4), 447–457.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241*(1–2), 34–42. http://dx.doi.org/10.1016/j.heares.2008.04.013

Neely, S. T., Norton, S. J., Gorga, M. P., & Jesteadt, W. (1988). Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli. *Journal of the Acoustical Society of America, 83*(2), 652–656.

O'Dell, M. (2003). *Intrinsic timing and quantity in Finnish. Acta Universitatis Tamperensis 979.* Tampere: Tampere University Press.

Picton, T. W. (2011). Introduction: past, present, and potential. In *Human auditory evoked potentials.* pp. 1–23. San Diego: Plural Publishing.

Rønne, F. M., Dau, T., Harte, J., & Elberling, C. (2012). Modeling auditory evoked brainstem responses to transient stimuli. *Journal of the Acoustical Society of America, 131*(5), 3903–3913. http://dx.doi.org/10.1121/1.3699171

Rees, A., & Langner, G. (2005). Chapter 12: Temporal coding in the auditory midbrain. In J. A. Winer, & C. E. Schreiner (Eds.), *The inferior colliculus* (pp. 346–376). New York, NY: Springer Science + Business Media Inc.

Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear and Hearing, 31*(3), 302–324. http://dx.doi.org/10.1097/AUD.0b013e3181cdb272

Skoe, E., & Kraus, N. (2012). A little goes a long way: how the adult brain is shaped by musical training in childhood. *Journal of Neuroscience*, *32*(34), 11507–11510. http://dx.doi.org/10.1523/JNEUROSCI.1949–12.2012

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hearing Research*, *308*, 109–121. http://dx.doi.org/10.1016/j.heares.2013.08.004

Stuart, A., Yang, E. Y., & Botea, M. (1996). Neonatal auditory brainstem responses recorded from four electrode montages. *Journal of Communication Disorders*, *29*(2), 125–139.

Suomi, K. (1980). *Voicing in English and Finnish stops: a typological comparison with an interlanguage study of the two languages in contact*. Department of Finnish and General Linguistics of the University of Turku.

Tervaniemi, M., Jacobsen, T., Röttger, S., Kujala, T., Widmann, A., Vainio, M., et al. (2006). Selective tuning of cortical sound-feature processing by language experience. *European Journal of Neuroscience*, *23*(9), 2538–2541. http://dx.doi.org/10.1111/j.1460-9568.2006.04752.x

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron*, *62*, 463–469. http://dx.doi.org/10.1016/j.neuron.2009.05.002

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *Neuroimage*, *31*(3), 1284–1296.

Vainio, M., Järvikivi, J., Aalto, D., & Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *The Journal of the Acoustical Society of America*, *128*, 1313.

Wible, B., Nicol, T., & Kraus, N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain*, *128*(2), 417–423.

Wiese, R. (1996). *The phonology of German*. Oxford, UK: Clarendon Press.

Winer, J. A., & Schreiner, C. E. (2005). Chapter 1: the central auditory system: a functional analysis. In J. A. Winer, & C. E. Schreiner (Eds.), *The inferior colliculus* (pp. 1–68). New York, NY: Springer Science + Business Media Inc.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422.