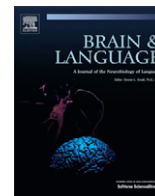


Contents lists available at [SciVerse ScienceDirect](#)

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates

Xiaochen Hu^{a,b,c,*}, Hermann Ackermann^a, Jason A. Martin^f, Michael Erb^c, Susanne Winkler^e, Susanne M. Reiterer^{d,e}

^a Research Group Neurophonetics, Department of General Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

^b MR Research Group, Department of Diagnostic and Interventional Neuroradiology, University of Tübingen, Germany

^c Department of Psychiatry and Psychotherapy, University of Bonn, Germany

^d Centre for Language Learning and Teaching Research, Faculty of Philological and Cultural Studies, University of Vienna, Austria

^e Department of English Studies, Centre of Linguistics, University of Tübingen, Germany

^f Center of Neurology, Division of Neuropsychology, Hertie-Institute for Clinical Brain Research, University of Tübingen, Germany

ARTICLE INFO

Article history:

Available online xxxx

Keywords:

Second language acquisition

Language aptitude

Pronunciation

individual differences

Phonological working memory

Empathy

ABSTRACT

Individual differences in second language (L2) aptitude have been assumed to depend upon a variety of cognitive and personality factors. Especially, the cognitive factor phonological working memory has been conceptualised as language learning device. However, strong associations between phonological working memory and L2 aptitude have been previously found in early-stage learners only, not in advanced learners. The current study aimed at investigating the behavioural and neurobiological predictors of advanced L2 learning. Our behavioural results showed that phonetic coding ability and empathy, but not phonological working memory, predict L2 pronunciation aptitude in advanced learners. Second, functional neuroimaging revealed this behavioural trait to be correlated with hemodynamic responses of the cerebral network of speech motor control and auditory-perceptual areas. We suggest that the acquisition of L2 pronunciation aptitude is a dynamic process, requiring a variety of neural resources at different processing stages over time.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Pronunciation (speech articulation) must be considered one of the most complex human motor skills (Levelt, 1989). Presumably, the process of acquiring such a complex speech-motor skill may be obscured for first language (L1) acquisition due to its early onset in infancy and may generate the impression of an easily acquired ability. However, adult second language (L2) learners face considerable and often lasting problems with pronunciation, contrasting, eventually, with excellent knowledge of vocabulary and grammar – a dissociation of capabilities known as the “Joseph Conrad Phenomenon” (Reiterer et al., 2011). Indeed, adults vary greatly in their L2 pronunciation aptitude – both with respect to segmental (speech sounds) and suprasegmental (intonation, rhythm etc.) manifestations of spoken language (Golestani & Zatorre, 2009; Jilka, 2009a, 2009b). And it has been reported that only between 5% and 15% of adult L2 learners still manage to reach native-like

or “accent-free” speech (Birdsong, 1999, 2005; Novoa, Fein, & Obler, 1988; Seliger, Krashen, & Ladefoged, 1975; Selinker, 1972).

Language aptitude is traditionally defined as a largely innate, relatively fixed talent for learning language (Abrahamsson & Hyltenstam, 2008) and is considered independent of other cognitive abilities, including intelligence (Carroll, 1993; Skehan, 1989). Among the four subcomponents of language aptitude¹ proposed by Carroll (1981), phonetic coding ability (PCA) mainly relates to pronunciation skills. PCA is defined as an ability to identify distinct sounds, to form associations between these sounds and the symbols representing them, and to retain these associations in memory. Any subject of low PCA abilities will have troubles not only in remembering phonetic material or word form, but also in mimicking speech sounds (Carroll, 1962). Sparks and colleagues (Sparks & Ganschow, 1991; Sparks, Ganschow, Javorsky, & Pohlman, 1992) have argued that tests of PCA bear a close relationship to tests for mild dyslexia. Although language aptitude as specified by Carroll’s model has proved to be a good predictor for language learning ability, it does

* Corresponding author. Address: Department of Psychiatry and Psychotherapy, University Hospital of Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany. Fax: +49 228 287 19419.

E-mail address: xiaochen.hu@ukb.uni-bonn.de (X. Hu).

¹ Construct of language aptitude proposed by Carroll (1981): (1) phonetic coding ability; (2) grammatical sensitivity; (3) rote learning ability; (4) inductive language learning ability.

not provide an explanation for individual differences in L2 acquisition (Robinson, 2005).

Among the various cognitive and personality factors related to pronunciation talent (Dörnyei, 2006), phonological working memory (PWM) is considered one of the most salient predictors of L2 aptitude (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998; Miyake & Friedman, 1998). PWM specialises in the retention of verbal information over short periods of time (Baddeley, 1986), and comprises both a phonological store that holds information in phonological form, and sub-vocal articulatory rehearsal that is capable of refreshing the memory trace to prevent its decay (Baddeley, 1990). A variety of behavioural data indicate PWM to be associated with the ability of learning unfamiliar phonological forms (Atkin & Baddeley, 1998; Baddeley, 1993; Baddeley, Papagno, & Vallar, 1988; Gathercole, Service, Hitch, Adams, & Martin, 1999; Papagno, Valentine, & Baddeley, 1991; Service, 1992). As a consequence, this system might serve, among other things, as a language learning device (Baddeley et al., 1998). The network of brain areas bound to PWM include a storage component located in the left supramarginal gyrus (SMG)/inferior parietal lobe, and a rehearsal mechanism involving Broca's area (Awh, Smith, & Jonides, 1995; Baldo & Dronkers, 2006; Chein & Fiez, 2001; Fiez et al., 1996; Henson, Burgess, & Frith, 2000; Koelsch et al., 2009; Paulesu, Frith, & Frackowiak, 1993; Vallar, 2006; Vallar, Di Betta, & Silveri, 1997). The PWM circuit has also been found to overlap with cerebral networks engaged in speech perception and speech production (Aboitiz, Garcia, Bosman, & Brunetti, 2006; Acheson, Hamidi, Binder, & Postle, 2011; Acheson & MacDonald, 2009; Ackermann, 2008; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Hickok & Poeppel, 2007; Reiterer, Erb, Grodd, & Wildgruber, 2008). Currently, the association between PWM and language learning is widely accepted in the fields of psycholinguistics and neurosciences. However, the PWM model cannot account for two important issues of L2 acquisition: (a) the differences between early and late language learners and (b) the role of overt speech articulation.

(a) Since language learning represents a dynamic process, somewhat different predictors for early and more advanced stages of L2 acquisition must be expected (Carroll, 1990; Robinson, 2005). A strong relationship between behavioural measures of PWM, such as digit span or pseudo-word repetition, and language learning abilities could be documented for early-stage learners only, while this association was found to decline in advanced learners (Gathercole, 2006; Masoura & Gathercole, 2005). A previous study of our group was able to confirm that PWM and hemodynamic activation of the respective brain regions, i.e., SMG and Broca's area, predict pronunciation aptitude of early-stage L2 learners (Reiterer et al., 2011). However, the cognitive and personality factors impacting advanced learners still remain to be determined. Previous literature refers to music aptitude (Nardo & Reiterer, 2009; Novoa et al., 1988; Slevc & Miyake, 2006), intelligence (Novoa et al., 1988; Rota & Reiterer, 2009), as well as personality factors (Dörnyei, 2006; Hu & Reiterer, 2009) such as empathy (Guiora & Acton, 1979), extraversion (Dewaele & Furnham, 1999, 2000) and openness to experience (Verhoeven & Vermeer, 2002). Thus, cognitive and personality factors other than PWM may influence advanced L2 learning.

Music aptitude was found to predict both perceptive and productive L2 phonological ability in a group of Japanese immigrants in USA (Slevc & Miyake, 2006). Phonological and musical processing might have similar neural underpinnings. Musicians with absolute pitch showed significantly more hemodynamic activation – compared to those without absolute pitch – in speech-relevant areas such as superior-temporal regions during perception tasks addressing speech prosody (Oechslin, Meyer, & Jäncke, 2010). Passive listening to classical music activated the

language area inferior frontal gyrus (IFG) (Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001). Working memory tasks with verbal materials and with musical materials showed similar activation patterns (Koelsch et al., 2009). Moreover, professional musicians were found to have morphological changes in the speech relevant cortical regions such as planum temporale (PT) and Heschl's gyrus (HG) (Keenan, Thangarai, Halpern, & Schlaug, 2001; Luders, Gaser, Jäncke, & Schlaug, 2004; Schlaug, Jäncke, Huang, & Steinmetz, 1995; Schneider et al., 2005; Wilson, Lusher, Wan, Dudgeon, & Reutens, 2008).

An association between empathy and the capacity for mastery of L2 pronunciation was reported in several studies (Guiora, 1967; Guiora, Beit-Hallahmi, Brannon, Dull, & Scovel, 1972; Guiora, Taylor, & Brandwin, 1968; Taylor, Guiora, Catford, & Lane, 1969). It was suggested that both empathy and L2 pronunciation capacity were influenced by the same underlying process – permeability of “ego boundary” (Guiora & Acton, 1979). More recently the discovery of mirror neurons has provided another possible explanation to the process of language acquisition. Other than the elusive concept “ego boundary”, the neurophysiologically grounded “mirror neuron system” might be the underlying link. The Mirror neuron system has been suggested to be important in language development (Rizzolatti, 2005) and the neural network for empathy is suggested to be composed of the mirror neuron system, the insula, and the limbic system (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Iacoboni, 2009). Empirically, regions of prosody perception and production were found in premotor regions, and activity in these regions was further found to correlate with individual differences in empathy (Aziz-Zadeh, Sheng, & Gheytanchi, 2010). Interestingly, imitation of accent improved language comprehension (Adank, Hagoort, & Bekkering, 2010), and the language comprehension in social context largely relates to empathy (Van den Brink et al., 2010).

As its first purpose, the present study tries to determine the behavioural factors predicting L2 pronunciation aptitude in more advanced learners, i.e., adults with an onset of L2 learning at around an age of 10 years and more than 10 years of exposure to L2 (for the association between language/pronunciation aptitude and other cognitive/personality factors see Dewaele, 2009; Dörnyei, 2006; Hu & Reiterer, 2009; Sparks & Ganschow, 2001). Previous neuroimaging studies of language learning can be classified into investigations of early-stage learning or more advanced learning. fMRI Experiments with early-stage learners have shown negative relationships between phonetic learning performance and the activation in insula/IFG areas and temporal areas during phonetic identification/perception tasks (Golestani & Zatorre, 2004; Wilson & Iacoboni, 2006; Wong, Perrachione, & Parrish, 2007), also in IFG and SMG during imitation tasks (Reiterer et al., 2011). For the advanced learners, better performance in phonetic identification was associated with higher hemodynamic activation in a network encompassing left insula/IFG and temporal areas, SMG, cerebellum and basal ganglia during a phonetic identification task (Callan, Jones, Callan, & Akahane-Yamada, 2004). Highly proficient bilingual subjects also displayed enhanced responses – as compared to low-aptitude individuals – at the level of the insula (Chee, Soon, Lee, & Pallier, 2004) and premotor areas (Majerus et al., 2008) during working memory tasks. The different directions of the activation-aptitude association might indicate different underlying neural mechanisms between early-stage and advanced learners.

(b) Learning to speak essentially represents a motor skill acquisition task (Hickok & Poeppel, 2007) – reaching beyond the elaboration of auditory perception and memory capabilities into the domain of motor development (Seibert, 1927). Language teaching experiments demonstrated, e.g., that L2 learning could be improved by enhancing an overt rehearsal strategy (Ellis &

Beaton, 1993; Iacoboni, 2008). However, the role of overt articulatory processes was neglected so far in models of PWM as a language learning device (Baddeley et al., 1998). Baddeley (2003) even explicitly denied sub-vocal rehearsal to depend upon the capacity for overt speech production. Therefore, the latter processing stages were not included into PWM-based models of a language learning device. As a second aim, the present study tries to elucidate the role of overt speech production in advanced L2 learners. L2 pronunciation aptitude was assessed, therefore, by reading aloud sentences in a second language, i.e., English, and the quality of the pronunciation was rated by native speakers. Since there is evidence for an interplay between language production and perception (Hickok et al., 2003; Iacoboni, 2008; Nasir & Ostry, 2009), we investigated, additionally, neural substrates of both perception and production processes in our advanced L2 learners. In the above-mentioned studies of individual differences in advanced L2 learners, enhanced hemodynamic responses both in the network of speech motor control (insula/IFG, premotor areas, cerebellum, basal ganglia) (Ackermann, 2008) and in the network of speech perception (temporal areas and SMG) (Hickok & Poeppel, 2007) were found for the higher ability group. We hypothesised that advanced L2 learners with high L2 pronunciation aptitude would have higher activities than those with low aptitude in the speech-motor and auditory-perceptual networks.

2. Methods

2.1. Subjects

One-hundred and nine German university students and graduates (42 males and 67 females; age = 26 ± 5 ; age range: 20–40 years), who began learning English at the age of 10 ± 1 years, took part in the behavioural part of the study. According to the English pronunciation aptitude score described below, subjects outside one standard deviation from the mean were classified as high and low aptitude speakers (high aptitude speakers: $n = 18$; low aptitude speakers: $n = 17$). Two subgroups (eight subjects for each group; gender and age matched) were selected for two functional MRI experiments, after providing written informed consent in accordance with the Declaration of Helsinki. The study was approved by the local ethics committee (University of Tübingen, Germany). All subjects were neurologically normal, and right-handed as determined by the Edinburgh handedness inventory (Oldfield, 1971).

2.2. Behavioural assessment (summarised in Table 1)

2.2.1. English pronunciation aptitude

A reading task was used to assess the English pronunciation aptitude, because it facilitates a controlled coverage of the phoneme inventory and specific phonotactic constellations, as well

as a reasonably controlled elicitation of pitch accent distribution and tunes associated with specific discourse situations (Jilka, 2009a, 2009b). All subjects were required to read a standard International Phonetic Alphabet text titled “The Northwind and the Sun” in the best English accent they could “do”. They were free to choose the variant of English (e.g. General American or British English) they were most comfortable with. The pronunciation of the readings was rated by 30 native English speakers (15 females) via the internet on a visual analogue scale from 0 to 10 (least to most representation of “native-speaker-likeness”). Higher scores indicated that the speakers have higher pronunciation aptitude in the L2 English.

2.2.2. Phonetic coding ability

Participants completed a subtest of the Modern Language Aptitude Test – spelling cues (Carroll & Sapon, 2002). Subjects are required to decode the possible pronunciation for an alternative spelled word. As an example, subjects are required to guess the pronunciation of a pseudo-word “luv”, and to choose a possible synonym or related term from following five words – “carry”, “exist”, “affection”, “wash”, “spy”, for which “affection” is the right answer.

2.2.3. Phonological working memory

Phonological working memory (PWM) was assessed using two tasks – a digit span task and a non-word repetition task. The digit span task is a subtest of a German version of Wechsler Adult Intelligence Scale (Tewes, 1994). The non-word repetition task was constructed similarly to the digit span task, using the German non-words derived from a database at the Institute for Natural Language Processing at the University Stuttgart, Germany (IMS, Stuttgart). While both tasks are useful indications of PWM capacity, the non-word repetition task also provides a relative pure measure of the phonological storage (Gathercole, 2006).

2.2.4. Music aptitude

Participants completed the Advanced Measure of Music Audiation (Gordon, 1989). In each test question, subjects are required to listen to two short musical pieces one after the other, and to decide whether these two pieces are the same or different. If the answer is different, then they need to decide whether the difference is a result of a tonal change or a rhythm change.

2.2.5. Intelligence

Verbal and non-verbal intelligence was assessed by two widely used tasks respectively – the multiple choice vocabulary test (Lehrl, Triebig, & Fischer, 1995) and the German version of the Raven’s matrices test (Kratzmeier & Horn, 1980).

Table 1

Summary of the behavioural assessment of the current study.

Test name	Variables	Sources
Reading task	English pronunciation aptitude	Jilka (2009b)
Spelling cues (MLAT)	Phonetic coding ability	Carroll and Sapon (2002)
Digit span (HAWIE-R)	Phonological working memory	Tewes (1994)
Non-word repetition	Phonological working memory	Database, IMS, Stuttgart
AMMA	Music aptitude	Gordon (1989)
MWT-B	Verbal IQ	Lehrl, Triebig and Fischer (1995)
Raven	Non-verbal IQ	Kratzmeier and Horn (1980)
NEO-FFI	Neuroticism, Extraversion, Openness to experience, Agreeableness & Conscientiousness	Borkenau and Ostendorf (1993)
EQ	Empathy	Baron-Cohen and Wheelwright (2004)

Note: HAWIE-R: German version of Wechsler Adult Intelligence Scale; MLAT: Modern Language Aptitude Test; AMMA: Advanced Measures of Music Audiation; MWT-B: Multiple Word Choice Test (Form B); Raven: Raven’s Matrices Test; NEO-FFI: NEO-Five-Factor-Inventory; EQ: Empathy Quotient.

2.2.6. Personality

Five-factor personality was assessed by a German version of the Five-Factor Model personality questionnaire (Borkenau & Ostendorf, 1993). The questionnaire describes personality traits along five dimensions – neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Empathy was assessed by an empathy quotient task (Baron-Cohen & Wheelwright, 2004).

2.2.7. Behavioural data analysis

All the behavioural data analyses were performed by SPSS17 (SPSS, Inc., Chicago, IL). Raw values were used for the following behavioural analysis. Pearson correlations were calculated between the English pronunciation aptitude (EPA) and all the other 12 behavioural scores, including PCA (spelling cues), PWM capacity (digit span, non-word repetition), music aptitude, intelligence (verbal and non-verbal IQ), and personality scores (neuroticism, extraversion, openness to experience, agreeableness, conscientiousness, and empathy). A threshold of $p < 0.0041$ ($p < 0.05/12$) was applied for indicating a significant correlation and a threshold of $p < 0.0083$ ($p < 0.1/12$) was applied for indicating a trend of significant correlation after Bonferroni correction. The variables which passed these thresholds were further entered into a stepwise multiple linear regression analysis (MLR) as independent variables. The EPA score was entered into the MLR as dependent variable. The order of entering the independent variables into the MLR depended on their statistical contribution in explaining the variance in the dependent variable (R^2). The criterion of entering independent variables was a probability of F -change ≤ 0.05 . And the criterion of removing independent variables was a probability of F -change ≥ 0.1 . If more than one variable were found to best predict the EPA, then the interaction effect between the independent variables were tested by adding the product terms into the best model previously found by the stepwise MLR. Additionally, using independent-sample t tests, the two subgroups selected for the fMRI experiments were compared in all the behavioural scores described previously, as well as the behavioural accuracy of both fMRI experiments.

2.3. Functional neuroimaging

2.3.1. fMRI experiment design

A sentence imitation task was used as a production task, since it is the closest way of learning a language – children spontaneously mimic the words of others from the first year of infancy onwards (Gathercole, 2006). In the sentence imitation (Fig. 1A), subjects were required to listen to a sentence (either in German or English) and immediately imitate the presented stimulus with the best mimicking capacity at command. All together 50 sentences (50% German, 50% English; 11 syllables; mean duration = 2.5 s), balanced for syntactic complexity and semantic content, were used as stimuli. In order to avoid movement artefacts, a sparse sampling paradigm was used with sentences presented and imitated during the delay period.

In the perception task – sentence discrimination (Fig. 1B), subjects were required to listen to pairs of sentences in quick succession (sentence 1, sentence 2), which are identical in words, but may be different in prosodic information. And then they were asked to decide, whether the prosody of these two sentences are the same or not. All together 48 sentence pairs (50% German, 50% English; 15 syllables; mean sentence duration = 3.2 s) were used. A continuous sampling paradigm was used for this task.

Stimuli were jittered for both tasks. For the imitation task, the task trials were interleaved with rest trials, and the stimulus onset of each trial was presented randomly between 0 and 2.5 s after the onset of TR delay. Thus the inter-trial interval for the imitation task ranged randomly from 15 to 17.5 s. For the perception task, the

inter-trial interval ranged randomly from 8 to 16 s. Stimuli were also presented in pseudo-randomised order for both tasks. Recently it could be shown that switching behaviour underlies even individual differences in language aptitude (Festman & Münte, 2012). However, in our case the experimental condition focused on phonological processes only (and not naming), and switching between languages was not very demanding under the long inter-trial interval. Thus we ensured that it could not have influenced our results.

The stimulus materials were programmed and presented on the commercially available software “Eprime” and were projected into the scanner via projector and mirror. Stimuli were binaurally presented over MR-compatible earphones (Sennheiser) and the produced speech was recorded by a commercially available MR-compatible optical microphone (www.optoacoustics.com). The overt speech responses of the subjects in the imitation task were recorded in the scanner for later quality control. Before the start of the fMRI experiments, subjects were familiarised with sample stimuli.

2.3.2. Image acquisition

All MR-image acquisitions were performed on a Siemens Vision 1.5 T MRI scanner. An 8-channel head coil was used. For acquiring blood-oxygen-level-dependent signal during the tasks, two different echo planar imaging gradient echo sequences were used. A sparse sampling sequence was used for sentence imitation task: repetition time (TR) = 12 s, delay in TR = 9 s, echo time (TE) = 48 ms, number of slices = 36 transversal, flip angle (FA) = 90°, slice thickness = 3 mm + 1 mm gap, voxel size = $3 \times 3 \times 4 \text{ mm}^3$, field of view (FoV) = $192 \times 192 \times 143 \text{ mm}^3$, matrix = 64×64 . For each trial, sentences were presented and imitated by subjects during the delay period (9 s) of the TR (Fig. 1A, Listen and Imitation) and followed with three seconds of data acquisition (Fig. 1A, Task scan). In this way the movement (articulation) induced artefacts in the data can be reduced.

The continuous sampling sequence was used for sentence discrimination task: TR = 2, delay in TR = 0, TE = 40 ms, number of slices = 27, FA = 90°, slice thickness = 3.5 mm + 1 mm gap, voxel size = $3 \times 3 \times 4.5 \text{ mm}^3$, FoV = $192 \times 192 \times 120 \text{ mm}^3$, matrix = 64×64 . Field maps were acquired to correct the geometric distortions in EPI caused by static-field inhomogeneity using a gradient echo sequence with two echos: TR = 483 ms, TE(1) = 5.19 ms, TE(2) = 9.95 ms, number of slices = 36 transversal, FA = 60°, slice thickness = 3 mm + 0.75 mm gap, voxel size = $3 \times 3 \times 3.75 \text{ mm}^3$, FoV = $192 \times 192 \times 135 \text{ mm}^3$. A T1-weighted MDEFT (modified driven-equilibrium Fourier transform) (Deichmann, Schwarzbauer, & Turner, 2004; Lee et al., 1995; Thomas, De Vita, Deichmann, Turner, & Ordidge, 2005) sequence was used for 3D structural MRI scans: TR = 7.92 ms; TE = 2.48 ms; TI = 910 ms; FA = 16°; voxel size: $1 \times 1 \times 1 \text{ mm}^3$, FoV = $256 \times 224 \times 176 \text{ mm}^3$, matrix = 256×224 .

2.3.3. Image preprocessing

Imaging data were preprocessed and analysed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) executed in MATLAB R2009b (Mathworks Inc.). The origin of each data set was manually set at the anterior commissure. The preprocessing steps for the EPI images were the same for the sentence imitation task and sentence discrimination task, except that the discrimination task included slice time correction to the middle (i.e. 14th slice). Voxel displacement maps (VDM) were created for data sets from both tasks using field maps within the Fieldmap toolbox. EPI images were spatially realigned to the first image, and were field map distortion corrected using VDM. Individual T1-images were registered with EPI images. T1 images were then segmented using “new-segment” tool in the SPM8. Together with another set of T1-weighted MDEFT images from a previous project (Hu et al., 2011), a DARTEL tem-

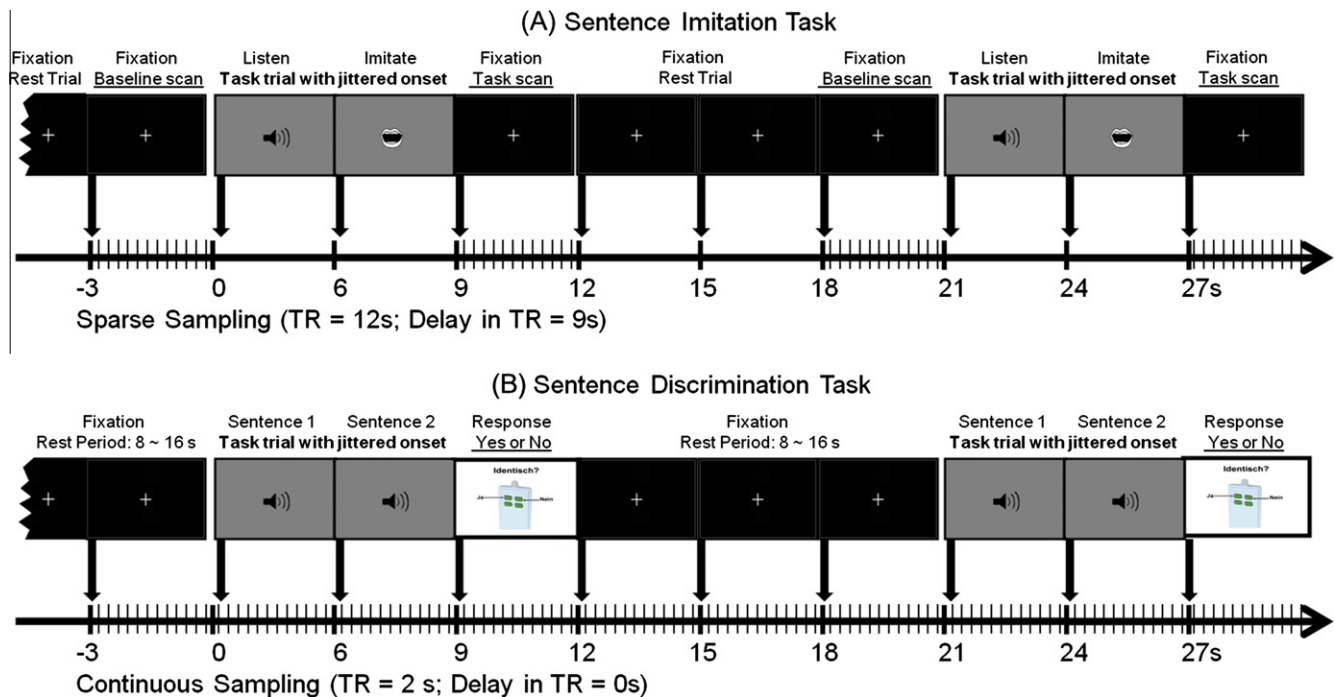


Fig. 1. Graphical depiction of the sentence imitation task and the sentence discrimination task. In the sentence imitation task (A), subjects were required to listen to a sentence and immediately imitate the presented stimulus with the best mimicking capacity at command. In order to avoid movement artefacts, a sparse sampling sequence with TR = 12 s and delay in TR = 9 s was used. The stimuli were jittered and presented every two TRs within the delay period of the TR. In the sentence discrimination task (B), subjects were required to listen to pairs of sentences in quick succession (sentence 1, sentence 2), and were asked to decide, whether the prosody of these two sentences are identical. A continuous sampling sequence with TR = 2 s was used. The stimuli were jittered.

plate of 68 subjects was generated and flow field was calculated for each subject in the current study within DARTEL toolbox (Ashburner, 2007). EPI images were then normalised to Montreal Neurological Institute (MNI; resample to voxel size = $1.5 \times 1.5 \times 1.5 \text{ mm}^3$) space, and also modulated by the DARTEL flow field. They were then smoothed using a 10 mm full-width at half maximum (FWHM) Gaussian filter.

2.3.4. fMRI statistical analysis

For each subject, two separated General Linear Model (GLM) were estimated for each task. A GLM of the imitation task comprised two conditions (German imitation, English imitation) as regressors of interest. The recorded speech of subjects during the imitation task was *post hoc* controlled and the wrongly imitated sentences were considered as error trials. The following criteria were used for identifying wrong imitation: (1) the imitation was not completed; (2) the words or phrases within the sentence were not identical to that of the original sentence in the sense of word form, not pronunciation; (3) the subjects stuttered or hesitated too much during the imitation. Therefore the error trial and six motion correction parameters calculated during the rigid-body realignment were also included in this GLM as regressors of non-interest. The other GLM of the discrimination task comprised four conditions (German sentence 1, German sentence 2, English sentence 1, and English sentence 2) as regressors of interest. The accuracy of the response during the discrimination task was *post hoc* controlled. The error trials and six motion correction parameters were also included in this GLM as regressors of non-interest.

At the group level, two separated full factorial analysis of variance (ANOVA) design matrices were set up for each task. For the imitation task, a 2×2 ANOVA was performed, with one between-subject factor "aptitude" (low vs. high aptitude), and one within-subject factor "language" (German vs. English). Main effect for aptitude and the aptitude \times language interaction effect were

calculated. For the discrimination task, a $2 \times 2 \times 2$ ANOVA was performed with one between subject factor "aptitude" (low vs. high aptitude), one within-subject factor "language" (German vs. English), and one within-subject factor "sentence" (sentence 1 vs. sentence 2). Main effect for aptitude and the aptitude \times language interaction effect separately for sentence 1 and sentence 2 were calculated. For all the analyses, a statistical threshold of $p < 0.05$ (whole brain cluster level family-wise error (FWE) corrected) was obtained. Results were overlaid on the rendered image of the previously obtained DARTEL template and on the normalised mean T1 image of the group.

3. Results

3.1. Behavioural results

All 30 native speaker ratings of the EPA were averaged for all 109 subjects (range = 1.52–9.69, mean = 5.84 ± 1.69). The intra-class correlation coefficient of the ratings was 0.96, indicating a high inter-rater reliability. Pearson correlation revealed that EPA significantly correlated with phonological coding ability (PCA; $r = 0.56, p < 0.001$), music aptitude ($r = 0.27, p = 0.002$), and openness to experience ($r = 0.29, p = 0.001$). Empathy also correlated with EPA at a trend level ($r = 0.23, p = 0.008$). These variables were further entered into the stepwise regression analysis and this analysis revealed that, PCA and empathy together significantly predicted the EPA scores (Table 2, model 2; $R^2 = 0.34$). No significant interaction was found between the PCA and the empathy scores (Table 2, model 3; $R^2 \Delta = 0.01, p > 0.05$). Independent-sample *t* tests showed that, the high and low EPA subgroups, which were selected for fMRI examination, were significantly different from each other in their EPA ($t = 13.43, p < 0.001$), PCA ($t = 5.08, p < 0.001$), empathy ($t = 2.77, p < 0.05$) and the behavioural accuracy of the sentence imitation task ($t = 2.40, p < 0.05$). No group differences were found

Table 2
Summary of stepwise regression results for English pronunciation aptitude.

Model	DV	IVs	β	R	R ²	R ² Δ	F-change (df)
1	EPA	PCA	0.56***	0.56	0.31	0.31***	48.86 (1, 107)
2	EPA	PCA	0.54***	0.58	0.34	0.03 ^a	4.36 (1, 106)
		Empathy	0.17 [*]				
3	EPA	PCA	0.06	0.59	0.35	0.01 ^a	2.89 (1, 105)
		Empathy	-0.45				
		PCA \times Empathy	0.84				

Note: DV: dependent variable; IVs: Independent variables; EPA: English pronunciation aptitude; PCA: Phonetic coding ability; PCA \times Empathy: The products of PCA and Empathy.

^{*} $p < 0.05$.

^{***} $p < 0.001$.

^a $p > 0.05$.

for other cognitive and personality factors, or the behavioural accuracy of the sentence discrimination tasks ($t \leq 1.65$, $p > 0.05$).

3.2. fMRI results

In the imitation task, there was a main effect for aptitude (high vs. low) in a network (Fig. 2A; Table 3, tag 1–7) including left superior temporal gyrus (STG), left superior temporal pole (STP), left IFG, left Insula, left rolandic operculum (RO), left postcentral and precentral gyri, left superior and middle frontal gyrus (SFG & MFG), bilateral paracentral gyrus, bilateral supplementary motor areas (SMA), right posterior STG, right RO, right HG, right inferior temporal gyrus (ITG), right cerebellum lobule 1 (crus 1), and the basal ganglia cluster of left caudate and bilateral thalamus. No language \times group interaction was found for the imitation task.

When listening to the first sentence (sentence 1) of the discrimination task (Fig. 2B; Table 3, tag 8), there was a main effect for aptitude (high vs. low) in a left-sided network including STG, STP, IFG, RO, HG and middle temporal gyrus (MTG). When listening to the second sentence (sentence 2; Fig. 2C; Table 3, tag 9–10), there was also a main effect for aptitude (high vs. low) in a left-sided network including insula, IFG, STP, STG, RO, SFG, MFG, and precentral gyrus. An aptitude \times language interaction effect were found for sentence 1 of the discrimination task (contrast English vs. German for high EPA group with English vs. German for low EPA group; Fig. 2D; Table 3, tag 11–13) in a left-sided network including STP, STG, IFG, MTG, precentral gyrus, posterior MTG, and posterior STG. The cluster of IFG, STG, STP and RO overlapped across the imitation task and both the sentence 1 and sentence 2 of discrimination task (Fig. 2E). And the cluster of SFG and MFG overlapped across the imitation task and sentence 2 of discrimination task (Fig. 2E).

4. Discussion

The first purpose of the current study is to find out the behavioural predictors of L2 pronunciation aptitude in advanced language learners. For the current sample of native German speakers, English pronunciation aptitude (EPA) was measured as the L2 pronunciation aptitude. The high inter-rater reliability for the EPA scores indicates that our L2 pronunciation aptitude scores for the current sample are very reliable. The correlation analysis revealed that, not PWM, but other cognitive and personality factors are related to EPA, including phonetic coding ability (PCA), music aptitude, empathy, and openness to experience. These results are consistent with previous literature that the association between PWM and L2 learning ability was only expected to be strong in the early-stage learners, and this association was expected to be declined in advanced learners (Gathercole, 2006; Masoura & Gathercole, 2005). Among all the cognitive and personality factors, multiple regression analysis revealed that, only PCA and empathy

together significantly predicted the EPA scores. Although music aptitude and openness were also associated with the EPA scores, their contribution in the multiple regression analysis overlaps with that of the PCA and empathy scores.

The fMRI data of the imitation task revealed a network of speech-motor control and speech-auditory perception, which is more activated in the high EPA group. Part of this activated network for high EPA group overlaps with the whole network of speech-motor control. The left IFG, left premotor areas, left anterior insula, and right superior cerebellum (crus 1) are among the preparative loop of the speech-motor control network and are responsible for planning of speech movement sequences (Ackermann, 2008; Ackermann & Riecker, 2010; Riecker et al., 2005). SMA is also part of the preparative loop and is related to preparedness for initiation of vocal tract movements (Ackermann, Hertrich, & Ziegler, 2010; Ackermann & Ziegler, 2010; Brendel et al., 2010). Motor cortex, basal ganglia regions (caudate and thalamus) are among the executive loop and are responsible for on-line innervations of vocal tract muscles (Ackermann, 2008; Riecker et al., 2005). Previous meta-analysis revealed that the same speech-motor control network was involved in different types of overt speech production tasks (Eickhoff, Heim, Zilles, & Amunts, 2009). This result indicates that participants with better L2 pronunciation aptitude highly rely on the speech-motor or articulation network during the imitation task.

Another part of the activated network for the high EPA group during the imitation task overlaps with the speech-auditory perception network, including left anterior superior temporal areas (STP and anterior STG), right HG, the adjacent posterior STG, and right ITG. Superior temporal areas are important for representing and/or processing phonological information (Hickok & Poeppel, 2007) and for prelexical processing of auditory speech (Price, 2010). The anterior part of the STG is involved in perception of intelligible or familiar phonetic sounds (Binder et al., 2000; Davis & Johnsruide, 2003; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Matsumoto et al., 2011; Obleser, Wise, Dresner, & Scott, 2007; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007), while the posterior part of the STG and adjacent HG is involved in simple temporally encoded auditory information, which has no preference for intelligible speech sounds (Binder et al., 2000; Liebenthal et al., 2005; Obleser, Zimmermann, et al., 2007; Scott & Johnsruide, 2003). HG is the “core” auditory cortex, and activated when listening to any type of sound compared to silence (not specific to speech) (Scott & Johnsruide, 2003). The higher activation for the high aptitude group in right HG and the adjacent STG area may be related to analysing the fundamental frequency and tonal information of one’s own voice for feedback-guided modulation during the task (Perry et al., 1999; Zatorre, 2001). Furthermore, right pSTG was also involved in both explicit and implicit processing of affective prosody information (Frühholz, Ceravolo, & Grandjean, 2012; Frühholz & Grandjean, 2012). Although affective prosody was not

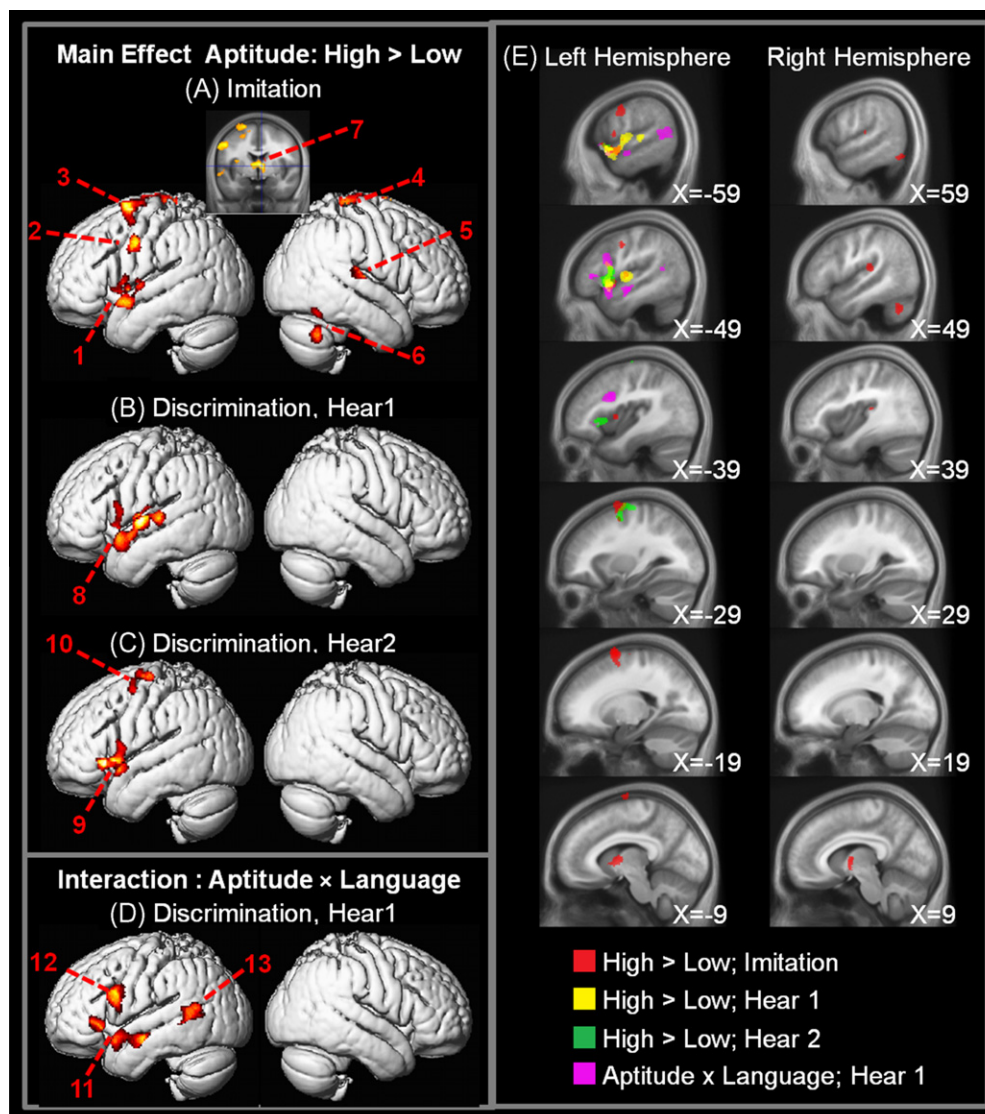


Fig. 2. The main result of the current study. (A) The network of brain areas activated for the main effect for aptitude (high vs. low) for the imitation task in seven clusters: (1) left-sided inferior frontal gyrus (IFG), superior temporal pole (STP), superior temporal gyrus (STG), rolandic operculum (RO), and insula; (2) left-sided postcentral and precentral gyri; (3) left-sided superior frontal gyrus (SFG), superior middle frontal gyrus (SFG), supplementary motor area (SMA); (4) bilateral paracentral gyri and SMA; (5) right-sided STG, RO, and Heschl's gyrus (HG); (6) right-sided superior cerebellum lobule 1 (crus1), and inferior temporal gyrus (ITG); (7) the basal ganglia regions including left caudate and bilateral thalamus. (B) The network of brain areas activated for the main effect for aptitude (high vs. low) for sentence 1 of the discrimination task in the eighth cluster including left-sided IFG, STP, STG, RO, HG, and middle temporal gyrus (MTG). (C) The network of brain areas activated for the main effect for aptitude (high vs. low) for sentence 2 of the discrimination task in two clusters: (9) left-sided IFG, STP, STG, RO, insula; (10) left-sided SFG, MFG, precentral gyrus. D: The network of brain areas activated for the interaction effect for sentence 1 of the discrimination task in three clusters: (11) left-sided IFG, STP, STG, MTG; (12) left-sided IFG, and precentral gyrus; (13) left-sided STG, and MTG. E: Overlaps of the activations from A–D. Threshold for all activations are $p < 0.05$ (whole brain cluster level FWE corrected).

a designed experimental condition in the current study, it cannot be excluded that the high aptitude group might be implicitly processing some general affective prosody elements during the sentence imitation task. Right ITG may correspond to the lexical interface, which links phonological and semantic information (Hickok & Poeppel, 2007). This result indicates that participants with better L2 pronunciation aptitude also highly rely on the auditory-perceptual network during the imitation task.

During the discrimination task, listening to the first sentence (sentence 1) is a relatively pure perceptual process, while listening to the second sentence (sentence 2) involves phonological working memory (PWM) processing so that the second sentence can be compared with the first one. Our results show that the contrast high versus low aptitude groups yields different activation patterns for sentence 1 and sentence 2. The high EPA group had more activities than the low EPA group in left-sided IFG, anterior superior

temporal areas (STP, and anterior STG) for both sentence 1 and sentence 2. Left HG and anterior MTG were only activated for sentence 1, and left insula, dorsal premotor areas (SFG, and MFG) were only activated for sentence 2. This result suggests that participants with better L2 pronunciation aptitude recruit more of the auditory-perceptual network during sentence 1, and recruit more of the speech-motor network during sentence 2.

An aptitude \times language Interaction effect was found during sentence 1 in left-sided IFG, precentral gyrus, anterior temporal regions (STP, anterior STG, and anterior MTG), and posterior temporal regions (posterior STG and MTG). Further analysis of these activation levels showed that ROIs for sentence 1 of the discrimination task included activation clusters either with or without main effect for aptitude or aptitude \times language interaction effect. The regions with both main effect for aptitude and interaction effect (left IFG, and STP) indicates that the high EPA group had more activa-

Table 3
Loci of BOLD activation clusters.

Tag	Location	Peak MNI coordinates			T	Z	Cluster-level	
		x	y	z			Size	<i>p</i> _{FWE}
<i>A: Imitation task, clusters of main effect for aptitude: high > low</i>								
1	L IFG, STP, STG, RO, Insula	−56	6	−9	4.91	4.14	1196	−***
2	L postcentral, precentral gyri	−54	−3	40	4.77	4.04	445	−*
3	L SFG, MFG, SMA	−18	5	70	6.11	4.83	1139	−***
4	B paracentral gyri, SMA	2	−34	70	4.75	4.04	754	−**
5	R STG, RO, HG	44	−25	13	5.37	4.42	463	−*
6	R crus1, ITG	53	−60	−35	5.16	4.29	427	−*
7	L Caudate, B thalamus	8	−6	6	4.55	3.91	1249	−***
<i>B: Discrimination task, hear1, cluster of main effect for aptitude: high > low</i>								
8	L IFG, STP, STG, MTG, RO, HG	−50	−9	3	5.15	4.64	2514	−***
<i>C: Discrimination task, hear2, clusters of main effect for aptitude: high > low</i>								
9	L IFG, STP, STG, RO, Insula	−48	11	−2	5.03	4.55	1618	−***
10	L SFG, MFG, precentral gyrus	−35	−16	66	3.85	3.61	513	−*
<i>D: Discrimination task, hear1, clusters of interaction effect: aptitude × language</i>								
11	L IFG, STP, STG, MTG	−50	−13	−14	5.38	4.81	1469	−***
12	L IFG, precentral gyrus	−41	9	25	5.06	4.57	859	−***
13	L STG, MTG	−62	−51	4	4.51	4.15	671	−**

Note: L, left; R, Right; B, Bilateral; PCA: phonetic coding ability; STG: superior temporal gyrus; STP: superior temporal pole; IFG: inferior frontal gyrus; RO: rolandic operculum; SFG: superior frontal gyrus; MFG: middle frontal gyrus; SMA: supplementary area; HG: heschl's gyrus; crus1: cerebellum lobule 1; ITG: inferior temporal gyrus; MTG: middle temporal gyrus.

* *p*_{FWE} < 0.05.

** *p*_{FWE} < 0.01.

*** *p*_{FWE} < 0.001.

tions than the low EPA group in both first and second language perception in these regions. Additionally, the hemodynamic responses in these regions were stronger for L2. The regions with only main effect for aptitude (left RO, HG, the upper part of the anterior STG, and the more anterior part of the anterior MTG) indicates that the higher activations for the high EPA group in these regions were not language specific. They were recruited regardless of the language under processing (L1 or L2). The regions with only an interaction effect (left precentral gyrus and IFG, left posterior STG and MTG, the lower part of the anterior STG, and the posterior part of the anterior MTG) indicated that these regions were more specialised for English phoneme processing in the high EPA group. No aptitude × language interaction effects were found for the sentence imitation task and the sentence 2 of the discrimination task, which indicates that for these functions higher activations for the high EPA group in the activated regions were not language specific.

To summarise, it seems that, the neural bases of excellence in L2 perception and production are mostly supported by the excellence in L1 perception and pronunciation, except that for the L2 perception, extra regions are required for English phoneme processing (regions with only an interaction effect but not a main effect for group such as left posterior STG). It becomes increasingly clear (for review see Sparks, 2012) that L1 aptitude is strongly tied to L2 aptitude, regardless whether this effects the phonetic system or aptitude for other subsystems of language (e.g. grammar learning). On the basis of this fact and common source of L1 and L2 learning ability or 'faculty of language' we also argue that a common left-hemisphere focused network is responsible for language processing *per se*, be it L1, L2 or a further language. Language processing draws always on partially the same resources (Reiterer, 2010). In earlier research we found that regardless of the language of the actual task (L1 or L2) activation patterns used to be similar for certain ability or proficiency groups – even during L1 processing individual differences were found (Reiterer, Hemmelmann, Rappelsberger, & Berger, 2005).

These results are consistent with our hypothesis that advanced L2 learners with high L2 pronunciation aptitude would have higher activations than those with low aptitude in both the speech-motor and the auditory-perceptual networks. However, the results are only partly consistent with previous neuroimaging studies for ad-

vanced L2 learners (Callan et al., 2004; Chee et al., 2004; Majerus et al., 2008). Callan and colleagues (Callan et al., 2004) suggested another region, the supramarginal gyrus (SMG), which subserves individual differences in the performance of L2 phonetic identification. Nevertheless, this region was not found in the current study. This may be related to the specific role of SMG in phonological processing. SMG is one of the overlapping terminologies that describe the areas surrounding the Sylvian fissure at the parieto-temporal boundary (SMG/PT/angular gyrus), and considered to play a key role in the phonological store (Baldo & Dronkers, 2006; Buchsbaum, Olsen, Koch, & Berman, 2005; Henson et al., 2000; McGettigan et al., 2011; Paulesu et al., 1993; Vallar et al., 1997). The current study revealed no aptitude group differences in the activations of any of these areas, as well as no differences in the behavioural scores of both PWM measures (digit span and non-word span) or the accuracy of the discrimination task. This suggests that L2 pronunciation aptitude in advanced learners does not rely on the brain regions that subserves the phonological store component of the PWM (SMG/PT/angular gyrus).

Previous literature (Buchsbaum et al., 2005) suggested separate roles for temporal areas and the phonological store area (PT) during PWM processing. Temporal area is sensitive to the auditory modality, and is involved in a kind of transient "echoic" delay period activity. PT is insensitive to the source of modality, and is involved in a sustained delay period activity. The group differences of the auditory-perceptual network in the current study were within temporal regions (e.g. STG, STP, MTG), and not in the PT. Similarly, superior temporal regions are activated in the immediate imitation of pseudo-words (shadowing task), indicating its role in preparation of a translation of perceptual into motor information (Peschke, Ziegler, Kappes, & Baumgaertner, 2009). Especially, left posterior STG has previously been shown to be activated during both speech perception and production tasks and has been proposed to serve as an interface between speech perception, lexical recall and speech production (Buchsbaum, Hickok, & Humphries, 2001; Wise et al., 2001). We suggest that, the high EPA group might be more capable of "catching" the phonological information processed in the transient delayed period and using it as the input information for the speech-motor or articulatory network. Our high EPA group had higher activation in the whole speech-motor

control network, including the preparative and executive loops (Ackermann, 2008; Ackermann & Riecker, 2010), but only a part of the auditory-perceptual network. We suggest that, the purpose of greater recruitment of the auditory-perceptual network in the high EPA group might be for better functioning in pronunciation production (speech-motor output).

The widely accepted PWM model for language acquisition proposes that PWM is a language learning device and plays a crucial role in learning the novel phonological forms of new words (Baddeley et al., 1998). However, the association between the PWM and L2 aptitude were only strong in early-stage L2 learners (Gathercole, 2006; Masoura & Gathercole, 2005). The strong association between PWM and L2 aptitude for pronunciation in early-stage L2 learners has been confirmed through both behavioural and brain imaging data of our previous study (Reiterer et al., 2011). For advanced learners, the current study demonstrates that there was no association between the capacity of behavioural PWM (digit span and non-word span) measures and L2 pronunciation aptitude. The presented imaging data also shows that the regions subserving the phonological storage, an important component of the PWM, was not related to the individual differences in L2 pronunciation aptitude.

On the other hand, the PWM model neglects the role of overt articulation in the language learning process (Baddeley, 2003; Baddeley et al., 1998). Baddeley's PWM model only includes the sub-vocal rehearsal, not the overt articulatory rehearsal component. On the contrary, the current results suggest that L2 pronunciation aptitude in advanced learners relies on the development of the whole speech-motor control network, including areas related to overt articulation or speech motor executive loop (Ackermann, 2008; Eickhoff et al., 2009). Indeed, several studies have shown that the speech motor executive regions (motor cortex, brain stem, basal ganglia, and cerebellum) were also involved in the sub-vocal rehearsal process of the PWM (Awh et al., 1995; Koelsch et al., 2009; Vallar, 2006). Therefore, the exclusion of the overt articulation from the PWM model as a language learning device is rather arbitrary.

The reason for excluding the overt articulatory component from PWM might be that the previous language learning studies with working memory measures did not focus on the speech-motor or pronunciation aspect of language learning. The classical PWM model holds the role of the sub-vocal rehearsal to be preventing the decay of the phonological representation (Baddeley et al., 1998). However, it was not clear as to the degree of such prevention – only to avoid the decay of the “core” phonological structure (segmental information), or also to avoid the decay of the variation in the phonological representation (prosodic or suprasegmental information). The latter aspect cannot be tested by the classical PWM measures such as the simple span tasks, because only the quantitative amount of retained phonological representation is measured, not the quality of that representation. There is evidence for individual differences of quality of phonological representation and that influences long-term phonological learning (Service, Maury, & Luotoniemi, 2007). Moreover, the rehearsal process during the classical non-word repetition task can be still carried out without perfectly imitating the phonetic and prosodic details (Kappes, Baumgaertner, Peschke, Goldenberg, & Ziegler, 2010), such that the involvement of the overt articulation process may enhance the latter part of the prevention.

5. Conclusions

The current study shows that phonetic coding ability and empathy together, but not the classical measures of phonological working memory, predict language pronunciation aptitude in advanced

L2 learners. The observed higher levels of hemodynamic responses for the high aptitude group are the whole speech-motor network, including regions for speech-motor preparation/planning as well as speech-motor execution, and parts of the auditory-perceptual network, including the areas for perception of familiar/intelligible phonemes. This fMRI result indicates that these components crucially contribute to the excellence of L2 pronunciation aptitude in advanced learners. In line with the behavioural result, the vital region for phonological storage component was not involved in this network. It is different from the early-stage learning, for which the phonological working memory plays more important role. We suggest that, the acquisition of L2 pronunciation aptitude is a dynamic process, requiring different neural resources for different stages.

Acknowledgments

We wish to thank Mathias Röger and Maïke Borutta for the help with the scanning and acquisition; Davide Nardo for the help with preparing the questionnaire; and English native speakers recruited via the internet for evaluating the English pronunciation aptitude of our subjects. This research was funded by the German Research Foundation (DFG grant AC 55/7-1) and Centre for Integrative Neuroscience (CIN) in Tübingen, Germany.

Author contributions: Preparations and designs of experiment: SMR, HA, XH. Experiment conduction: SMR & XH. Data analysis: XH & ME. Manuscript preparations: XH, SMR, HA, JAM, SW.

References

- Aboitiz, F., Garcia, R. R., Bosman, C., & Brunetti, E. (2006). Cortical memory mechanisms and language origins. *Brain and Language*, 98, 40–56.
- Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30, 481–509.
- Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23, 1358–1367.
- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135, 50–68.
- Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, 31, 265–272.
- Ackermann, H., Hertrich, I., & Ziegler, W. (2010). Dysarthria. In J. S. Damico, N. Müller, & M. J. Ball (Eds.), *The handbook of language and speech disorders* (pp. 362–390). Malden, MA: Wiley-Blackwell.
- Ackermann, H., & Riecker, A. (2010). The contribution(s) of the insula to speech production: A review of the clinical and functional imaging literature. *Brain Structure and Function*, 214, 419–433.
- Ackermann, H., & Ziegler, W. (2010). Brain mechanisms underlying speech motor control. In W. J. Hardcastle, J. Laver, & F. E. Gibbons (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 202–250). Malden, MA: Wiley-Blackwell.
- Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, 21, 1903–1909.
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, 38, 95–113.
- Atkin, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19, 537–552.
- Awh, E., Smith, E. E., & Jonides, J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. *Annals of the New York Academy of Sciences*, 769, 97–118.
- Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *Plos One*, 5, 1–8.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. D. (1993). Short-term phonological memory and long-term learning: A single case study. *Journal of Cognitive Psychology*, 5, 129–148.
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27, 586–595.
- Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 54–73). Cambridge: Cambridge University Press.

- Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20, 529–538.
- Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163–175.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–528.
- Birdsong, D. (1999). Introduction: Whys and why not of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161–181). Mahwah, NJ: Erlbaum.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives* (pp. 109–127). New York: Cambridge University Press.
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae: Handanweisung*. Göttingen: Hogrefe.
- Brendel, B., Hertrich, I., Erb, M., Lindner, A., Riecker, A., Grodd, W., et al. (2010). The contribution of mesofrontal cortex to the preparation and execution of repetitive syllable productions: An fMRI study. *Neuroimage*, 50, 1219–1230.
- Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25, 663–678.
- Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, 48, 687–697.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage*, 22, 1182–1194.
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of USA*, 100, 5497–5502.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.
- Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training and research in education* (pp. 87–136). Pittsburgh, PA: University of Pittsburgh Press.
- Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–29). Englewood Cliffs, NJ: Prentice Hall.
- Carroll, J. B., & Sapon, S. (2002). *Modern language aptitude test*. Bethesda, MD: Second Language Testing, Inc.
- Chee, M. W. L., Soon, C. S., Lee, H. L., & Pallier, C. (2004). Left insula activation: A marker for language attainment in bilinguals. *Proceedings of the National Academy of Sciences of USA*, 101, 15265–15270.
- Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, 11, 1003–1014.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, 23, 3423–3431.
- Deichmann, R., Schwarzbauer, C., & Turner, R. (2004). Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 T and 3 T. *Neuroimage*, 21, 757–767.
- Dewaele, J. M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49, 509–544.
- Dewaele, J. M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, 28, 355–365.
- Dewaele, J. M. (2009). Individual differences in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 623–646). Emerald: Bingley.
- Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, 19, 42–68.
- Eickhoff, S., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A*, 367, 2399–2421.
- Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology A*, 46, 533–558.
- Festman, J., & Münte, T. F. (2012). Cognitive control in Russian–German bilinguals. *Frontiers in Psychology*, 3(116), 1–6.
- Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *The Journal of Neuroscience*, 16, 808–822.
- Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, 22, 1107–1117.
- Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expression. *Neuroimage*, 62, 1658–1666.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13, 65–77.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage*, 21, 494–506.
- Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, 109, 55–67.
- Gordon, E. E. (1989). *Advanced measures of music audition*. Chicago: GIA Publications.
- Guiora, A. Z. (1967). Toward a systematic study of empathy. *Comprehensive Psychiatry*, 8, 375–385.
- Guiora, A. Z., & Acton, W. R. (1979). Personality and language behavior: A restatement. *Language Learning*, 29, 193–204.
- Guiora, A. Z., Beit-Hallahmi, B., Brannon, R. C. L., Dull, C. Y., & Scovel, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study. *Comprehensive Psychiatry*, 13, 421–428.
- Guiora, A. Z., Taylor, L. L., & Brandwin, M. A. (1968). The role of empathy in second language behavior. In *Proceedings of the 16th international congress of applied psychology*. Amsterdam: Swets and Zeitlinger.
- Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recording, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia*, 38, 426–440.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, 15, 673–682.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Hu, X., Erb, M., Ackermann, H., Martin, J. A., Grodd, W., & Reiterer, S. M. (2011). Voxel-based morphometry studies of personality: Issue of statistical model specification—Effect of nuisance covariates. *Neuroimage*, 54, 1994–2005.
- Hu, X., & Reiterer, S. M. (2009). Personality and pronunciation talent in second language acquisition. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and the brain* (pp. 97–129). Berlin: de Gruyter.
- Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology – Paris*, 102, 31–34.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.
- Jilka, M. (2009a). Talent and proficiency in language. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and the brain* (pp. 1–16). Berlin: de Gruyter.
- Jilka, M. (2009b). Assessment of phonetic ability. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and the brain* (pp. 17–66). Berlin: de Gruyter.
- Kappes, J., Baumgaertner, A., Peschke, C., Goldenberg, G., & Ziegler, W. (2010). Imitation of para-phonological detail following left hemisphere lesions. *Neuropsychologia*, 48, 1115–1124.
- Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage*, 14, 1402–1408.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, 30, 859–873.
- Kratzmeier, H., & Horn, R. (1980). *Raven-Matrizen-test. Advanced progressive matrices (APM)*. Weinheim: Beltz.
- Lee, J. H., Garwood, M., Menon, R., Adriany, G., Andersen, P., Truwit, C. L., et al. (1995). High contrast and fast three-dimensional magnetic resonance imaging at high fields. *Magnetic Resonance in Medicine*, 34, 308–312.
- Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, 91, 335–345.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.
- Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *Neuroimage*, 20, 2142–2152.
- Lieenthal, E., Binder, J. R., Spitzer, S. M., Posing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, 15, 1621–1631.
- Luders, E., Gaser, C., Jäncke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage*, 22, 656–664.
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: An MEG study. *Nature Neuroscience*, 4, 540–545.
- Majerus, S., Belayachi, S., De Smedt, B., Leclercq, A. L., Martinez, T., Schmidt, C., et al. (2008). Neural networks for short-term for order differentiate high and low proficiency bilinguals. *Neuroimage*, 42, 1698–1713.
- Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory*, 13, 422–429.
- Matsumoto, R., Imamura, H., Inouchi, M., Nakagawa, T., Yokoyama, Y., Matsushashi, M., et al. (2011). Left anterior temporal cortex actively engages in speech perception: A direct cortical stimulation study. *Neuropsychologia*, 49, 1350–1354.
- McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., & Scott, S. K. (2011). Neural correlates of sublexical processing in phonological working memory. *Journal of Cognitive Neuroscience*, 23, 961–977.
- Miyake, A., & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

- Nardo, D., & Reiterer, S. M. (2009). Musicality and phonetic language aptitude. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and the brain* (pp. 213–255). Berlin: de Gruyter.
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of USA*, 106, 20470–20475.
- Novoa, L., Fein, D., & Obler, L. K. (1988). Talent in foreign languages: A case study. In L. K. Obler & D. Fein (Eds.), *The exceptional brain: The neuropsychology of talent and special abilities* (pp. 294–302). London, New York: Guilford Press.
- Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, 27, 2283–2289.
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17, 2251–2257.
- Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch – Functional evidence of speech-relevant auditory acuity. *Cerebral Cortex*, 20, 447–455.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331–347.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.
- Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport*, 10, 3979–3984.
- Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, A. (2009). Auditory–motor integration during fast repetition: The neuronal correlates of shadowing. *Neuroimage*, 47, 392–402.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 62–88.
- Reiterer, S., Erb, M., Grodd, W., & Wildgruber, D. (2008). Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging and Behavior*, 2, 1–10.
- Reiterer, S., Hemmelmann, C., Rappelsberger, P., & Berger, M. L. (2005). Characteristic functional networks in high- versus low second language speakers detected also during native language processing: An explorative EEG coherence study in 6 frequency bands. *Cognitive Brain Research*, 25, 566–578.
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., et al. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2, 271.
- Reiterer, S. (2010). The cognitive neuroscience of second language acquisition and bilingualism. In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 307–321). Amsterdam, Philadelphia: John Benjamins.
- Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., et al. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, 64, 700–706.
- Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, 210, 419–421.
- Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46–73.
- Rota, G., & Reiterer, S. M. (2009). Cognitive aspects of pronunciation talent. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and the brain* (pp. 67–96). Berlin: de Gruyter.
- Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267, 699–701.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, 8, 1241–1247.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107.
- Seibert, L. C. (1927). An experiment in learning French vocabulary. *Journal of Educational Psychology*, 18, 294–309.
- Seliger, H. W., Krashen, S., & Ladefoged, P. (1975). Maturational constraints in the acquisition of second language accent. *Language Sciences*, 36, 20–22.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209–231.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology A*, 45, 21–50.
- Service, E., Maury, S., & Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Memory and Cognition*, 35, 1122–1135.
- Skehan, P. (1989). *Individual differences in second language learning*. London: Arnold.
- Slevc, L., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17, 675–681.
- Sparks, R. L. (2012). Individual differences in L2 learning and long-term L1–L2 relationships. *Language Learning*, 62, 5–27.
- Sparks, R. E., & Ganschow, L. (1991). Foreign language learning differences: Affective filter or native language aptitude differences. *Modern Language Journal*, 75, 3–15.
- Sparks, R. E., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, 21, 90–111.
- Sparks, R. E., Ganschow, L., Javorsky, J., & Pohlman, J. (1992). Test comparisons among students identified as high-risk, low-risk, and learning disabled in high-school foreign language courses. *Modern Language Journal*, 76, 142–158.
- Taylor, L. L., Guiora, A. Z., Catford, J. C., & Lane, H. L. (1969). The role of personality variables in second language behavior. *Comprehensive Psychiatry*, 48, 285–290.
- Tewes, U. (1994). *HAWIE-R. Hamburg-Wechsler-Intelligenztest für Erwachsene, Revision 1991*. Bern, Göttingen, Toronto, Seattle: Verlag Hans Huber.
- Thomas, D. L., De Vita, E., Deichmann, R., Turner, R., & Ordidge, R. J. (2005). 3D MDEFIT imaging of the human brain at 4.7 T with reduced sensitivity to radiofrequency inhomogeneity. *Magnetic Resonance in Medicine*, 53, 1452–1458.
- Vallar, G. (2006). Memory systems: the case of phonological short-term memory. A festschrift for cognitive neuropsychology. *Cognitive Neuropsychology*, 23, 135–155.
- Vallar, G., Di Betta, A. M., & Silveri, M. C. (1997). The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. *Neuropsychologia*, 35, 795–812.
- Van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., et al. (2010). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, 7, 173–183.
- Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics*, 23, 361–374.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage*, 33, 316–325.
- Wilson, S. J., Lusher, D., Wan, C. Y., Dudgeon, P., & Reutens, D. C. (2008). The neurocognitive components of pitch processing: Insights from absolute pitch. *Cerebral Cortex*, 19, 724–732.
- Wise, R. J. S., Scott, S. K., Blank, C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain*, 124, 83–95.
- Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, 28, 995–1006.
- Zatorre, R. J. (2001). Neural specializations for tonal processing. *Annals of the New York Academy of Sciences*, 930, 193–210.