

# Neuroimaging of phonetic perception in bilinguals\*

NARLY GOLESTANI  
University of Geneva

(Received: September 15, 2014; final revision received: July 3, 2015; accepted: August 7, 2015; first published online 6 October 2015)

*This review addresses the cortical basis of phonetic processing in bilinguals and of phonetic learning, with a focus on functional magnetic resonance imaging studies of phonetic perception. Although results vary across studies depending on stimulus characteristics, task demands, and participants' previous experience with the non-native/second-language sounds, taken together, the literature reveals involvement of overlapping brain regions during phonetic processing in the first and second language of bilinguals, with special involvement of regions of the dorsal audio-motor interface including frontal and posterior cortices during the processing of new, or 'difficult' speech sounds. These findings converge with the brain imaging literature on language processing in bilinguals more generally, during semantic and syntactic processing of words and of connected speech. More brain imaging work can serve to better elucidate the precise mechanisms underlying phonetic encoding and its interaction with articulatory processes, in particular where multiple phonetic repertoires have been or are being acquired.*

Keywords: neuroimaging, phonetic, perception, bilinguals, non-native, learning

## Introduction

When a new language is proficiently learned in early or late bilinguals, novel phonetic categories are established (MacKain, Best & Strange, 1981), although there are individual differences in how well these sounds are perceived and produced (Flege, Munro & MacKay, 1995, Pallier, Bosch & Sebastián-Gallés, 1997, Bosch, Costa & Sebastián-Gallés, 2000, Sebastián-Gallés, Rodríguez-Fornells, Deigo-Balaguer & Díaz, 2006). Laboratory training studies also show that adults can learn to hear and to produce foreign speech sounds, again with large individual differences (Golestani & Zatorre, 2009, Hattori & Iverson, 2009, Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2015). This review will address the cortical basis of phonetic processing in bilinguals and of phonetic learning, with a focus on functional magnetic resonance imaging (fMRI) studies of phonetic perception. An overview of the neural basis of phonetic processing per se will precede the review of the bilingual and phonetic learning literature.

## Cortical bases of phonetic processing

Functional brain imaging studies using methods such as PET and fMRI in adults have examined the neural

underpinnings of phonetic perception using words, speech syllables, and meaningless speech sounds, and using passive listening, phoneme monitoring, discrimination, identification, and rhyming tasks. Several existing papers offer well-established models of the neural underpinnings of language processing and learning more generally (Hickok & Poeppel, 2007, Rodríguez-Fornells, Cunillera, Mestres-Misse & Deigo-Balaguer, 2009, Price, 2012). With respect to phonetic processing specifically, these models highlight the role of the dorsal audio-motor interface, or of the dorsal stream, including auditory, frontal and parietal regions, in mapping sounds onto articulatory-based representations (Hickok & Poeppel, 2007, Rodríguez-Fornells et al., 2009). This network is especially relevant for phonological processing and working memory (Aboitiz, 2012), in contrast with the ventral stream, which is thought to be more implicated in lexical processing and in processing meaning, or semantics (Hickok & Poeppel, 2007, Rodríguez-Fornells et al., 2009). Within the dorsal audio-motor network, the left pars opercularis, which lies in the posterior portion of Broca's area, and the adjacent left insula/frontal operculum (FO) of the left inferior frontal gyrus (LIFG) are involved even during purely receptive (i.e., perceptual) phonetic tasks when there are specific task demands such as phonetic segmentation and analysis (Démonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, Rascol & Frackowiak, 1992, Zatorre, Evans, Meyer & Gjedde, 1992, Fiez, Raichle, Miezin, Petersen, Tallal & Katz, 1995, Poldrack, Wagner, Prull, Desmond, Glover &

\* This work was supported by the Swiss National Science Foundation (PP00P3\_133701). I would like to thank Christophe Pallier and an anonymous reviewer, who provided helpful comments on this review.

Address for correspondence:

Brain and Language Lab, Department of Clinical Neuroscience, Campus Biotech, 9 Chemin des Mines, 1202 Genève, Switzerland.

[Narly.Golestani@unige.ch](mailto:Narly.Golestani@unige.ch)

Garbrieli, 1999, Burton, Small & Blumstein, 2000, Golestani & Zatorre, 2004). The left pars opercularis and the left supramarginal gyrus (SMG) are implicated in verbal working memory, or in the phonological loop, with the left pars opercularis and the adjacent left premotor area being involved in subvocal rehearsal, and the left SMG being involved in phonological storage (Paulesu, Frith & Frackowiak, 1993, Smith, Jonides, Marshuetz & Koeppel, 1998, Nixon, Lazarova, Hodinott-Hill, Gough & Passingham, 2004, Koelsch, Schulze, Sammler, Fritz, Mueller & Gruber, 2009). The implication of left motor cortex in addition to premotor regions during phonetic perception is thought to reflect subvocal articulatory demands (Pulvermuller, Huss, Kherif, Martin, Hauk & Shtyrov, 2006, Lee, Turkeltaub, Granger & Raizada, 2012, Rogers, Mottonen, Boyles & Watkins, 2014), in line with the motor theory of speech perception (Liberman & Mattingly, 1985).

The bilateral auditory cortex activations observed in the superior temporal gyrus (STG) during phonetic perception are typically localized to secondary auditory cortices anterior and posterior to Heschl's gyrus (HG), including the planum temporale (PT) (Binder, Rao, Hammeke, Yetkin, Jesmanowicz, Bandettini, Wong, Estkowski, Goldstein, Haughton & Hyde, 1994, Jancke, Shah, Posse, Grosse-Ryken & Muller-Gartner, 1998, Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman & Possing, 2000, Hickok & Poeppel, 2000, Kilian-Huetten, Valente, Vroomen & Formisano, 2011). However, these regions are also involved in processing complex sounds such as amplitude modulated noise (Giraud, Lorenzi, Ashburner, Wable, Johnsrude, Frackowiak & Kleinschmidt, 2000) as well in the analysis of spectral and temporal information more generally (Oleser, Eisner & Kotz, 2008, Santoro, Moerel, De Martino, Goebel, Ugurbil, Yacoub & Formisano, 2014), whereas earlier, primary auditory regions respond preferentially to simpler stimuli such as pure tones (Wessinger, VanMeter, Tian, Van Lare, Pekar & Rauschecker, 2001). When the processing of complex auditory (i.e., non-phonetic) stimuli is controlled for, or when across category phonetic conditions are compared to within category ones, phonetic perception is localised to the more downstream left middle/anterior superior temporal sulcus (STS) (Liebenthal, Binder, Spitzer, Possing & Medler, 2005) and to the adjacent left middle temporal gyrus (Zhang, Xi, Xu, Shu, Wang & Li, 2011), respectively. This latter study, which investigated lexical tonal stimuli in native speakers of Chinese, and other studies having examined the learning of lexical tone in people who were not native speakers of Chinese (Wong, Perrachione & Parrish, 2007), demonstrate convergence in terms of the left-lateralized neural underpinnings of lexical tone processing and of phonetic processing in non-tonal languages. Consistent with the hierarchical

view that more downstream regions respond to phonetic information per se, it has been proposed that speech perception is robust due to the presence of multiple, complementary representations of the input, which operate both on acoustic-phonetic features but also in articulatory-gestural domains (Scott & Johnsrude, 2003, Oleser, Leaver, VanMeter & Rauschecker, 2010). Bilateral temporal regions are involved in the processing of phonology, and higher levels of linguistic information in the speech signal (e.g., semantics, syntax) are processed in higher-level, left-lateralized frontal and parietal association cortices (Scott & Johnsrude, 2003, Peelle, 2012). Interestingly however, recent electrical recordings in humans (electrocorticography, or ECoG) during surgical planning have shown neural response patterns within the posterior STG (pSTG) which correspond to phonetic category boundaries (Chang, Rieger, Johnson, Berger, Barbaro & Knight, 2010), and to the speech sound features which map onto particular articulatory dimensions (Mesgarani, Cheung, Johnson & Chang, 2014). In other words, the pSTG does more than process spectro-temporal information in complex auditory input, and is likely also engaged in functional interaction with higher-level frontal and parietal regions that are involved in the categorical perception (CP) of speech sounds, an idea that is supported by recent developmental fMRI work on CP (Conant, Liebenthal, Desai & Binder, 2014), and by fMRI adaptation (Raizada & Poldrack, 2007) and pattern classification studies on CP (Lee et al., 2012). Similarly, the adjacent left temporo-parietal junction (area Spt) is thought to be involved in the interface, or mapping between sensory and motor representations during speech processing (Hickok & Poeppel, 2007). Finally, there is growing evidence for involvement of partially overlapping frontal (i.e., Broca's area) and posterior (i.e., Wernicke's area) brain regions classically associated with speech production and perception, respectively, during phonological and speech perception and production (Paus et al., 1996, Buchsbaum, Hickok & Humphries, 2001, Heim, Opitz, Muller & Friederici, 2003, Hickok & Poeppel, 2007, Meister, Wilson, Deblieck, Wu & Iacoboni, 2007, Price, Crinion & Macsweeney, 2011, Agnew et al., 2013), lending further support to the idea of interdependency of phonetic perception and production in the human brain.

### ***Functional brain imaging studies on bilingual phonetic processing and on phonetic learning***

#### ***Studies involving words***

In an early PET study in late, proficient bilinguals, overlapping activations were observed in regions including the pars triangularis and the pars orbitalis of the LIFG in the first (L1) and second language (L2) during rhyme and synonym generation tasks, where

phonological and semantic cues guided word selection, respectively (Klein, Milner, Zatorre, Meyer & Evans, 1995). These frontal regions are more typically associated with semantic processing and memory (Binder, Frost, Hammeke, Rao & Cox, 1996, Dapretto & Bookheimer, 1999, Liebenthal, Desai, Ellingson, Ramachandran, Desai & Binder, 2010) than with phonetic processing, which is more typically localized to the left pars opercularis (Poldrack, Wagner, Prull, Desmond, Glover & Gabrieli, 1999). The implication of semantic regions also during phonologically guided word retrieval might be expected given that a word generation task was used where semantic and lexical processes are likely also at play, especially when new words are generated. The findings of this study were interpreted as reflecting shared neural representations during phonetic and also semantic processing, in proficient bilinguals (Klein et al., 1995).

In a later longitudinal fMRI study on phonetic learning, minimal word pairs were used to test and to train Japanese individuals to hear the /r/ - /l/ contrast. These participants had previously been extensively exposed to this contrast, during 6 years of English-language instruction. After training, increased activation was found in regions including the bilateral superior temporal gyrus/sulcus (STG/STS), IFG, insula, SMG, premotor cortex, supplementary motor area and subcortical regions. It was proposed that these increases reflect the acquisition of auditory-articulatory mappings for the difficult /r-l/ contrast, in particular since this network was broader than that observed during perception of an easy phonetic contrast (/b-g/) (Callan, Tajima, Callan, Kubo, Masaki & Akahane-Yamada, 2003). Given that training was extensive and that it involved words, the functional plasticity results could in part have arisen from changes in semantic processing. It is interesting, however, that activation in primary and secondary auditory areas was also increased after training, reflecting functional plasticity in relatively low-level auditory regions (Callan et al., 2003). More generally, greater overall activation during perception of the difficult compared to the easy contrast is consistent with the idea of greater neural recruitment during effortful task performance, an explanation that has been offered for bilingual language processing more generally, in particular in the left IFG (Frith, Friston, Liddle & Frackowiak, 1991, Chee, Hon, Lee & Soon, 2001, Golestani & Zatorre, 2004, Golestani, Alario, Meriaux, Le Bihan, Dehaene & Pallier, 2006). However, the above studies did not isolate phonetic processing per se, and as such the interpretation of the findings is limited.

#### ***Studies having used isolated phonemes or syllables***

Studies on bilingual phonetic processing and on phonetic learning having used isolated phonemes or syllables converge with the idea that phonetic processing in L1

and L2 generally overlap, with greater neural recruitment during non-native, or effortful phonetic processing. For example in a magnetoencephalographic study on preattentive neural responses to stimulus change, English and Japanese listeners were tested during exposure to the /ra/ and /la/ syllables. The processing of non-native speech sounds in the Japanese group recruited greater neural resources and was associated with longer periods of brain activation in bilateral superior temporal and inferior parietal regions (Zhang et al., 2005).

Other phonetic perception studies have required active task performance. In one such fMRI study, native (English) and non-native (Japanese) listeners identified syllables starting with /r/ and /l/ (Callan, Jones, Callan & Akahane-Yamada, 2004). The Japanese listeners had previously studied English for at least 6 years, and accordingly, they performed above chance on this task, but still more poorly than the English participants. In line with the above-described longitudinal study by the same group (Callan et al., 2003), brain imaging revealed greater activation in the non-native listeners in an articulatory-auditory network comprising Broca's area, the anterior insula, the anterior STS/STG, the PT, the temporo-parietal junction, the SMG and the cerebellum, once again consistent with greater neural recruitment during more effortful, non-native phonetic processing. There was also a weak, positive correlation between performance on the /r-/ /l/ contrast and activation in the above-reported network in the non-native listeners (Callan et al., 2004). In other words, between groups, higher activation was associated with poorer performance (i.e., in the non-native compared to native listeners), but within the non-native (Japanese) group, the opposite was observed.

In line with the above-described study (Callan et al., 2004) and with the related longitudinal study by the same group (Callan et al., 2003), a second longitudinal study also found greater recruitment of auditory and articulatory brain regions after learning to hear a difficult non-native phonetic contrast (Golestani & Zatorre, 2004). In this latter study, listeners were trained to hear the difficult dental-retroflex contrast. After training, the pattern of brain activation came to resemble that observed during identification of a native contrast, with greater recruitment of the left IFG, the right insula / FO, the STG bilaterally and the left caudate nucleus (Golestani & Zatorre, 2004). There was also a positive relationship between behavioural improvement and post-training brain activation in the left angular gyrus, as well as a negative relationship between improvement and activation in the left insula/FO. This latter result suggests that the degree of success in phonetic learning is accompanied by more efficient neural processing in frontal speech regions implicated in phonetic processing, and conversely, that more effortful processing in the poorer learners is accompanied by greater recruitment of the left insula/FO

(Golestani & Zatorre, 2004). The negative correlation with performance is in the opposite direction to that found in this and other brain regions by Callan and colleagues (2004). One factor that could explain the discrepancy is that in Golestani and Zatorre (2004), participants were completely naïve to the contrast before training, and after 5 hours of training, only about half of the participants performed above chance (Golestani & Zatorre, 2004), whereas in the study by Callan and colleagues (2004), all the Japanese participants performed above chance even before scanning.

This raises the important question of the interaction between performance/effort and the degree of neural recruitment of relevant brain regions. Specifically, it is likely that some individuals can easily hear the contrast, that others can do so but with difficulty (i.e., with uncertainty and effort), and that yet others cannot hear it at all. In this latter subgroup, due to perceptual assimilation of non-native with native sounds, one can expect that participants eventually make less effort (i.e., they might give up on performing the task), and one can also expect greater neural adaptation in these individuals (Grill-Spector & Malach, 2001) due to the fact that they effectively hear the same sound across different trials. Such differences across individuals and also across studies (e.g., related to aptitudes, but also to previous exposure to the contrast of interest) might modulate the observed neural response in brain regions involved, resulting in discrepancies across studies in terms of the direction of the training effects, and in terms of the direction of correlations between activation and performance.

Interestingly, an electroencephalography (EEG) study has uncovered an important finding in relation to individual differences in phonetic perception. Using a pre-attentive oddball paradigm on vowels, it was found that good and poor phonetic perceivers differed in their electrophysiological response indexing change detection (i.e., the mismatch negativity, or MMN response) not only to non-native but also to native phonetic contrasts (Díaz, Baus, Escera, Costa & Sebastián-Gallés, 2008). In other words, people who are particularly good or poor in non-native vowel perception also differ in their neural response to native vowel contrasts. This finding may arise from the partially shared neural resources underlying L1 and L2 phonetic processing (Golestani & Zatorre, 2004), and suggests that there exist individual differences even in how native speech sounds are perceived, at least in bilinguals. This could in part be due to the influence of learning a new phonetic inventory on characteristics of the native inventory (Chang, 2012, Kartushina, Hervais-Adelman, Frauenfelder & Golestani, unpublished manuscript). Possibly related to a relationship between L1 and L2 phonetic perception is recent behavioural evidence for a relationship between L1 and L2 phonetic production (Kartushina and Frauenfelder, 2014).

The studies reviewed thus far reported results of univariate analyses, and generally converge in showing greater recruitment of frontal and/or of posterior brain regions during the processing of new or of difficult speech sounds. Different, complementary results have been obtained using multi-voxel pattern analysis (MVPA, aka ‘pattern classification’), which is better suited for differentiating neural representations within spatially overlapping brain regions. In one such study, English and Japanese listeners were tested on their perception of the /r/ - /l/ distinction. It was found that the statistical separability of fMRI activation patterns in the right primary auditory cortex predicted subjects’ ability to tell the sounds apart, both across and within groups (Raizada, Tsao, Liu & Kuhl, 2010). This result is consistent with functional brain imaging (Binder et al., 1994, Jancke et al., 1998, Binder et al., 2000, Hickok & Poeppel, 2000, Kilian-Huetten et al., 2011) and with electrocorticography studies showing temporal cortex involvement during phonetic processing (Chang et al., 2010, Mesgarani et al., 2014), and demonstrates that further work is needed involving more fine-grained analyses of differences in neural recruitment within spatially overlapping brain regions. This opens the question of the contributions of top-down versus bottom-up influences on auditory cortex activation differences in relation to phonetic processing.

A recent adaptation fMRI study partially addressed this question (Myers & Swan, 2012). Involvement of temporal and inferior frontal brain regions was shown in phonetic processing (Myers & Swan, 2012), and additionally, the bilateral middle frontal gyri were implicated specifically during the processing of a newly learned phonetic category. This suggests that top-down information about new categories may reshape perceptual sensitivities via attentional or executive mechanisms (Myers & Swan, 2012), and demonstrates that there is a complex interplay between low-level, perceptual aspects of the input and higher-level knowledge about phonetic categories, in particular when they are newly learned. Related to this are the results of a longitudinal training study with synthetic, phonetic and non-speech but voice-like continua, which showed that the left posterior STS may play a role in the short-term representation of sound features relevant for learning new sound categories (Liebenthal et al., 2010). This provides evidence for a lower-level, temporal cortex mechanism that may mediate subsequent consolidation during the learning of novel speech sounds.

### Conclusions and future reading

In conclusion, although a limited number of functional imaging studies have examined the neural underpinnings of bilingual phonetic processing per se, the results of these studies generally converge in showing overlapping brain regions during phonetic processing in the L1 and

L2 of bilinguals, with greater recruitment of frontal and posterior brain regions during the processing of new or of 'difficult' non-native sounds. This converges with findings on bilingual language processing more generally, where it has been shown that at early stages of L2 learning there is relatively greater engagement of anterior and parietal portions of the language network including Broca's area as well as of higher level executive and language control regions, and that, as increased proficiency is attained in the second language, the two languages recruit more overlapping brain networks (Indefrey, 2006, Abutalebi, Cappa & Perani, 2001, Stowe & Sabourin, 2005, Abutalebi, 2008, Sebastian, Laird & Kiran, 2011). Further, studies having examined the question of phonetic perception and learning per se using univariate approaches and at the macroscopic level using fMRI suggest that largely overlapping regions of the auditory cortex are recruited when processing familiar versus novel speech sounds, or when processing different speech sounds of one language. More advanced image analysis methods (i.e., MVPA) and invasive approaches such as intracranial recordings, however, reveal differences in the neural response pattern within overlapping regions of auditory cortex in response to L1 versus L2 speech sounds, and also in relation to specific phonetic features such as place of articulation, and in relation to cross versus within category differences (i.e., categorical perception). These more fine-grained auditory cortex differences, which are likely modified during the acquisition of new speech sounds, are likely mediated a) by regions including the left middle to posterior STS in the short-term representation of sound features defining new sound categories; b) by increased involvement of the left temporo-parietal junction related to increased demands on sensori-motor mapping of the new sounds; and c) by the additional involvement of frontal brain regions in the top-down reshaping of lower-level, perceptual phonetic encoding in the auditory cortex. These findings are convergent with the known roles of these respective components of the dorsal audio-motor stream in spectro-temporal analysis (bilateral dorsal STG), in phonological processing (bilateral middle to posterior STS), in the sensori-motor interface (left temporo-parietal junction) and in subvocal articulation (posterior LIFG) (Hickok & Poeppel, 2007). Outstanding questions remain regarding the precise mechanisms underlying differential encoding of L1 versus L2 (or foreign) speech sounds in primary and secondary auditory cortices, in particular in light of interactions of these bottom-up, auditory processes with top-down, frontal and temporo-parietal ones. These can be addressed using, among other approaches, ultra-high resolution (i.e., 7 Tesla) functional mapping, advanced data analysis methods including MVPA and computational modelling, and invasive methods such as intracranial recordings.

Recommendations for further reading that relate to the neural bases of phonetic processing in bilingualism and to phonetic learning include developmental work on native and non-native speech sound processing in infants (Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho & Naatanen, 1998, Rivera-Gaxiola, Silva-Pereyra & Kuhl, 2005, Minagawa-Kawai, Mori, Naoi & Kojima, 2007, Petitto, Berens, Kovelman, Dubins, Jasinska & Shalinsky, 2012, Ortiz-Mantilla, Hamalainen, Musacchia & Benasich, 2013, Fava, Hull & Bortfeld, 2014), on foreign-language syllable production in children (Hashizume, Taki, Sassa, Thyreau, Asano, Asano, Takeuchi, Nouchi, Kotozaki, Jeong, Sugiura & Kawashima, 2014), and on the neural bases of lexical tone processing in individuals whose first language was tonal but was subsequently forgotten (Pierce, Klein, Chen, Delcenserie & Genesee, 2014). There is also a large electrophysiological (EEG and magnetoencephalography, or MEG) literature and some functional near infrared spectroscopy (fNIRS) work on the cortical and subcortical bases of phonetic perception and learning (Alain, Reinke, McDonald, Chau, Tam, Pacurar & Graham, 2005, Zhang, Kuhl, Imada, Iverson, Pruitt, Stevens, Kawakatsu, Tohkura & Nemoto, 2009, Kumar, Hegde & Mayaleela, 2010, Xi, Zhang, Shu, Zhang & Li, 2010, Zhang et al., 2011, Chandrasekaran, Kraus & Wong, 2012, Brandmeyer, Farquhar, McQueen & Desain, 2013, Kaan, Wayland & Keil, 2013, Skoe, Chandrasekaran, Spitzer, Wong & Kraus, 2014, Zinszer, Chen, Wu, Shu & Li, 2015). Also, given the growing evidence for the importance of syllable-level speech processing (Morillon, Liegeois-Chauvel, Amer, Bener & Giraud, 2012, Edwards & Chang, 2013, Doelling, Arnal, Ghitz & Poeppel, 2014), studies on the neural basis of bilingual phonotactic processing are recommended (Dehaene-Lambertz, Dupoux & Gout, 2000, Jacquemot, Pallier, Le Bihan, Dehaene & Dupoux, 2003, Minagawa-Kawai, Cristia, Long, Vendelin, Hakuno, Dutat, Filippin, Cabrol & Dupoux, 2013), although only a limited number of studies have addressed this.

Other literature that is relevant to bilingual phonetic processing and learning is a body of work on the brain structural correlates of individual differences in phonetic processing and also in language processing more generally (see Golestani, 2014, for a recent review). These include studies on the brain structural correlates of phonetic perception (Golestani, Paus & Zatorre, 2002, Golestani, Molko, Dehaene, Le Bihan & Pallier, 2007, Wong, Chandrasekaran, Garibaldi & Wong, 2011, Lebel & Beaulieu, 2009, Wong, Warrier, Penhune, Roy, Sadehh, Parrish & Zatorre, 2008, Sebastián-Gallés, Soriano-Mas, Baus, Díaz, Ressel, Pallier, Costa & Pujol, 2012, Burgaleta, Baus, Díaz & Sebastián-Gallés 2014) and production (Golestani & Pallier, 2007), on foreign speech imitation (Reiterer, Hu, Erb, Rota, Nardo, Grodd, Winkler & Ackermann, 2011), on bilingualism (Mechelli, Crinion,

Noppeney, O'Doherty, Ashburner, Frackowiak & Price, 2004, Ressel, Pallier, Ventura-Campos, Díaz, Roessler, Avila & Sebastián-Gallés, 2012, Klein, Mok, Chen & Watkins, 2014) and on expertise in phonetics (Golestani, Price & Scott, 2011).

## References

- Aboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Frontiers in Evolutionary Neuroscience*, 4, 2.
- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica (Amst)*, 128, 466–78.
- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4, 179–190.
- Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *Neuroimage*, 73, 191–9.
- Alain, C., Reinke, K., McDonald, K. L., Chau, W., Tam, F., Pacurar, A., & Graham, S. (2005). Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage*, 26, 592–599.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–528.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119, 1239–1247.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., Wong, E. C., Estkowski, L. D., Goldstein, M. D., Houghton, V. M., & Hyde, J. S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, 35, 662–72.
- Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12, 189–221.
- Brandmeyer, A., Farquhar, J. D. R., McQueen, J. M., & Desain, P. W. M. (2013). Decoding Speech Perception by Native and Non-Native Speakers Using Single-Trial Electrophysiological Data. *PLoS One*, 8:e68261.
- 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.
- Burgalata, M., Baus, C., Díaz, B., & Sebastián-Gallés, N. (2014). Brain structure is related to speech perception abilities in bilinguals. *Brain Structure & Function*, 219, 1405–1416.
- Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679–90.
- 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.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, 19, 113–124.
- Chandrasekaran, B., Kraus, N., & Wong, P. C. M. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, 107, 1325–1336.
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40, 249–268.
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13, 1428–U169.
- Chee, M. W. L., Hon, N., Lee, H. L., & Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *Neuroimage*, 13, 1155–1163.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Naatanen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351–353.
- Conant, L. L., Liebenthal, E., Desai, A., & Binder, J. R. (2014). fMRI of phonemic perception and its relationship to reading development in elementary- to middle-school-age children. *Neuroimage*, 89, 192–202.
- Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, 24, 427–432.
- Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, 12, 635–47.
- Démonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753–68.
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 16083–16088.
- Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, 85, 761–768.
- Edwards, E., & Chang, E. F. (2013). Syllabic (approximately 2–5 Hz) and fluctuation (approximately 1–10 Hz) ranges in speech and auditory processing. *Hearing Research*, 305, 113–34.
- Fava, E., Hull, R., & Bortfeld, H. (2014). Dissociating Cortical Activity during Processing of Native and Non-Native Audiovisual Speech from Early to Late Infancy. *Brain Sciences*, 4, 471–87.

- Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET Studies of Auditory and Phonological Processing: Effects of Stimulus Characteristics and Task Demands. *Journal of Cognitive Neuroscience*, 7, 357–75.
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97, 3125–34.
- Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. J. (1991). A PET study of word finding. *Neuropsychologia*, 29, 1137–1148.
- Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, 84, 1588–1598.
- Golestani, N. (2014). Brain structural correlates of individual differences at low-to high-levels of the language processing hierarchy: A review of new approaches to imaging research. *International Journal of Bilingualism*, 18, 6–34.
- Golestani, N., Alario, F. X., Meriaux, S., Le Bihan, D., Dehaene, S., & Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia*, 44, 1029–40.
- Golestani, N., Molko, N., Dehaene, S., Le Bihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, 17, 575–82.
- Golestani, N., & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*, 17, 929–34.
- Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, 35, 997–1010.
- Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an Ear for Dialects? Structural Plasticity in the Expert Phonetician Brain. *Journal of Neuroscience*, 31, 4213–4220.
- 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.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107, 293–321.
- Hashizume, H., Taki, Y., Sassa, Y., Thyreau, B., Asano, M., Asano, K., Takeuchi, H., Nouchi, R., Kotozaki, Y., Jeong, H., Sugiura, M., & Kawashima, R. (2014). Developmental changes in brain activation involved in the production of novel speech sounds in children. *Human Brain Mapping*, 35, 4079–89.
- Hattori, K., & Iverson, P. (2009). English vertical bar r vertical bar-vertical bar l vertical bar category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, 125, 469–479.
- Heim, S., Opitz, B., Muller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*, 16, 285–296.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.
- Hickok, G., & Poeppel, D. (2007). Opinion - The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning*, 56, 279–304.
- Jacquemot, C., Pallier, C., Le Bihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 23, 9541–9546.
- Jancke, L., Shah, N. J., Posse, S., Grosse-Ryken, M., & Muller-Gartner, H. W. (1998). Intensity coding of auditory stimuli: an fMRI study. *Neuropsychologia*, 36, 875–883.
- Kaan, E., Wayland, R., & Keil, A. (2013). Changes in oscillatory brain networks after lexical tone training. *Brain Sciences*, 3, 757–80.
- Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, 5. <http://doi.org/10.3389/fpsyg.2014.01246>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U., & Golestani, N. (2015). The effect of production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America*, 138, 817–832.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U., & Golestani, N. (unpublished manuscript). Mutual influences between native and non-native vowels in production: evidence from short-term articulatory feedback training. *Journal of Phonetics*.
- Kilian-Huetten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory Cortex Encodes the Perceptual Interpretation of Ambiguous Sound. *Journal of Neuroscience*, 31, 1715–1720.
- Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation - a bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 2899–2903.
- Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: a cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, 131, 20–4.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Mueller, K., & Gruber, O. (2009). Functional Architecture of Verbal and Tonal Working Memory: An fMRI Study. *Human Brain Mapping*, 30, 859–873.
- Kumar, A. U., Hegde, M., & Mayaleela, (2010). Perceptual learning of non-native speech contrast and functioning of the olivocochlear bundle. *International Journal of Audiology*, 49, 488–496.
- Lebel, C., & Beaulieu, C. (2009). Lateralization of the Arcuate Fasciculus from Childhood to Adulthood and its Relation

- to Cognitive Abilities in Children. *Human Brain Mapping*, 30, 3563–3573.
- Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical Speech Processing in Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis. *Journal of Neuroscience*, 32, 3942–3948.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, 15, 1621–1631.
- Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the Left Superior Temporal Sulcus for Auditory Categorization. *Cerebral Cortex*, 20, 2958–2970.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /t/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369–390.
- Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: structural plasticity in the bilingual brain. *Nature*, 431 (7010), 757.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology: CB*, 17, 1692–6.
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, 343, 1006–1010.
- Minagawa-Kawai, Y., Cristia, A., Long, B., Vendelin, I., Hakuno, Y., Dutat, M., Filippin, L., Cabrol, D., & Dupoux, E. (2013). Insights on NIRS Sensitivity from a Cross-Linguistic Study on the Emergence of Phonological Grammar. *Frontiers in Psychology*, 4, 170.
- Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience*, 27, 315–321.
- Morillon, B., Liegeois-Chauvel, C., Amer, L. H., Bener, C.-G., & Giraud, A.-L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Frontiers in Psychology*, 3.
- Myers, E. B., & Swan, K. (2012). Effects of Category Learning on Neural Sensitivity to Non-native Phonetic Categories. *Journal of Cognitive Neuroscience*, 24, 1695–1708.
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The inferior frontal gyrus and phonological processing: An investigation using rTMS. *Journal of Cognitive Neuroscience*, 16, 289–300.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, 28, 8116–23.
- Obleser, J., Leaver, A. M., Vanmeter, J., & Rauschecker, J. P. (2010). Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Frontiers in Psychology*, 1, 232.
- Ortiz-Mantilla, S., Hamalainen, J. A., Musacchia, G., & Benasich, A. A. (2013). Enhancement of Gamma Oscillations Indicates Preferential Processing of Native over Foreign Phonemic Contrasts in Infants. *Journal of Neuroscience*, 33, 18746–18754.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–5.
- Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *The European Journal of Neuroscience*, 8, 2236–46.
- Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Frontiers in Human Neuroscience*, 6.
- Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The “Perceptual Wedge Hypothesis” as the basis for bilingual babies’ phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language*, 121, 130–143.
- Pierce, L. J., Klein, D., Chen, J. K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 17314–9.
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, 10, 15–35.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62, 816–47.
- Price, C. J., Crinion, J. T., & Macsweeney, M. (2011). A Generative Model of Speech Production in Broca's and Wernicke's Areas. *Frontiers in Psychology*, 2, 237.
- Pulvermuller, F., Huss, M., Kherif, F., Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 7865–7870.
- Raizada, R. D. S., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56, 726–740.
- Raizada, R. D. S., Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2010). Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences. *Cerebral Cortex*, 20, 1–12.
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., Winkler, S., & Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2, 271.
- Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Avila, C., & Sebastián-Gallés, N. (2012). An Effect of Bilingualism on the Auditory Cortex. *Journal of Neuroscience*, 32, 16597–16601.
- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts



- in 7- and 11-month-old American infants. *Developmental Science*, 8, 162–172.
- Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 3711–3735.
- Rogers, J. C., Mottonen, R., Boyles, R., & Watkins, K. E. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology*, 5, 754.
- Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Computational Biology*, 10, e1003412.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107.
- Sebastián-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, 18, 1277–1291.
- Sebastián-Gallés, N., Soriano-Mas, C., Baus, C., Díaz, B., Ressel, V., Pallier, C., Costa, A., & Pujol, J. (2012). Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics*, 25, 150–162.
- Sebastian, R., Laird, A. R., & Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Applied Psycholinguistics*, 32, 799–819.
- Skoe, E., Chandrasekaran, B., Spitzer, E. R., Wong, P. C. M., & Kraus, N. (2014). Human brainstem plasticity: The interaction of stimulus probability and auditory learning. *Neurobiology of Learning and Memory*, 109, 82–93.
- Smith, E. E., Jonides, J., Marshuetz, C., & Koeppel, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 876–882.
- Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, 43, 329–353.
- Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 13, 1–7.
- Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White Matter Anisotropy in the Ventral Language Pathway Predicts Sound-to-Word Learning Success. *Journal of Neuroscience*, 31, 8780–8785.
- 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.
- Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadeh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex*, 18, 828–836.
- Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170, 223–31.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846–9.
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical Dynamics of Acoustic and Phonological Processing in Speech Perception. *PLoS One*, 6.
- Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage*, 46, 226–240.
- Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *Neuroimage*, 26, 703–720.
- Zinszer, B. D., Chen, P., Wu, H., Shu, H., & Li, P. (2015). Second language experience modulates neural specialization for first language lexical tones. *Journal of Neurolinguistics*, 33, 50–66.