

Brain networks involved in accented speech processing

Mireia Hernández^{a,b}, Noelia Ventura-Campos^{c,d}, Albert Costa^{e,f}, Anna Miró-Padilla^c,
and César Ávila^c

^a Section of Cognitive Processes, Department of Cognition, Development, and Educational Psychology, University of Barcelona, Barcelona, Spain.

^b Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Spain.

^c Neuropsychology and Functional Imaging Group, Department of Basic Psychology, Clinical Psychology, and Psychobiology, Universitat Jaume I, Castellón, Spain.

^d Department of Education and Specific Didactics, Universitat Jaume I, Castellón, Spain.

^e Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain.

^f Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Abstract

We investigated the neural correlates of accented speech processing (ASP) with an fMRI study that overcame prior limitations in this line of research: we preserved intelligibility by using two regional accents that differ in prosody but only mildly in phonetics (Latin American and Castilian Spanish), and we used independent component analysis to identify brain networks as opposed to isolated regions. ASP engaged a speech perception network composed primarily of structures related with the processing of prosody (cerebellum, putamen, and thalamus). This network also included anterior fronto-temporal areas associated with lexical-semantic processing and a portion of the inferior frontal gyrus linked to executive control. ASP also recruited domain-general executive control networks related with cognitive demands (*dorsal attentional* and *default mode networks*) and the processing of salient events (*saliency network*). Finally, the reward network showed a preference for the native accent, presumably revealing people's sense of social belonging.

Keywords: accented speech; brain networks; fMRI; ICA; regional accent; speech perception

1. Introduction

At the height of fierce battles in World War II, a British soldier meets an Australian soldier in a trench.

British soldier: “Oh! Did we all come here to die?”

Australian soldier: “Nah, we arrived yesterday.”

The butt of this popular joke is how, to the ears of the Australian soldier, the British soldier pronounces the phrase “to die” as the word “today”, reflecting differences in the phonetic repertoires of the two speakers. Our aim here is to study the neural correlates of accented speech processing (ASP) by assessing differences in neural activity when listening to a non-native regional accent (Latin American Spanish), as compared with a native accent (Castilian Spanish), where regional accent is understood to be a distinct form of a language spoken in different geographical sites. This design allows us to complement prior neuroimaging studies on ASP and, as we argue below, to go beyond some of their potential shortcomings.

There is some consensus that ASP broadly engages the neural circuit elicited when listening to speech under adverse conditions (e.g., speech in background noise, time-compressed speech, etc.). In this respect, it is as if ASP resembles speech processing when the signal is deteriorated. Importantly, ASP recruits auditory and speech planning areas to a greater extent than does native speech processing (NSP) (see Adank, Nuttall, Banks, & Kennedy-Higgins, 2015 for a review). Auditory regions involve the bilateral auditory association cortex—large parts of the temporal gyrus and sulcus (STG/STS; BA22), including Wernicke’s area in the dominant side and extending to a posterior portion of the middle temporal gyrus (MTG; BA21). Planning regions involve the cerebellum, portions of the somatosensory cortex, and frontal areas, including the inferior frontal

gyrus (IFG; BA44, 45, 47), insula, supplementary motor area (SMA), and portions of the motor and premotor cortices. These areas are thought to interact to help find the best match between the speaker's and the listener's phonetic templates.

However, a closer look at the six prior studies on ASP (Adank, Davis, & Hagoort, 2012; Adank, Noordzij, & Hagoort, 2012; Adank, Rueschemeyer, & Bekkering, 2013; Bestelmeyer, Belin, & Ladd, 2015; Callan, Callan, & Jones, 2014; Yi, Smiljanic, & Chandrasekaran, 2014) reveals quite considerable heterogeneity in the results (Figure 1). Compared to NSP, ASP only recruited auditory association regions in Adank et al. (2012a), whereas the studies by Adank et al. (2012b, 2013) also observed the recruitment of frontal speech planning regions. Two other studies have found that, relative to NSP, ASP recruited speech planning regions plus regions typically involved in domain-general executive control (EC) (Callan, et al., 2014; Yi et al., 2014). Finally, Bestelmeyer et al. (2015) found no differences in neural activity during ASP compared to NSP.

=====
Figure 1 around here
=====

This heterogeneity probably stems from differences in the specific conditions used in each experiment¹. Consider, for example, the involvement of auditory association regions. This observation comes from those studies (Adank et al., 2012a,b, 2013) in which ASP always involved reduced intelligibility (similar to the effects of background noise at +2 decibels, as measured by the authors). Thus, the involvement of auditory association regions in these experiments might be more attributable to reduced intelligibility (Davis & Johnsrude, 2003; Hickok & Poeppel, 2007) than to ASP per se. Indeed, when intelligibility is not compromised, neural activity in auditory association regions is similar in ASP and NSP (Bestelmeyer et al., 2015; Callan et al., 2014; Yi et al., 2014).

Another source of controversy relates to the interpretation given to the involvement of frontal speech planning regions in ASP, especially Broca's area (BA44, 45). According to some authors, this involvement does not reflect speech planning processes but, rather, the involvement of domain-general EC processes typically engaged in difficult processing situations (Venezia, Saberi, Chubb, & Hickok, 2012). This interpretation would be consistent with the observation that people's ability to adapt to novel accents correlates with their EC capacities (Adank & Janse, 2010; Banks, Gowen, Munro, & Adank, 2015). One way to better understand the role of these frontal regions is to establish the networks in which these regions are recruited during ASP: that is, whether they are recruited as part of linguistic networks, executive control networks, or both. However, all studies to date have used traditional analyses based on the general linear model (GLM) (Friston et al., 1995), and thus the researchers could only detect isolated regions involved in ASP, as opposed to entire networks.

It should also be noted that ASP and NPS seem to elicit a different emotional response. In this regard, Bestelmeyer et al. (2015) observed higher activity in the amygdala during NSP than during ASP, presumably due to in-group versus out-group biases. At present, however, the study by Bestelmeyer et al. is the only one to provide evidence about differences between NSP and ASP in the modulation of emotionally-related areas.

1.1. The current study

We assessed the brain networks associated with ASP in a design with the following properties.

Equating for intelligibility. To this end, we used two regional variations of Spanish: Castilian (the variant spoken in northern-east central Spain) and Latin American (the variant spoken in Caribbean regions). Participants had the Castilian variant as their own. The difference between these two variants of Spanish is roughly comparable to those

between the English varieties of London and Chicago. In terms of pronunciation, abrupt phonetic contrasts are not encountered between Castilian and Latin American Spanish. The most evident phonetic difference between the two variants is the use of the phonemes /s/ and /θ/ to pronounce the consonant [c]: the phoneme /θ/ only exists in Castilian, which means that native speakers of Latin American Spanish replace it with /s/: for instance, *gra/s/ias* instead of *gra/θ/ias* to pronounce the word *gra[c]ias* (the Spanish word for “thanks”). The critical point in this regard is that the phoneme /s/ exists in both dialects, which means that native speakers of Castilian (our participants in this study) can easily identify it as part of their phonetic repertoire. It is only the fact of encountering /s/ in a word where it is not supposed to be (e.g., *gra/s/ias*) that makes native speakers of Castilian perceive it as phonetically deviant. Hence, differences in intelligibility at the phonetic level between the two variants are minimal. The greatest difference between these two varieties relates to prosody—suprasegmental properties of speech such as syllable duration and sentence intonation. However, this kind of variation does not hamper lexical recognition and, hence, it does not significantly affect intelligibility: TV shows and movies are broadcast in either Spain or Latin American countries with no need for dubbing². It should also be noted here that Castilian and Latin American Spanish do differ in a few lexical forms: for example, the word “carro” instead of “coche” is used in Latin American to refer to “car”. However, the interest of the present study was restricted to pronunciation and prosody. Therefore, we avoided differences at the word-form level in our materials.

Naturalistic experimental setting. To create a naturalistic setting we asked participants to watch two different types of movie clips that critically differed in the Spanish variant in which they were dubbed (Castilian vs. Latin American); we also included a baseline condition consisting of clips dubbed in a language that participants could not understand

(Dutch). The choice of an audiovisual format (as opposed to auditory-only materials) was made in order to use a common real-life context, as is watching soaps, movies, or TV shows today. Also in line with a naturalistic scenario, we did not focus on specific phonetic or prosodic deviations of Latin American relative to Castilian Spanish: we simply allowed the type and frequency of such deviations to occur randomly throughout the clips.

Exploring brain networks by means of independent component analyses (ICA). Unlike previous studies that used GLM-based analyses, we used ICA (see Calhoun, Liu, & Adali, 2009 for a review), an approach which allowed us to detect the brain networks involved in ASP, rather than only isolated regions. This is because, unlike GLM-based analyses, ICA can identify a brain region as a component of a different network (Xu, Potenza, Calhoun, 2013). GLM-based analyses are blind to this possibility because they cannot detect certain types of changes in the neural signal such as transient task-related, slow varying modulations (Calhoun et al., 2009). For the sake of completeness, we also performed GLM-based analyses, which we will describe and discuss only briefly in the supplementary material.

2. Materials and methods

2.1. Participants

Thirty native speakers of Castilian, who have always lived in northern-east Spain (within the region of Valencia), took part in this study (21 female; mean age = 22.97, SD = 2.31; mean years of education = 14.67, SD = 1.18). We did not include any participants who had seen *Alice in Wonderland* (Burton, 2010), the movie from which the clips were extracted. All participants reported being right-handed. The mean score on the Edinburgh Handedness Inventory (Oldfield, 1971) was 19.8 (SD = 2.91, range: 14-25; scores on this scale range from 10 to 50, with lower values for strong right-handedness).

No participant reported neurological or psychiatric disorders. They all received 20 EUR for participating in this study. All procedures were approved by the ethics committee of the *Universitat Jaume I* (Spain).

Linguistic background. All participants had Castilian Spanish as a native and dominant language and Valencian as a second language, the latter being typologically similar to Castilian (both are Romance languages). The two languages have co-official status in the region of Valencia, with Castilian having a more prominent presence in social, academic, and official contexts. Participants' proficiency level in Valencian varied from native-like to intermediate levels. Additionally, most participants (26 of them) had English as a foreign language. Ten of these 26 participants had a second foreign language (besides English), which was French for nine of them and German for the remaining one. The most common proficiency level in these foreign languages (English, French, and German) was intermediate, with only a minority of participants having achieved an upper-intermediate or advanced level (Table 1). Participants were required to estimate the percentage of exposure they had on a daily basis—in family, work, and social contexts—to each of the languages they knew, including Latin American Spanish (the percentages given had to sum 100%). They were most frequently exposed to Castilian (60.93%, SD = 15.82), followed by Valencian (30.4%; SD = 14.47). The amount of exposure to foreign languages was modest (English: 6.17%, SD = 6.44; French/German: 0.2%; SD = 0.92). Importantly, participants had little exposure to Latin American Spanish (2.3%, SD = 5.94). They were also asked to estimate the percentage of series, movies, and TV shows they watched in different languages (as before, the percentages had to sum 100%): they watched most series, movies, and TV shows in either Castilian (79.5%, SD = 23.9) or English (16.7%, SD = 7.6), and only 3.8% (SD = 7.6) of them in Latin American Spanish.

=====
Table 1 around here
=====

It is important to highlight here that participants' perception of Latin American Spanish was unlikely to be affected by their knowledge of other languages besides Castilian (i.e., Valencian, English, French, or German). The fact that only a few participants had knowledge of French and German, with a rather low level of proficiency and little exposure to them, led us to dismiss any concerns about these languages. Regarding Valencian and English, these languages have a richer phonetic repertoire than does Spanish, whether Castilian or Latin American. For instance, both Valencian and English have the phonemes /z/ (like in “zero”), /ʒ/ (like in “gel”), and /ʃ/ (like in “shell”), whereas neither Castilian nor Latin American do. Therefore, no Valencian/English novel phoneme (with respect to Spanish) would help a native speaker of Castilian to perceive the deviant phonemes of Latin American as more native-like. For example, *gra/ʃ/ias* would be perceived as deviant from *gra[θ]ias* regardless of whether the person has knowledge of Valencian/English. In a similar vein, the fact that someone is familiar with the prosody of Valencian or English should not help him or her to perceive that of Latin American as less deviant. With respect to Valencian, this language is very similar to Castilian in terms of the relative duration of syllables and intonation, and no Valencian prosodic particularities resemble those of Latin American Spanish (see Prieto et al. (2013) for a detailed description of Valencian prosodic traits). As for English prosody, it is clearly different to that of Spanish, whether Castilian or Latin American (Prieto, Vanrell, Astruc, Payne, & Post, 2010). Indeed, neither variant of Spanish shows the most characteristic English traits, such as: (i) prominent stressed syllables being much longer than the unstressed ones, producing the auditory impression of a Morse-like rhythm; (ii) the marking of the most important word within a thought group by a major pitch change on

the stressed syllable of that word; and (iii) the speaker leaving the pitch up between thought groups to indicate that the sentence continues.

2.2. Stimuli and experimental design

Stimuli consisted of 21 clips from the movie *Alice in Wonderland* (Burton, 2010). Participants were assigned to one of the three existing dubbing lists, which varied in the language and/or accent in which the same clip had been dubbed: Castilian, Latin American, and Dutch (no participant had any knowledge of this language). The order of each specific clip was kept constant across lists, which means that the same scanning onset was used across language and/or accent. For instance, the scanning onset of 6 seconds was always used for movie clip 1, which was dubbed in Castilian in list 1, Dutch in list 2, and Latin American in list 3. Given that participants were native speakers of Castilian, those three dubbing languages/accents gave rise to two experimental conditions (native: Castilian; accented: Latin American) and a baseline condition (Dutch). There were seven clips per condition, each lasting 30 seconds. Two clips of the same condition were never presented consecutively. Clips were separated by a 12-second interval of rest (i.e., black screen). Participants were instructed to pay attention to each clip even if they could not understand the dialogs (which were those in the baseline condition).

Procedure. Participants watched the computer screen through MRI-compatible goggles and listened to the dialogues through MRI-compatible headphones. After the scanning session, participants answered a multiple-choice “movie quiz” about the content of the dialogues. There were four response choices, including a “don’t know” option. For example, Aunt Emogine said she was waiting for: (a) her daughter, (b) two friends of hers, (c) her fiancé, (d) “I don’t know”. There was one question per clip, and thus seven questions per condition (native, accented, and baseline), making a total of 21 questions. A separate score was computed for each condition (native, accented, and Dutch) as the

sum of correct responses (maximum score per condition = 7). Responses could not be guessed by the visual context of the clip. Thus, the “don’t know” response was the only correct option in the Dutch condition.

2.3. MRI scanning parameters

Functional MRI data were collected on a 1.5T Siemens Symphony MRI scanner (Erlangen, Germany). Participants were placed in a supine position in the MRI scanner and their heads were immobilized with cushions to reduce motion artifacts. Functional images were acquired using a T2*-weighted echo-planar MR sequence covering the entire brain (TR/TE = 2500/49 ms, matrix = 64x64x28, flip angle = 90°, voxel size = 3.5x3.5x4.48; slice thickness = 4 mm; slice gap = 0.48 mm). A total of 298 images were recorded. The slices were made parallel to the anterior-posterior commissure plane covering the entire brain. Before obtaining the functional magnetic resonance images, a high-resolution structural T1-weighted MPRAGE sequence was acquired (TR = 2200 ms, TE = 3 ms, flip angle = 90°, matrix = 256x256x160, voxel size = 1x1x1 mm).

2.4. Image preprocessing

Data were analyzed using SPM12 software (<http://www.fil.ion.ucl.ac.uk/spm/>). Functional data were slice time-corrected, realigned to the first volume, motion-corrected, normalized into standard stereotactic space using the Montreal Neurological Institute (MNI) template, and smoothed with a Gaussian kernel (8 mm FWHM). The time series were high-filtered to eliminate low-frequency components (filter 128 s).

2.5. Independent component analysis (ICA)

We performed a GLM analysis of the component time courses estimated by independent component analysis (ICA; Calhoun, Adali, Pearlson, & Pekar, 2001) to determine the different brain networks modulated by the experimental conditions (native, accented, and

baseline). We then carried out the second-level analyses, which consisted of within-group comparisons using the beta-weights obtained from the GLM.

To obtain the functional brain networks underlying the fMRI data we performed a group spatial ICA using the GIFT toolbox (<http://icatb.sourceforge.net>) and an Infomax algorithm (Bell & Sejnowski, 1995). A final number of 20 independent components (ICs) was reached with the minimum description length criteria (Li, Adali, & Calhoun, 2007). Twenty iterations of the ICA analysis were performed by ICASSO software (Himberg, Hyvärinen, & Esposito, 2004) to ensure the stability of the estimated ICs. After ICA decomposition, individual IC maps and time courses were then computed using the GICA-3 back-reconstruction approach (Calhoun et al., 2001; Erhardt et al., 2011). The voxel values for the individual maps represent their contribution to the component time course. Therefore, the brain regions that were significantly related with each component time course were determined in the whole group through one-sample t-tests in the second-level analyses with SPM12 (at $p < .05$ FWE-corrected for multiple comparisons).

In order to study how functional networks were modulated during the task, we applied a GLM on the component time courses using a design matrix representing the task. The analyses yielded a set of beta-weights representing the modulation of component time courses by the GLM regressors in relation to the baseline. The GLM design matrix included separate regressors to model the three task conditions (native, accented, and baseline). The six head-motion parameters were included as additional regressors of no interest. Regressors were convolved with the canonical hemodynamic response function and included time derivatives. The beta-weights associated with the three conditions were then used to perform the second-level analyses.

As ICA is a data-driven approach and some components may represent motion-related or physiological signals, we used a three-step IC selection criterion based on prior studies

(Sambataro et al., 2010; Ye, Doñamayor, & Münte, 2014): 1) ICs with a stability index < 0.9 in ICASSO were removed; 2) ICs not primarily located on gray matter (GM) were discarded; this was determined by correlating the spatial map of the components with the prior probabilistic maps of GM, white matter (WM), and cerebral spinal fluid (CSF) provided by the MNI templates of SPM12. We rejected those ICs with a high spatial correlation with WM ($r^2 \geq 0.02$) or CSF ($r^2 \geq 0.05$) and those with greater spatial correlation with WM than with GM; and 3) ICs that were not task-related were excluded. A separate one-way repeated measures ANOVA for the beta-weights of each IC with the factor “condition” (native, accented, and baseline) was performed. We considered that an IC was not modulated during the task if the ANOVA showed no significant main effect of condition at $p < .05$ FDR-corrected.

After component selection, we performed second-level analyses through pairwise comparisons to examine whether and how each condition (native, accented, and baseline) differentially modulated the BOLD signal in each task-related IC. P-values were adjusted based on Holm-Bonferroni correction for multiple comparisons (Holm, 1979; Ludbrook, 1998) taking into account three comparisons: native vs. baseline, accented vs. baseline, and accented vs. native. Additionally, we calculated Spearman’s correlations with the beta-weights of different ICs and with the movie quiz scores.

3. Results

3.1. Movie quiz

As the scores of the movie quiz were not normally distributed, we used the Wilcoxon Rank test to perform all the analyses of these scores. We first examined whether the median (Mdn) of each condition (native, accented, and baseline) differed from every other. The results showed that participants understood accented (Mdn = 4.25) and native

(Mdn = 4.75) dialogues equally well ($V = 192, p = .12$). The results also showed no difference between the baseline (Mdn = 5) and the accented ($V = 123.5, p = .12$) or the native ($V = 170.5, p = .46$) conditions. This means that participants were unable to guess dialogue contents in the baseline condition, that is, they mainly gave the correct “don’t know” response to questions concerning Dutch clips. The fact that participants understood dialogue contents in the native and accented conditions but not in the baseline one was further confirmed with three separate one-sample tests. The results showed that the number of real correct responses in the baseline condition (i.e., not the “don’t know” option but the option with the actual content of the dialog; Mdn = 1) did not differ from chance ($V = 110, p = .99$). In contrast, the number of correct responses did differ from chance in the accented ($V = 465, p = .0001$) and native ($V = 465, p < .0001$) conditions.

3.2. ICA results

Six ICs passed the selection criteria and were therefore taken as ICs of interest (Table 2 and Figure 2); see Supplementary Table 2 and Supplementary Figure 2 for a description of another eight ICs that passed the first two component selection criteria (i.e., identifying them as stable and primarily located on gray matter) but not the third criterion testing task-relatedness. Below, we describe the spatial maps (based on prior studies: Allen et al., 2011; Segall et al., 2012) of the six ICs of interest and differences among conditions (Figure 2).

=====
Table 2 & Figure 2 around here
=====

Medial visual network. This was composed of the middle occipital lobe, including the lingual gyrus and extending into the precuneus. Recent studies have implicated the medial visual cortex in audio-visual integration during the processing of non-native speech: it is

thought to seek the support of visual cues (mouth movements) to enhance speech perception (Barrós-Loscertales et al., 2013). The second-level analyses showed a higher positive modulation for the native and accented conditions—which did not differ from one another—relative to the unknown language condition (baseline). These differences across conditions, however, are difficult to interpret in terms of audio-visual integration. This is because this integration could not occur in the present study (at least not efficiently) because the mouth movements corresponded to the original language of the movie (English). For this reason, we will not discuss this network any further.

Cerebellar speech perception network. This was mainly composed of the cerebellum, probably reflecting differences in prosodic analysis (Ackermann, 2008; Ackermann, Mathiak, & Riecker, 2007; Ivry & Keele, 1989), which is the most relevant distinction between the two regional accents used. There was also a portion of the left IFG (BA46) corresponding to the DLPFC, which is typically associated with EC processes such as attention and working memory (Nee et al., 2013). This network also included the thalamus and the putamen, two subcortical structures previously associated with articulatory planning in the neuropsychological (Avila, González, Parcet, & Belloch, 2004; Craig-McQuaide, Akram, Zrinzo, & Tripoliti, 2014; Wise, Greene, Büchel, & Scott, 1999) and neuroimaging (Abutalebi et al., 2013; Brown et al., 2009; Klein, Zatorre, Milner, Meyer, & Evans, 1994) literature. The temporal pole (TP; BA38), extending into anterior portions of the middle and inferior temporal gyri (ant-MTG/ITG; BA20, 21) and the fronto-polar prefrontal cortex (Fp-PFC; BA10), was also involved in this network, which may have contributed to lexical-semantic processing of auditory patterns (Schwartz et al., 2009; Walker et al., 2011)³. This IC showed a greater negative modulation for the native condition relative to both the baseline and the accented condition. There was also a reduced negative modulation for the accented relative to the

native condition. These results indicate that accented speech represented a listening challenge compared to native speech, but not as much as an unknown language (baseline) did.

Posterior default mode network (DMN). This was predominantly composed of the precuneus and the angular and supramarginal (SMG) gyri. It also involved the posterior cingulate cortex (PCC), middle frontal gyrus (MFG), and MTG. This IC mainly constitutes the posterior part of the DMN (Abou-Elseoud et al., 2010), which is typically activated during mind-wandering (Mason et al., 2007). This network exhibited a reduced negative modulation for the native relative to both the accented and the unknown language (baseline) conditions. The accented and baseline conditions differed marginally in the magnitude of this negative modulation ($p = .051$). These results suggest that processing dialogs in an unfamiliar accent (accented condition) or trying to follow the events in clips with unintelligible dialogs (baseline condition) was task-engaging and that this effort left little room for mind-wandering, as compared with the processing of clips with native dialogues.

Dorsal attentional network (DAN). This was composed of the precuneus, portions of the premotor cortex, MFG, superior frontal gyrus (SFG), ACC, and left IFG. The DAN is typically related to the EC process of guiding attentional processing in a top-down manner (Corbetta & Shulman, 2002). This IC displayed a greater negative modulation for the native condition relative to both the accented and the baseline conditions, and also a greater negative modulation for the accented condition relative to the baseline. This pattern of results indicates that processing clips with accented dialogues demanded more attentional control than did those with native dialogues, but not as much as trying to follow clips whose dialogues participants could not understand. In both the native and the accented conditions, the DAN correlated positively with the *cerebellar speech perception*

network (native condition: $r = .452, p < .012$; accented condition: $r = .46, p < .01$). This may indicate that the better the match between the phonetic properties of intelligible speech (whether accented or native) and participants' templates, the less attentional control was needed.

Salience network (SN). This was predominantly composed of the ACC. It also comprised an extensive portion of the insula, extending towards auditory areas in the STG (including Heschl's gyrus) and also the SMG. There was also an evident bilateral involvement of the prefrontal cortex (PFC), including the MFG and IFG. Portions of the cerebellum, left lingual gyrus, and the precuneus were also included. The SN is typically associated with the processing of salient events (Seeley et al., 2007). This IC showed a negative modulation for both the native and accented conditions relative to baseline. It also showed a greater negative modulation for the native condition relative to the accented one. These results indicate that accented speech was more salient than native speech, but not as much as speech in a completely unfamiliar language. This is probably because both Dutch and accented speech were perceived as salient, the former being the most salient one. Consequently, clips with accented dialogues were more EC-demanding than were clips with native dialogues, but to a lesser extent than were Dutch clips. In addition, the SN in the accented condition correlated negatively with the movie quiz scores for the same condition ($r_s = -.43, p = .019$), such that the greater the engagement of the SN during accented speech perception, the worse the participant's performance on the movie quiz.

Reward network. This was predominantly composed of the thalamus and the striate body, especially the caudate but also the putamen and accumbens nuclei. It also comprised the ventral tegmental area (VTA) and the right IFG. All these regions are components of the frontal cortical-striatal-thalamic circuit involved in reward, motivation, and emotional drive (Haber & Knutson, 2010). Therefore, this IC constitutes part of the reward circuit.

This IC showed a positive modulation for the native condition relative to both the accented and the baseline conditions. There were no differences between the accented condition and the baseline, suggesting that this network is specifically engaged during NSP. This pattern of results is consistent with an emotional bias in favor of native speech.

4. Discussion

We aimed to identify the neural correlates of accented speech processing (ASP) by assessing how listening to speech in a non-native regional accent (Latin American) differentially modulated neural activity compared to listening to speech in a native accent (NSP, Castilian). We observed differences in the brain networks recruited during ASP and NSP. First, ASP recruited a cortical-subcortical network involving regions related to speech perception to a greater extent than did NSP. Second, compared to NSP, ASP differentially modulated three networks that are closely related to domain-general EC processes (DAN, posterior DMN, and SN). Third, NSP involved the reward system more than ASP did. We now discuss these three main observations in detail.

I. Modulation of the speech perception network associated with ASP (*cerebellar speech perception network*)

The results here showed that the mild differences of an unfamiliar accent relative to one's own are enough to challenge the speech perception system. The prominent implication of the cerebellum in this network is consistent with the primordial role that neurocognitive models of speech perception in adverse conditions give to this structure (Callan, Jones, Callan, & Akahane-Yamada, 2004; Guediche, Holt, Laurent, Lim, & Fiez, 2015)⁴. In this respect, it is surprising that only one prior study on ASP has reported cerebellar involvement (Callan et al., 2014). In a similar vein, the implication of the thalamus and putamen in ASP is a novel observation, since no prior study has reported this, despite the well-documented involvement of these structures in phonological-articulatory processes

(see Craig-McQuaide et al., 2014 for a review). Other speech perception regions that prior studies have associated with ASP were not, however, part of our *cerebellar speech perception network*. First, we did not observe that ASP differentially modulated neural activity in the auditory association cortex, as compared with NSP (see Adank et al., 2012a,b, 2013). This discrepancy probably arises from differences between studies in terms of intelligibility during ASP. As we noted in the Introduction, in those previous studies ASP only modulated neural activity in the auditory association cortex differentially to NSP when the intelligibility of accented speech was reduced. This reduction in intelligibility was unlikely to have occurred in the present study, at least not to a significant extent, since we used a non-native regional accent close to the native one in phonetic terms, and, furthermore, the results of the movie quiz indicated that participants understood the dialogues (whether accented or native). Second, ASP involved a portion of the IFG within the DLPFC, consistent with the hypothesis that EC processes are involved in response to the greater effort required by ASP compared to NPS (Venezia et al., 2012). However, Broca's area did not form part of this network, which is surprising given the frequent involvement of this area in ASP studies (Figure 1). We can only advance a tentative explanation for this discrepancy. Broca's area is especially sensitive to abrupt phonetic contrasts (e.g., /r/ vs. /l/), and less so to subtle acoustic differences between sounds belonging to the same phonetic category—for instance, it would be insensitive to the acoustic differences of /s/ in “Sue” and “see” (Lee, Turkeltaub, Granger, & Raizada, 2012; Myers, Blumstein, Walsh, & Eliassen, 2009). Thus, the fact that abrupt phonetic contrasts are not encountered between Castilian and Latin American Spanish could explain the lack of Broca's area involvement during ASP in our study.

Finally, it is also important to highlight the bilateral modulation of the network, whereas previous studies found a predominantly left-lateralized one. This bilateral pattern is

probably related to the fact that—contrary to any prior study—prosody rather than phonetics represented the most relevant distinction between the native and non-native accents. Indeed, it has been consistently shown that prosody is primarily processed in regions of the right hemisphere that are parallel to those composing the classical left language networks (Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015). Notably, the fact that Castilian and Latin American Spanish are relatively similar in phonetic terms does not mean that the left hemisphere of native speakers of Castilian (our participants) should respond to the unfamiliar traits of Latin American speech only mildly—which would have given rise to a predominantly right-lateralized pattern of activity (driven by unfamiliar prosodic traits) rather than the bilateral one that we actually observed. This is most likely because the parallel networks of the right and left hemispheres interact (via the corpus callosum) in order to coordinate the processing of prosodic and phonetic information during the course of speech comprehension (Friederici & Alter, 2004; Sammler, Kotz, Eckstein, Ott, & Friederici, 2010). In fact, it can be seen in Table 2 and Figure 2 that—except in the cases of the thalamus and DLPFC (BA46), where the activity was left- and right-lateralized respectively—the remaining clusters of activity related to ASP in this network (cerebellum; putamen; ant-MTG/ITG, TP; and Fp-PFC) arose bilaterally.

II. Modulation of the domain-general EC networks associated with ASP (DAN, posterior DMN, and SN)

Accented speech recruited domain-general EC processes throughout classical networks associated with two different EC attentional processes: those associated with the degree of cognitive effort required by the task at hand (instantiated through the DAN and posterior DMN), and those associated with the processing of salient stimuli (instantiated

through the SN). In general terms, ASP requires more attentional control and processing of salient stimuli, even in the absence of comprehension differences.

Cognitive effort (DAN and posterior DMN). The DAN is critical for keeping attention focused on the task (Corbetta and Shulman 2002), while the DMN is associated with mind-wandering and distraction (Mason et al., 2007). Consistent with this, a visual inspection of the graphs in Figure 2 reveals that task conditions (native, accented, and baseline) modulated the DAN and the posterior DMN antagonistically. In addition, the positive correlation between the DAN and the *cerebellar speech perception network* suggests that the degree of cognitive effort was related to speech perception difficulties: the more unfamiliar the accent is, the harder it is to process. Focusing on the accented versus native comparison, this means that processing dialogues in the non-native regional accent was cognitively more demanding than was processing dialogues in one's own accent: it required more attentional control (DAN) at the expense of distraction and greater internalization processes (posterior DMN). We do not associate this increased attentional control with language-specific processes activated during ASP, which were probably engaged through the IFG (BA46) in the *cerebellar speech perception network*. Rather, the fact that the DAN and posterior DMN are two amodal networks makes the attentional control more likely to be associated with the effortful processing of the message. This interpretation fits well with prior behavioral data. For example, in a study in which participants had to determine the veracity of different statements, Munro and Derwing (1995) showed that foreign accents delayed sentence processing relative to native accents. Similarly, in a study in which participants performed a lexical decision task on the last word of different sentences, Floccia et al. (2006) demonstrated that the processing of two similar regional accents delayed word recognition. This interpretation also fits well with the “cognitive demands” account (Venezia et al., 2012). Further

research is needed to clarify what exactly leads to this increase in cognitive demands during ASP. One possibility is that the increase has to do with the need to engage bilingual language control (bLC) mechanisms. To date, these mechanisms have been primarily associated with bilingual language production: they allow bilinguals to correctly speak in one language by avoiding interference from the lexical system of the unintended one (Abutalebi & Green, 2007, 2008). However, the fact that some of the brain regions involved in the DAN—the ACC, premotor sites (BA6), and the DLPFC—are considered bLC core neural substrates (Abutalebi et al., 2012) is consistent with the possibility that bLC mechanisms also intervene during ASP: they might aid ASP by preventing interference from the native’s phonetic/prosodic template. Finally, we acknowledge that this “cognitive effort” interpretation needs to be treated with caution, especially with regard to the DMN. This is because the results of recent studies cast some doubt on the DMN being exclusively associated with mind-wandering and distraction. In fact, it appears that the DMN may also be implicated in goal-directed behavior, through flexible coupling with task-relevant networks (Vatansever, Menon, Manktelow, Sahakian, & Stamatakis, 2015). We did not observe that the DMN correlated positively (not even though visual inspection) with any other network. This, in principle, would rule out any coupling with other networks. However, we cannot completely rule out the possibility that the DMN may play a task-related role during ASP, as the contribution of the DMN to the facilitation of task goals is still far from being understood.

Processing salient events (SN). The SN has been classically associated with the identification of salient events in order to guide flexible behavior. Its main components are the ACC and anterior insula, although it often includes portions of the fronto-parietal cortex as well (Menon, 2015). The type of salient events detected by the SN, however, are not necessarily restricted to infrequent oddities. It has been shown that they also

include behaviorally important events such as errors, the detection of which leads to online adjustments in behavioral control (Ham, Leff, de Boissezon, Joffe, & Sharp, 2013). In the specific context of our study, we propose that listeners treated as errors the phonetic/prosodic traits that did not match those represented in their native templates. These errors, in turn, would have prompted the strategy of adaptively adjusting the speech perception system to facilitate the perception of unfamiliar speech features. This adaptation may have been instantiated through the two auditory regions that additionally composed the SN in this study: the STG and Heschl's gyrus. The fact that both the baseline and the accented conditions enhanced neural activity relative to the native condition indicates that unfamiliar speech features drove this "error detection-adaptive adjustment" process. Notably, this enhancement in neural activity was more pronounced in the case of the baseline than in the accented condition. This probably reflects the greater challenge of trying to adapt the listener's speech perception system to the unfamiliar speech features of a completely unknown language. In addition, the negative correlation between movie quiz scores and neural activity in the accented condition suggests that the proposed "error detection-adaptive adjustment" process impoverished the processing of the contents of the message. This interpretation fits well with studies indicating that accented dialogues and conversations—even if completely intelligible—are more difficult to understand than are native ones (see Cristia et al., 2012 for a review), especially if the unfamiliar speech features have to do with prosody (Anderson-Hsieh & Koehler, 1988), as we argue to be the case in the present study.

III. NSP led to higher activity in the reward system than did ASP

This finding fits well with studies showing that participants tend to judge utterances in their own accent as more favorable (Coupland & Bishop, 2007) and trustworthy (Lev-Ari & Keysar, 2010) than those in non-native accents (see also, Frances, Costa, & Baus,

2018). This observation complements that of Bestelmeyer et al. (2015), who reported that a portion of the reward system—the amygdala—responded preferentially to participants’ own accent. In their study, the authors tentatively attributed this bias to the possibility that the native accent was triggering a social identity effect in their participants (Cohen, 2012). Since they intentionally used two accents belonging to two rival socio-cultural groups (Scottish vs. Southern English), it is possible that the amygdala was not included in the *reward network* modulated by ASP in our study because it only shows a favorable response to participants’ own accent if the non-native one evokes conflicting feelings related to arousal—we assume such rivalry feelings were not evoked in our study because confrontation is not generally a feature of social interactions between Latin Americans and Spaniards.

4.1. Limitations of the current study

Although our study overcomes some of the limitations of previous research (e.g., the control of unintelligibility and a more powerful network analysis; see the Introduction), it also has certain shortcomings of its own.

The first one is to do with the fact that we allowed the type and number of phonetic/prosodic differences between Castilian and Latin American Spanish to occur randomly across the movie clips. On the one hand, this strategy mimicked accented speech scenarios in real life, and so the results of this study may be safely extrapolated to daily contexts. We acknowledge, however, that the lack of systematic control over these phonetic/prosodic differences restricts the interpretation of the *cerebellar speech perception network* to general phonetic/prosodic differences between the two regional accents. This shortcoming also derives from the fact that the bulk of prosodic differences between Castilian and Latin American Spanish have yet to be investigated (Hualde & Prieto, 2015). A further two limitations concern the use of ICA. First, this technique

requires the inclusion of several participants in order to identify stable components, which raises the question of how true the results we report in this study might be at the individual level. Second, the functional interpretation of these components is relatively open, as the role of some brain networks—especially those associated with the DMN—is still far from being fully understood. A final shortcoming is the lack of a control group with Latin American participants. The inclusion of such a control group would have allowed us to determine which ASP effects were listener-specific (i.e., linked to participants' native language) and which are most likely universal.

5. Conclusion

This study provides the first evidence on how brain networks are differentially recruited during ASP as compared to NSP. We identified three different networks: (i) a network associated with speech perception, especially with the processing of prosody; (ii) three domain-general EC networks associated with cognitive effort and saliency detection; and (iii) a network linked to the reward system and probably related to differences in in-group versus out-group processing.

Funding

This work was supported by the Ramón y Cajal research program of the Spanish Ministry of Science, Innovation, and Universities (RYC-2016-19477 to MH), the MINECO (grant PSI2016-78805-R to CA), and a pre-doctoral graduate program grant from the Jaume I University (FPI to AMP).

Declarations of interest: none

References

- Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., & Kiviniemi, V. (2010). The effect of model order selection group PICA. *Human Brain Mapping, 31*(8), 1207–1216.
- Abutalebi, J., Della Rosa, P. A., Gonzaga, A. K., Keim, R., Costa, A., & Perani, D. (2013). The role of the left putamen in multilingual language production. *Brain and Language, 125*(3), 307–315.
- Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., et al. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*(9), 2076–2086.
- Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242–275.
- Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: neural evidence from language switching studies. *Language and Cognitive Processes, 23*(4), 557–582.
- Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neurosciences, 31*(6) 265–272.
- Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *The Cerebellum, 6*(3), 202–213.
- Adank, P., Davis, M. H., & Hagoort, P. (2012a). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia, 50*(1), 77–84.
- Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging, 25*(3), 736–740.
- Adank, P., Noordzij, M. L., & Hagoort, P. (2012b). The role of planum temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping, 33*(2), 360–372.
- Adank, P., Nuttall, H. E., Banks, B., & Kennedy-Higgins, D. (2015). Neural bases of accented speech perception. *Frontiers in Human Neuroscience, 9*:558.
- Adank, P., Rueschemeyer, S. A., & Bekkering, H. (2013). The role of accent imitation in sensorimotor integration during processing of intelligible speech. *Frontiers in Human Neuroscience, 7*:634.
- Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Frontiers in Systems Neuroscience, 5*:2.
- Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561–613.
- Avila, C., González, J., Parcet, M. A., & Belloch, V. (2004). Selective alteration of native, but not second language articulation in a patient with foreign accent syndrome. *Neuroreport, 15*(14), 2267–2270.
- Banks, B., Gowen, E., Munro, K., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America, 137*(4), 2015–2024.
- Barrós-Loscertales, A., Ventura-Campos, N., Visser, M., Alsius, A., Pallier, C., Ávila Rivera, C. et al. (2013). Neural correlates of audiovisual speech processing in a second language. *Brain and Language, 126*(3), 253–262.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.

- Bestelmeyer, P. E., Belin, P., & Ladd, D. R. (2015). A neural marker for social bias toward in-group accents. *Cerebral Cortex*, 25(10), 3953–3961.
- Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Tuerkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain and Cognition*, 70(1), 31–41.
- Burton, T. (2010). *Alice in Wonderland* (Movie). Hollywood: Walt Disney Pictures.
- Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140–151.
- Calhoun, V. D., Liu, J., & Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*, 45(1), S163–S172.
- Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Human Neuroscience*, 8:275.
- 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(3), 1182–1194.
- Cattaneo, L., & Rizzolatti, G. (2009). The mirror neuron system. *Archives of Neurology*, 66(5), 557-560.
- Cohen, E. (2012). The evolution of tag-based cooperation in humans: the case for accent. *Current Anthropology*, 53(5), 588–616.
- Corbetta, M., & Shulman, C. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215.
- Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics*, 11(1), 74–93.
- Craig-McQuaide, A., Akram, H., Zrinzo, L., & Tripoliti, E. (2014). A review of brain circuitries involved in stuttering. *Frontiers in Human Neuroscience*, 8:884.
- Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, 3:479.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423–3431.
- Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., & Calhoun, V. D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Human Brain Mapping*, 32(12), 2075–2095.
- Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276–1293.
- Frances, C., Costa, A., & Baus, C. (2018). On the effects of regional accents on memory and credibility. *Acta Psychologica*, 186, 63–70.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2(4), 189–210.

- Guediche, S., Holt, L., Laurent, P., Lim, S. J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, *25*(7), 1867–1877.
- Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*, 4–36.
- Ham, T., Leff, A., de Boissezon, X., Joffe, A., & Sharp, D. J. (2013). Cognitive Control and the Salience Network: An Investigation of Error Processing and Effective Connectivity. *Journal of Neuroscience*, *33*(16), 7091–7098.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Himberg, J., Hyvärinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage*, *22*(3), 1214–1222.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.
- Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish. In S. Frota & P. Prieto (Eds.), *Intonational variation in Romance* (pp. 350–391). Oxford: Oxford University Press.
- Ivry, R., & Keele, S. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, *1*(2), 136–152.
- Klein, D., Zatorre, R. J., Milner, B., Meyer, E., & Evans, A. C. (1994). Left putaminal activations when speaking a second language: Evidence from PET. *Neuroreport*, *5*(17), 2295–2297.
- Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience*, *32*(11), 3942–3948.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096.
- Li, Y.-O., Adali, T., & Calhoun, V. D. (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Human Brain Mapping*, *28*(11), 1251–1266.
- Ludbrook, J. (1998). Multiple comparisons procedures updated. *Clinical and Experimental Pharmacology and Physiology*, *25*, 1032–1037.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, *315*(5810), 393–395.
- Menon, V. (2015). Salience network. In Toga AW (Ed.), *Brain mapping: An encyclopedic reference* (Vol 2, pp 597–611). London (UK). Academic Press: Elsevier.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, *89*(2), 277–289.
- Munro, M. J., & Derwing, T. G. (1995). Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(Pt 3), 289–306.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895–903.

- Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., et al. (2013). A Meta-analysis of Executive Components of Working Memory. *Cerebral Cortex*, *23*(2), 264–282.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., et al. (2013). Intonational phonology of Catalan and its dialectal varieties. In: S. Frota, & P. Prieto (Eds.), *Intonational variation in Romance* (pp. 392–418). Oxford: Oxford University Press.
- Prieto, P., Vanrell, M. M., Astruc, L., Payne, E., & Post, B. (Eds.). (2010). Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish. *Speech Prosody*. Chicago, IL, 11–14 May 2010.
- Sambataro, F., Blasi, G., Fazio, L., Caforio, G., Taurisano, P., Romano, R., et al. (2010). Treatment with olanzapine is associated with modulation of the default mode network in patients with Schizophrenia. *Neuropsychopharmacology*, *35*, 904–912.
- Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*(23), 3079–3085.
- Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., & Friederici, A. D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain*, *133*, 2643–2655.
- Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2009). Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain*, *132*(12), 3411–3427.
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356.
- Segall, J. M., Allen, E. A., Jung, R. E., Erhardt, E. B., Arja, S. K., Kiehl, K., et al. (2012). Correspondence between structure and function in the human brain at rest. *Frontiers in Neuroinformatics*, *6*:10.
- Vatansver, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default mode network connectivity during task execution. *Neuroimage*, *122*, 96–104.
- Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology*, *3*:157.
- Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2011). Support for anterior temporal involvement in semantic error production in aphasia: new evidence from VLSM. *Brain and Language*, *117*(3), 110–122.
- Wise, R. J. S., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*(9158), 1057–1061.
- Xu, J., Potenza, M. N., Calhoun, V. D. 2013. Spatial ICA reveals functional activity hidden from traditional fMRI GLM-based analyses. *Frontiers in Neuroscience*, *7*, 54.
- Ye, Z., Doñamayor, N., & Münte, T. F. (2014). Brain network of semantic integration in sentence reading: Insights from independent component analysis and graph theoretical analysis. *Human Brain Mapping*, *35*(2), 367–376.
- Yi, H. G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, *8*:768.

Table 1. Participants' knowledge of languages other than Castilian.

	CEFR level						Total
	A1	A2	B1	B2	C1	C2	
Valencian	-	3	-	4	9	14	30
English	-	6	14	5	1	-	26
French	-	4	5	-	-	-	9
German	-	-	1	-	-	-	1

Note. CEFR = Common European Framework of Reference for Languages; A1 = Beginner, A2 = Elementary, B1 = Intermediate; B2 = Upper-intermediate; C1 = Advanced; C2 = Mastery (native-like)

Table 2. Brain regions belonging to the ICs of interest

Component	Region ^a	BA	MNI coordinates			K	<i>t</i> value	<i>p</i> value	
			x	y	z				
Medial Visual Network	L Cuneus, precuneus	18/19,7	-3	-76	11	2,975	31.49	.0001	
	R Cuneus, precuneus	18/19,7	15	-58	-4		27.06	.0001	
	Lingual gyrus	17/18	3	-76	17		27.63	.0001	
Posterior DMN	Precuneus	7	-6	-64	20	2,506	26.85	.0001	
	L PCC	31	-9	-52	23		23.37	.0001	
	R PCC	31	3	-61	26		23.07	.0001	
	L Angular gyrus	39	-39	-67	38		336	12.35	.0001
	R Angular gyrus	7	42	-61	41		281	11.61	.0001
	R SMG gyrus	40	48	-58	29		10.34	.0001	
	R Fp-PFC	10	3	47	-7		115	9.94	.0001
	L MTG	21	-57	-13	-16		19	7.44	.001
DAN	Precuneus	7	-18	-61	56	4,159	18.77	.0001	
	L Premotor cortex	6	-24	24	59		223	11.06	.0001
	R Premotor cortex	6	24	-4	53		158	9.89	.0001
	L DLPFC (MFG)	9	-39	32	29		184	9.61	.0001
	L DLPFC (IFG)	46	-45	35	14		8.58	.0001	
	L OFC (SFG)	10	-30	44	20		7.01	.003	
	R DLPFC (MFG)	9	36	44	26		56	7.70	.0001
	R OFC (SFG)	10	30	32	38		7.45	.001	
	ACC	24	0	-1	38		11	7.15	.002
Reward Network	R Thalamus	-	6	-13	17	1,495	21.84	.0001	
	L Thalamus	-	-6	-13	14		20.65	.0001	
	R Caudate	-	12	8	2		17.78	.0001	
	L Caudate	-	-9	11	2		16.15	.0001	
	VTA	-	3	-28	-7		10.49	.0001	
	Red nucleus	-	-6	-16	7		10.3	.0001	
	Accumbens nucleus	-	-21	-4	5		9.85	.0001	
	R DLPFC	9,46	45	26	17		32	8.57	.0001
Cerebellar Speech Perception Network	L Cerebellum	-	-9	-49	-28	2,394	19.73	.0001	
	R Cerebellum	-	15	-49	-25		17.18	.0001	
	L ant-MTG/ITG, TP	20,21,38	-39	-10	-28		103	10.03	.0001
	R ant-MTG/ITG, TP	20,21,38	42	5	-34		58	9.67	.0001
	R Putamen	-	30	-7	-1		34	9.43	.0001
	L Putamen	-	-27	-13	-7		26	7.16	.002
	L Fp-PFC	10	-30	59	17		20	8.16	.0001
	R Fp-PFC	10	24	62	16		51	7.61	.001
	R Fp -PFC	10	42	53	-1		48	7.25	.001
	R DLPFC (IFG)	46	51	44	5		7.09	.002	
	Thalamus	-	0	-19	2		12	6.73	.005
	Salience Network	R ACC	24,32	6	32		17	2,252	19.29
L ACC		24,32	-3	32	23	18.74	.0001		
L Insula		13,22	-39	11	-1	760	17.74		.0001
L SMG		40	-63	-28	20	10.87	.0001		

L STG (Heschl)	22,41,42	-51	-16	8		10.30	.0001
R SMG	40	54	-40	44	105	11.36	.0001
R OFC (MFG)	10	42	50	8	1,543	16.52	.0001
R OFC (IFG)	47	33	17	-4		15.61	.0001
R Insula	13,22	39	11	2		15.31	.0001
R STG (Heschl)	22,41,42	57	-10	5		12.55	.0001
L DLPFC (MFG)	46	-36	38	23	394	12.88	.0001
L DLPFC (IFG)	46	-45	41	14		12.41	.0001
L OFC (MFG)	10	-39	47	14		12.72	.0001
L SMG	40	54	-40	44	105	11.36	.0001
R Cerebellum	-	39	-55	-34	32	8.71	.0001
L Cerebellum	-	-30	-70	-19	37	7.90	.0001
L Lingual gyrus	19	-24	-76	-16		6.47	.011
R Precuneus	7	6	-79	38	60	7.14	.002
L Precuneus	7	-6	-76	38		6.92	.003

^a Regions with no lateralization (right, left) information showed bilateral activation. Statistical threshold at $p < .05$ FWE-corrected for multiple comparisons, $K > 10$. ACC = anterior cingulate cortex; ant- = anterior; BA = Brodmann areas; DAN = dorsal attentional network; DLPFC = dorsolateral prefrontal cortex; DMN = default mode network; Fp-PFC = fronto-polar prefrontal cortex; IC = independent component; IFG = inferior frontal gyrus; ITG = inferior temporal gyrus; L = Left; MFG = middle frontal gyrus; MNI = Montreal Neurological Institute; MTG = middle temporal gyrus; PCC = posterior cingulate cortex; PFC = prefrontal cortex; R = right; SFG = superior frontal gyrus; SMG = supramarginal gyrus; STG = superior temporal gyrus; TP = temporal pole; VTA = ventral tegmental area.

Captions to figures

Figure 1. Illustration of the approximate brain regions that prior fMRI studies associated with accented speech processing (A: left hemisphere, B: right hemisphere). The diamonds represent only the local peaks (maxima), regardless of the extent of the cluster reported in the original study. Brighter diamonds indicate that the peak maximum was close to the cortex surface, whereas blurred diamonds indicate a more medial location (IFG = inferior frontal gyrus, IPS = inferior parietal sulcus, MTG = middle temporal gyrus, SFS = superior frontal sulcus, SMA = supplementary motor area, STG = superior temporal gyrus, STS = superior temporal sulcus).

Figure 2. The images show the six task-related networks ($p < .05$ FWE-corrected for multiple comparisons; numbers in the images = Z MNI coordinates; $K > 10$). The graphs show the beta weights per condition (native, accented, and baseline) in each task-related network (* p -value < Holm-Bonferroni corrected p -value). Error bars represent 95% within-subject confidence intervals.

Footnotes

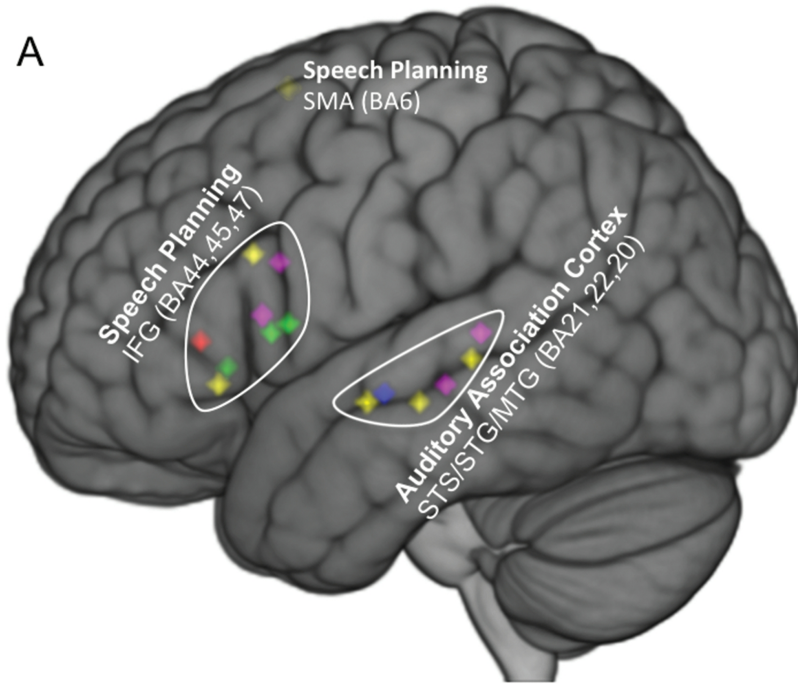
¹ ASP was not the exclusive focus in all prior studies, and thus some of them included types of listening adversities other than accent (e.g., background noise, Adank et al. 2012a, 2013; different voices, Adank et al., 2012b; a second language: Callan et al., 2014). These other listening adversities could have interfered with accented speech attunement, which, in turn, may have affected the pattern of brain activity in response to ASP.

² One of these differences has to do with the relative duration of the pre-tonic, tonic, and post-tonic syllables, making the same pitch accent to produce different auditory impressions in one and the other Spanish variant. Castilian and Latin American also differ in intonation. For instance, the pitch accent in Castilian statements is usually of the rising sort, with an initial valley commonly around the onset of the stressed syllable and a rise throughout it. In Latin American statements, by contrast, the stressed syllable is characterized by a low tone for most of its duration, with the rise being often confined to the next syllable. Much of the prosodic variation between Castilian and Latin American, however, is yet to be studied—see Hualde and Prieto (2015) for a detailed review of the current state of the art. This means that although it is easy to identify the two varieties by ear from their broad rhythmic and intonational differences (Hualde & Prieto, 2015), many of these differences still need to be formally described.

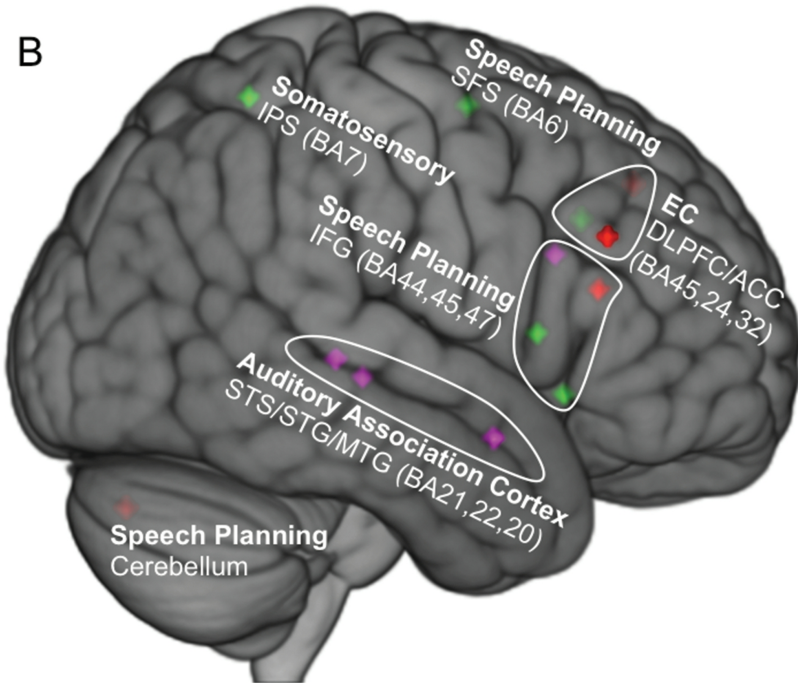
³ Theoretically, functional networks need to be interpreted as a whole. In this respect, please note that we do not propose to interpret the *cerebellar speech perception network* region-by-region—we mean only that we associate this network with speech perception given that it is composed of many areas that have previously been associated with different aspects of this concept.

⁴ The model mostly adopted in prior fMRI literature about ASP is the “articulatory-to-auditory feedback control system” (Callan et al., 2004), which is based on simulation theories of action perception (Cattaneo & Rizzolatti, 2009). Other models, however, are based on supervised learning mechanisms involving sensory prediction error signals rather than simulations (Guediche et al., 2015). As our data say nothing about the specific neurocognitive mechanisms behind the modulation that ASP exerts on the speech perception network, we will not discuss the hypotheses of each specific model.

A

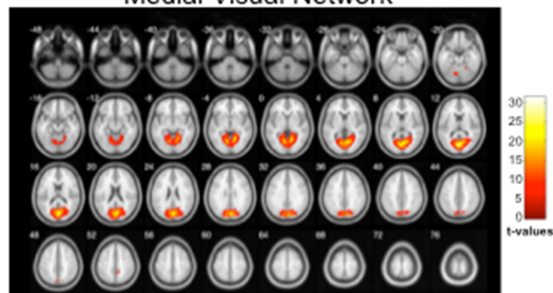


B

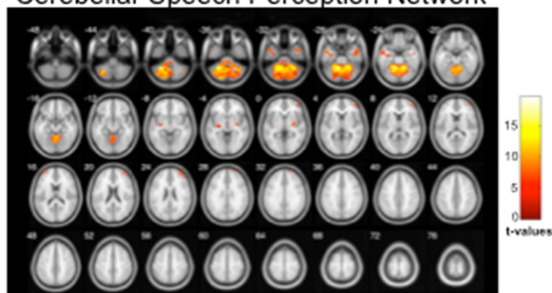


- ◆ Adank et al. 2012a
- ◆ Adank et al. 2012b
- ◆ Adank et al. 2013
- ◆ Callan et al. 2014
- ◆ Yi et al. 2014

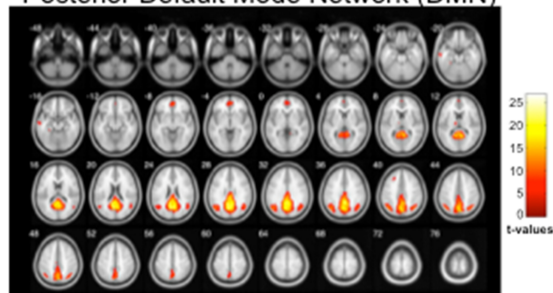
Medial Visual Network



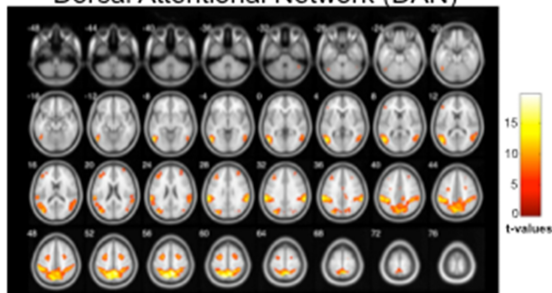
Cerebellar Speech Perception Network



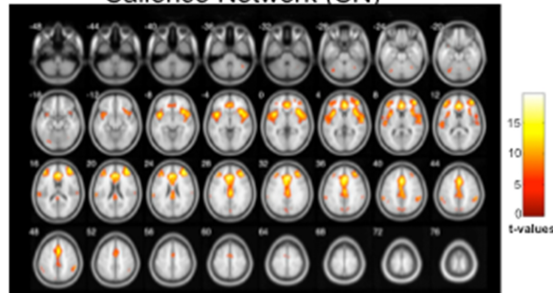
Posterior Default Mode Network (DMN)



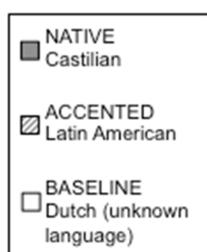
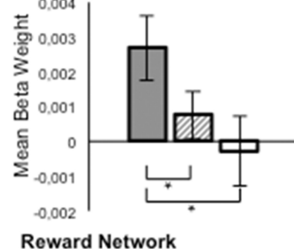
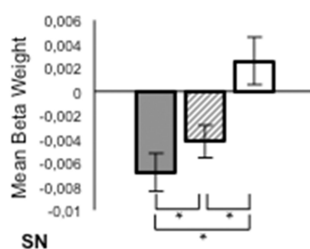
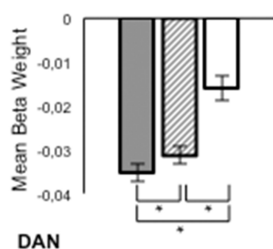
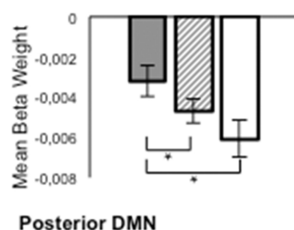
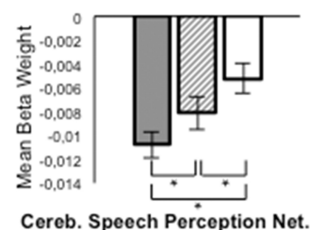
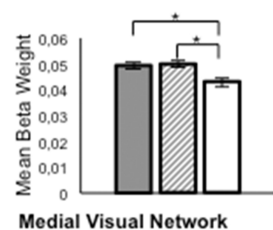
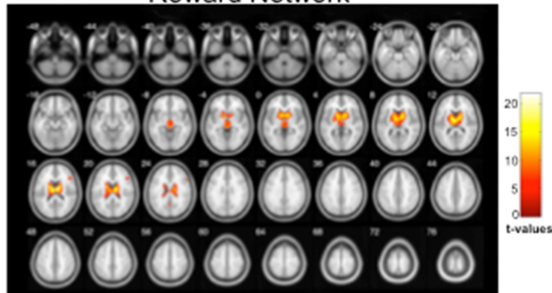
Dorsal Attentional Network (DAN)



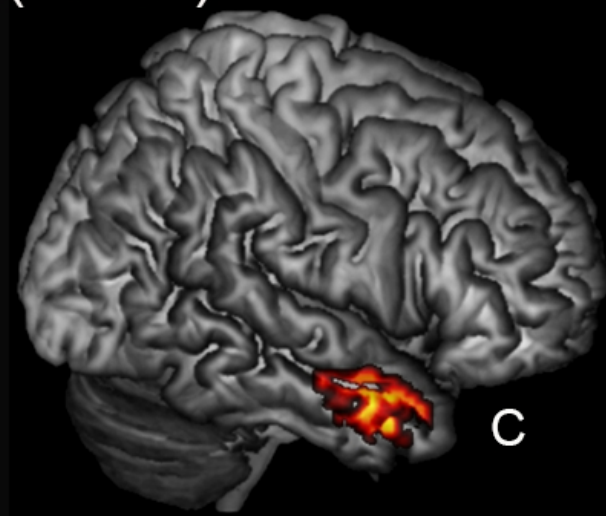
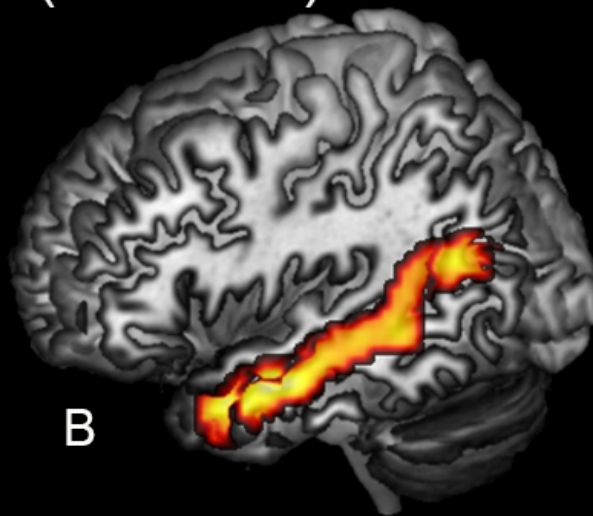
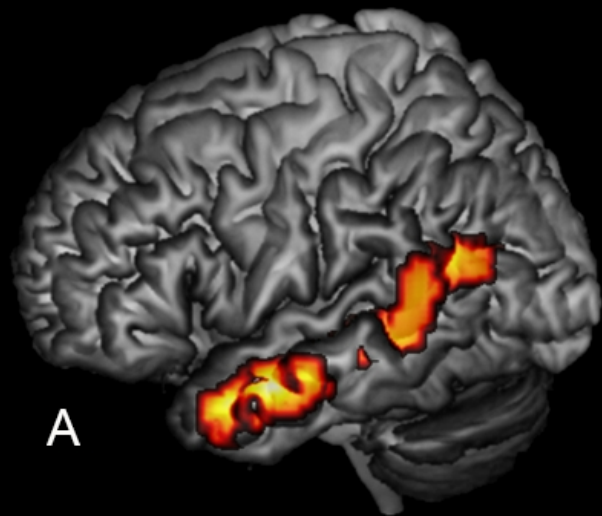
Salience Network (SN)



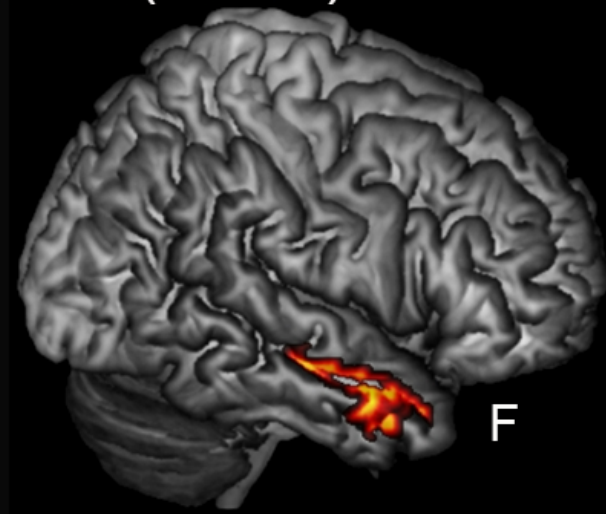
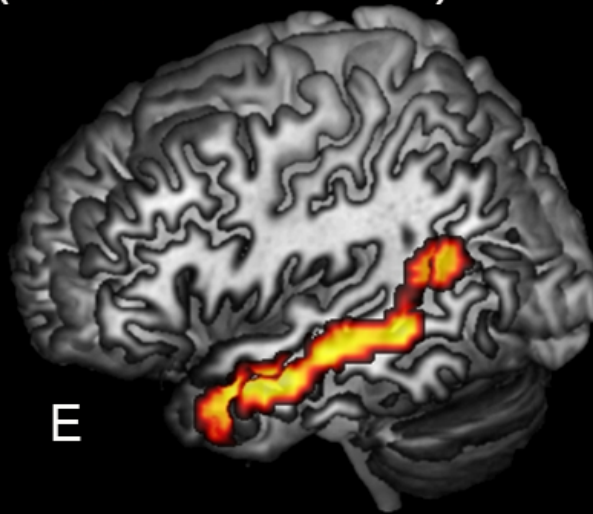
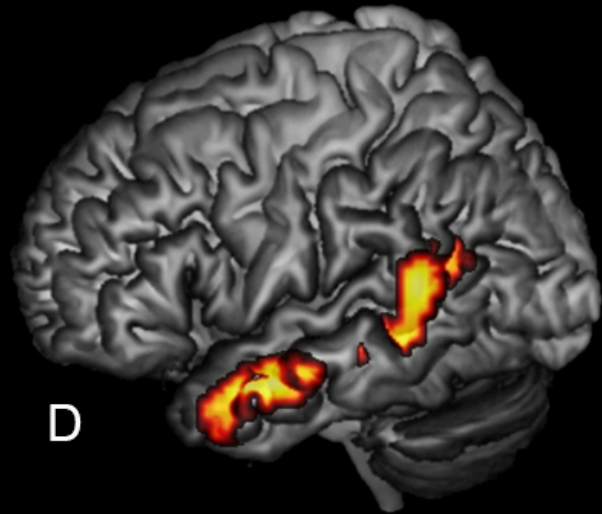
Reward Network



Native (Castilian) > Baseline (Dutch)



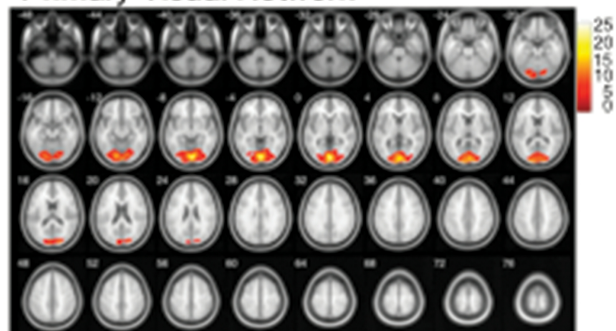
Accented (Latin American) > Baseline (Dutch)



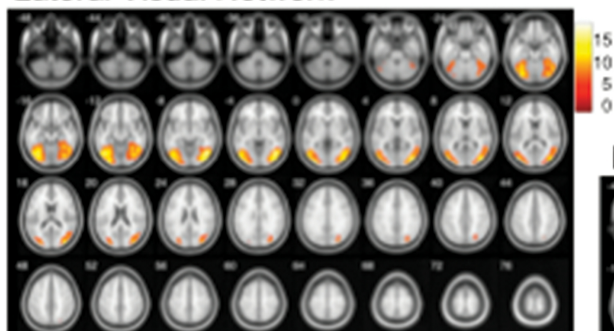
Caption to Supplementary Figure 1

Supplementary Figure 1. Rendered images showing the results of the whole brain analyses contrasting each experimental condition against baseline (A and D = lateral surface of the left hemisphere; B and E = a more medial view of the left hemisphere; C and F = surface of the right hemisphere). The statistical threshold is $p < .05$ FWE-corrected for multiple comparisons ($K > 20$).

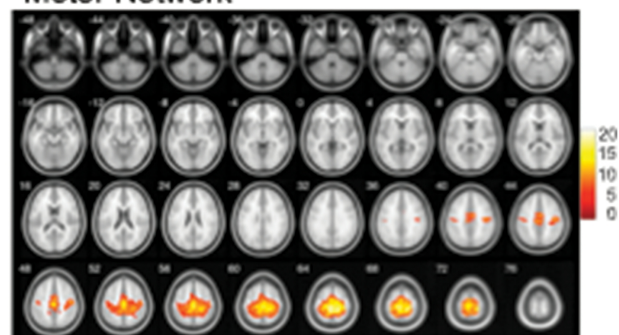
Primary Visual Network



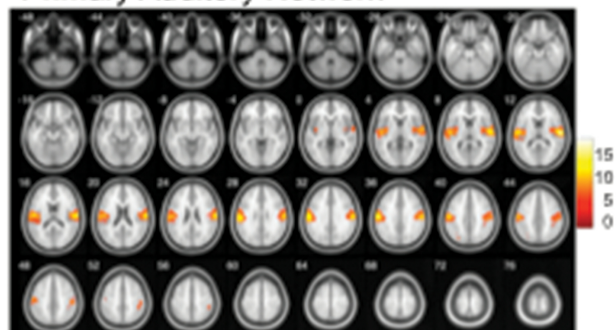
Lateral Visual Network



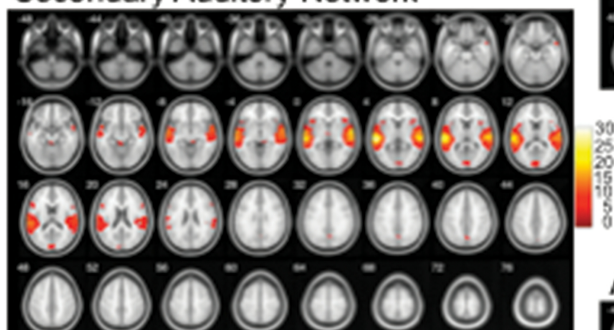
Motor Network



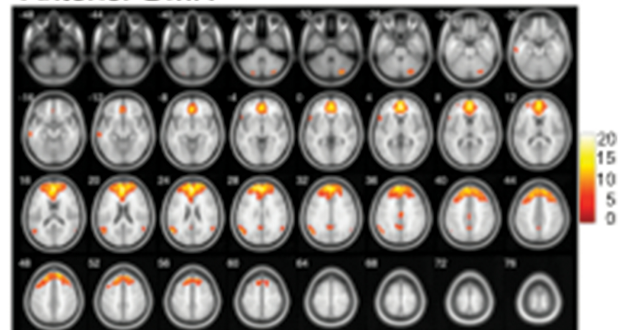
Primary Auditory Network



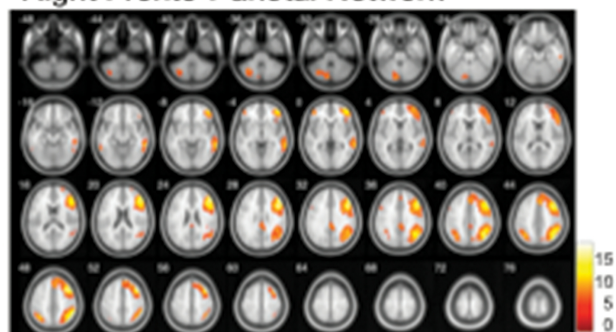
Secondary Auditory Network



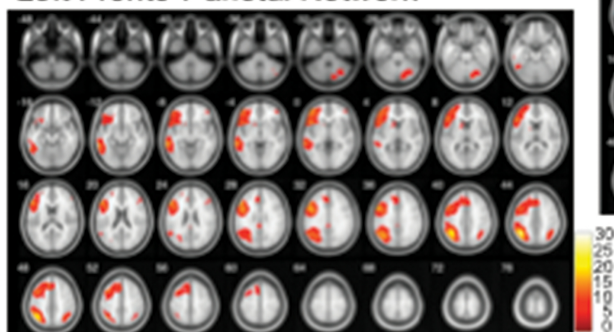
Anterior DMN



Right Fronto-Parietal Network



Left Fronto-Parietal Network



Caption to Supplementary Figure 2

Supplementary Figure 2. The eight ICs that passed the first two component selection criteria of the ICA (identifying them as stable and primarily located on gray matter) but not the third criterion testing task-relatedness. The statistical threshold is $p < .05$ FWE-corrected for multiple comparisons ($K > 10$). The color bars represent the t -values applicable to the images, while the numbers in the images correspond to the Z MNI coordinates.

Supplementary Table 2. Brain regions belonging to the ICs of no interest (i.e., task-unrelated)

Component	Region ^a	BA	MNI coordinates			K	<i>t</i> value	<i>p</i> value
			x	y	z			
Motor Network	R Postcentral gyrus	4	15	-37	65	2,980	20.34	.0001
	L Precuneus	4	-3	-40	65			
	L Precentral gyrus	4	-15	-28	62			
Primary Visual Network	R Lingual gyrus	18	3	-85	-7	1,904	29.27	.0001
	L MOG	18	-18	-94	11			
	L Lingual gyrus	18	-15	-85	-13			
Lateral Visual Network	R MOG	19	45	-64	-7	1,706	18.8	.0001
	R IOG	19	39	-82	-7			
	L Fusiform (temporal)	37	-42	-4	53	1,332	16.2	.0001
	L MOG	18	-33	-88	-1			
	L Fusiform (occipital)	19	-33	-76	-19			
Anterior DMN	L MFG	9	-3	50	20	3,427	20.69	.0001
	ACC	10	0	44	2			
	L SFG	10	-19	53	26			
	L Angular gyrus	39	-54	-61	23	168	12.01	.0001
	R Cerebellum	-	30	-82	-34	67	11.58	.0001
	L Cerebellum	-	-24	-82	-37	15	8.96	.0001
	L MTG	21	-63	-25	-16	38	8.78	.0001
	R MCC	24	6	-16	35	49	8.74	.0001
	L Precuneus	31	-3	-52	35	75	8.44	.0001
	L IFG	45	-57	20	2	36	7.79	.0001
	R Angular gyrus	40	57	-55	26	46	7.73	.0001
	L Fronto-Parietal Network	L Angular gyrus	40	-36	-55	44	1,050	30.69
L Postcentral gyrus		2	-51	-31	44			
L IFG		9	-51	8	35	2,845	19.92	.0001
L MFG		46,6	-51	23	26			
L MTG		37,21	-57	-40	-13	491	16.74	.0001
R Angular gyrus		40	33	-58	47	142	7.21	.0001
R Cerebellum		-	15	-79	-25	180	9.69	.0001
L Caudate nucleus		-	-12	11	5	39	9.38	.0001
L Putamen		-	-15	17	-7		6.60	.0008
PCC		23	0	-34	29	33	8.33	.0001
R MFG		46	51	35	20	26	7.88	.0001
R MFG		11	45	47	-4	22	7.63	.0001
R Fronto-Parietal Network		R MFG	8	36	14	50	2,301	18.76
	R IFG	9	45	17	29			
	R SFG	9	45	35	26			
	R Angular gyrus	40	39	-58	38	1,061	17.09	.0001
	L Angular gyrus	40	-36	-58	47	200	13.06	.0001
	R MTG	21,22	63	-46	-4	328	11.87	.0001
	L IFG	10	-45	44	1	40	10.64	.0001
	L Cerebellum	-	-33	-64	-34	204	10.34	.0001
	R MCC	9,39	3	-37	35	103	9.39	.0001
	L MTG	37	-57	-49	-13	12	8.93	.0001
	R Precuneus	7	9	-70	41	19	8.57	.0001
	L MFG	9	-51	11	35	18	7.11	.002
	Primary Auditory	R Heschl's gyri	41,42	57	-16	11	1,008	18.66

Network	R SMG	40	63	-19	23		13.93	.0001
	R Postcentral gyrus	2	48	-25	50		10.39	.0001
	L Postcentral gyrus	2	-54	-19	35	1,039	16.15	.0001
	L Insula (Heschl)	13	-51	-13	11		13.59	.0001
	L Heschl's gyri	41,42	-57	-19	14		11.77	.0001
	L Precuneus	19	-27	-76	41		7.29	.001
Secondary Auditory Network	L STG	22,21	-57	-28	2	1,567	30.48	.0001
	R STG	22,21,41	60	-25	2	1,542	25.6	.0001
	L Cuneus	18	-6	-91	17	118	10.42	.0001
	R Cuneus	18	3	-88	8		7.89	.0001
	L Lingual gyrus	27	-6	-34	-16	32	8.81	.0001
	R Lingual gyrus	27	6	-34	-7		7.76	.0001
	R Thalamus	-	3	-64	32	23	6.68	.006
	L Precuneus	7	-3	-64	38		6.67	.006

^aRegions with no lateralization (right, left) information showed bilateral activation. Statistical threshold at $p < .05$ FWE-

corrected for multiple comparisons, $K > 10$. ACC = anterior cingulate cortex; BA = Brodmann areas; DMN = default mode network; IC = independent component; IFG = inferior frontal gyrus; IOG = inferior occipital gyrus; L = Left; MCC = mid cingulate cortex; MFG = middle frontal gyrus; MNI = Montreal Neurological Institute; MTG = middle temporal gyrus; MOG = middle occipital gyrus; PCC = posterior cingulate cortex; R = right; SFG = superior frontal gyrus; SMG = supramarginal gyrus; STG = superior temporal gyrus.

Supplementary data analysis

GLM-based analysis

We performed a whole-brain GLM analysis (Friston et al., 1995). In the first-level analysis a statistical model was computed for each participant. The GLM design matrix included separate regressors for each condition (accented, native, and baseline) by applying a canonical hemodynamic response function and its time derivative. The six head-motion parameters were included as additional regressors of no interest. The time series were high-pass filtered with a cut-off frequency of 1/128 Hz to eliminate low-frequency components. Finally, statistical contrast images were generated by comparing different conditions. The second-level analyses consisted of one-sample t tests (at $p < .05$ FWE-corrected for multiple comparisons at whole-brain level) using the images from the first-level analyses.

Results and discussion

As expected, conspicuous neural activity in brain regions classically associated with language comprehension (Hickok & Poeppel, 2007) was observed when comparing either the native or the accented condition against the baseline (Dutch dubbed clips), indicating that participants understood the dialogues in Castilian and Latin American Spanish but not those in Dutch. This neural activity was found in substantial clusters throughout the primary and association auditory cortices in the left MTG and left STG, extending from Wernicke's area to the temporal pole (see Supplementary Table 1 and Supplementary Figure 1). This activity was bilateral in the most anterior portions of the MTG and the temporal pole. The opposite contrasts (i.e., baseline against either experimental condition) did not reveal any difference.

No differences were observed when comparing the native and accented conditions in either direction, which is consistent with Bestelmeyer et al. (2015), with the exception of the higher activity in the amygdala that those authors observed during NSP compared to ASP. The results of this analysis contrast with those obtained with ICA (see main text), suggesting that the latter analysis is better suited to the study of ASP.

Supplementary Table 1. Whole brain analyses contrasting each experimental condition against baseline.

Contrast	Region	BA	MNI			K	<i>t</i> value	<i>p</i> value
			coordinates					
			x	y	z			
Native > Baseline	L ant-MTG, TP	20,21	-54	-7	-16	556	11.33	.0001
	L STG, Wernicke's area	39	-45	-61	20		8.74	.0001
	R ant-MTG, STG, TP	20,21,38	57	2	-22	125	9.17	.0001
Accented > Baseline	L ant-MTG, TP	21,22	-54	2	-19	342	9.78	.0001
	L STG, Wernicke's area	22	-60	-49	17		8.47	.0001
	R ant-MTG, STG, TP	21,22,38	57	2	-22	114	7.58	.001

Statistical threshold at $p < .05$ FWE-corrected for multiple comparisons, $K > 20$. Contrast: Accented = Latin American;

Baseline = Dutch; Native = Castilian. ant- = anterior; BA = Brodmann areas; L = Left; MNI = Montreal Neurological

Institute; MTG = middle temporal gyrus; R = right; STG = superior temporal gyrus; TP = temporal pole.

We found that accented speech engaged speech perception and executive control networks. In addition, the reward network showed a preference for the native accent. These results help to advance knowledge about the neural basis of accented speech processing, which is an increasingly common listening condition in daily life.