NeuroImage 325 (2026) 121638

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Does nonsense make sense? A springboard to studying dynamic reconfiguration of large-scale networks during semantic and intonation speech processing

Irina Anurova [a,b,*], Katarzyna Ciesla [a,c,d], Amir Amedi [a,*]

[a] The Baruch Ivcher Institute for Brain, Cognition & Technology, Reichman University, Herzliya, Israel
[b] The BioMag Laboratory, HUS Diagnostic Center, University of Helsinki, and Helsinki University Hospital, Helsinki, Finland
[c] The Clinical Brain Imaging R&D Center (C.BIRD), Sheba Medical Center, Ramat Gan, Israel
[d] World Hearing Center, Warsaw, Poland

ARTICLE INFO

Keywords:
Speech
Semantics
Intonation
Prosody
Functional magnetic resonance imaging
Large-scale functional networks
Psychophysiological interaction (PPI)

ABSTRACT

Along with semantic content, prosody, which includes intonation, stress, and rhythm, plays an important role in speech comprehension by conveying paralinguistic information. While functional neuroimaging data yields valuable insights into the involvement of language-related grey matter structures in semantic and prosodic processing, the dynamic organization of large-scale functional networks supporting ecologically valid, continuous speech perception remains poorly understood. By selectively disrupting the semantic and/or prosodic structure of complex, continuous speech narratives, we investigated how semantic coherence and intonation modulate interactions between language network and domain-general systems. Our factorial fMRI design revealed involvement of classical left-lateralized language network in the processing of semantic content and bilateral superior temporal areas in the processing of intonation. Scrambled speech engaged the salience network and executive control regions, reflecting effortful meaning extraction. Psychophysiological interaction (PPI) analysis revealed a functional dissociation within the inferior frontal gyrus (IFG). Semantic coherence strengthened connectivity within left-hemispheric language network, including both IFG subregions, pars opercularis (IFGop) and pars triangularis (IFGtri), posterior temporal cortex, and inferior parietal lobule (IPL), and between right IFG and left IPL. Intonation enhanced right IFGop coupling with left temporal cortex, while reducing left IFGop connectivity with the right IPL. Thus, bilateral IFGop demonstrated context-dependent flexibility across semantic and prosodic networks, whereas bilateral IFGtri was specifically engaged in semantic processing. These findings illustrate a dynamic trade-off, where the engagement of core language networks for structured speech is complemented by the recruitment of domain-general evolutionary older systems when processing demands increase due to degraded input.

1. Introduction

Due to the social nature of human beings, verbal communication plays an essential role in daily life. Successful speech comprehension requires accurate interpretation of both semantic and prosodic cues. During this process, our brain integrates various sound features into a holistic auditory image and converts it into a meaningful message.

Prosody plays a dual role in communication. On the one hand, it provides information about emotional connotations of uttered words and the affective state of the interlocutor. On the other hand, it enables discrimination between different speech acts, such as questions or declarative statements. In addition, certain prosodic cues, e.g., variations of the voice fundamental frequency, loudness, speech rate, and rhythm, aid in judging whether the utterance is completed or expected to be continued (Xu, 2019), and in distinguishing between potentially ambiguous sentences (Meyer et al., 2002).

Prosody and semantics are processed in partially segregated brain networks (Baum and Pell, 1999; Binder et al., 2009; Fedorenko et al., 2015; 2011; Pasquinelli et al., 2023; Price, 2000; Turker et al., 2023; Wildgruber et al., 2004; Zhang et al., 2010), and are further integrated to ensure successful speech comprehension.

Semantic processing is conventionally linked to the language

* Corresponding authors.
E-mail addresses: anurova.irina@gmail.com, irina.anurova@hus.fi (I. Anurova), amir.amedi@runi.ac.il (A. Amedi).

https://doi.org/10.1016/j.neuroimage.2025.121638
Received 28 January 2025; Received in revised form 24 November 2025; Accepted 5 December 2025
Available online 5 December 2025

functional network, encompassing regions activated by linguistic inputs across various modalities. This network comprises association temporal, parietal, and inferior prefrontal areas (Binder et al., 2009; Fedorenko et al., 2015; Hagoort and Indefrey, 2014; Hodgson et al., 2021; Price, 2000; Turker et al., 2023). Dorsolateral prefrontal cortex (DLPFC) has been suggested to be an additional substrate for semantic processing (Bottini et al., 1994; Fedorenko et al., 2011), in a subregion partially distinct from the one serving general executive functions (Fedorenko et al., 2011). Multiple neuroimaging and lesion studies suggest that semantic processing is lateralized to the left hemisphere, although may recruit right-hemispheric sites, especially for the interpretation of figurative aspects of language (e.g., Binder et al., 2009; Bottini et al., 1994; Hodgson et al., 2021; Turker et al., 2023; Vigneau et al., 2011).

Prosodic processing shows more bilateral organization, although with some right-hemispheric bias, along the perisylvian (Baum and Pell, 1999; Fedorenko et al., 2015; Kandylaki et al., 2017; Sankaran et al., 2024; Tang et al., 2017; Themistocleous, 2025; Wildgruber et al., 2004; Witteman et al., 2011; Zhang et al., 2010) and in the inferior frontal cortex (Sihvonen et al., 2022; Van Lancker Sidtis et al., 2006; Wildgruber, 2009; Wildgruber et al., 2004).

Hemispheric specialization aligns with acoustic processing preferences. The cue-dependent hypothesis posits that temporal and spectral prosodic cues may be independently lateralized to the left and right hemispheres, respectively (Van Lancker and Sidtis, 1992). Although initially proposed for the prosodic processing, this hypothesis may be applied to other aspects of speech processing. It is consistent with more general models of auditory processing emerging from multiple neuroimaging and lesion studies. Zatorre and colleagues (Flinker et al., 2019; Zatorre and Belin, 2001; Zatorre and Gandour, 2008) associate left auditory cortex with higher temporal resolution, and hence fine temporal processing, whereas right one with better spectral resolution and analysis of pitch changes (see also Obleser et al., 2008). The asymmetric sampling in time (AST) model (Poeppel, 2003) suggests asymmetry of auditory processing in the time domain. It posits that non-primary auditory regions in the left hemisphere are specialized for extracting information over shorter temporal integration windows, corresponding to acoustic-phonetic speech cues, while homologous regions in the right hemisphere show preference for longer windows, corresponding to prosodic cues (Hesling et al., 2005a, 2005b; Liem et al., 2014). A recent meta-analytic review strongly supports this right-hemispheric preference for longer timescales and extends the evidence to include primary auditory areas, while also confirming, with partial support, the left-hemispheric preference for shorter windows (Oderbolz et al., 2025). Thus the left hemisphere may be more specialized for processing of segmental information at the level of phonemes and syllables, whereas the right hemisphere is tuned to sentence level suprasegmental information processing, which includes formation of intonation phrases, syntactic segmentation via boundary marking, and distinguishing declarative sentences from interrogatives and imperatives (Baum and Pell, 1999; Friederici, 2011). These functional interhemispheric asymmetries are embedded within a broader, structurally asymmetric connectome (i.e., the comprehensive map of neural connections), as revealed by an anatomical connectivity study (Mišic et al., 2018). Taken together, the models suggest a dynamic interaction between both hemispheres at different levels of speech processing, supporting the extraction of both semantic and prosodic cues.

Several ’dual-stream’ models further dissociate different aspects of speech processing along dorso-ventral anatomical axis (Friederici, 2020; Hickok and Poeppel, 2007, 2004; Rauschecker and Scott, 2009; Saur et al., 2008). A common anatomical framework describes a ventral stream, projecting ventro-laterally to middle/inferior temporal cortices, and a dorsal stream, projecting dorso-posteriorly via parietal cortex to frontal areas (Friederici, 2020; Hickok and Poeppel, 2007, 2004). While differing in specifics, these models generally agree that the left ventral stream primarily supports semantic processing (DeWitt and Rauschecker, 2012; Pylkkänen, 2019; Rauschecker and Scott, 2009) and building of local syntactic structures (Friederici et al., 2006), whereas the dorsal stream is implicated in speech production (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009) and complex syntactic integration (Bornkessel-Schlesewsky and Schlesewsky, 2013; Friederici, 2020, 2009; Giglio et al., 2024; Wilson et al., 2011). Some evidence also suggests dorsal stream involvement in lexico-semantic processing (e.g. Glasser and Rilling, 2008). In the right hemisphere, both streams contribute to prosodic processing (Frühholz et al., 2015; Sammler et al., 2018, 2015).

Both dorsal and ventral pathways converge on the inferior frontal gyrus (IFG), which is suggested to act as an integrative hub for higher-order speech and language processes (Bornkessel-Schlesewsky and Schlesewsky, 2013; Friederici, 2011; Hagoort, 2013). Within the IFG, a functional dissociation has been proposed between its posterior (pars opercularis) and anterior (pars triangularis) subregions. The IFGop is primarily implicated in phonological and syntactic processing, including hierarchical structure-building and parsing (Friederici, 2020; Hagoort and Indefrey, 2014; Zaccarella et al., 2017), while the IFGtri supports semantic processing (Binder et al., 2009; Friederici, 2020; Hagoort and Indefrey, 2014; van der Burght et al., 2023). Both subregions have also been implicated in prosodic processing (Belyk and Brown, 2014). However, the lateralization of intonation processing may depend on its role in syntactic processing: left-lateralized activity dominates when intonation is critical for syntactic resolution, while right IFG involvement is enhanced when intonation does not contribute to sentence comprehension (van der Burght et al., 2019).

The current study focuses on the effect of semantic content and intonation on neural networks engaged in speech perception. While numerous studies have examined these features in isolation, the present experiment advances the field by testing both at the same time. We employed carefully controlled speech stimuli. Specifically, the design manipulates both semantic content (connected vs. scrambled speech) and prosodic structure (intoned vs. monotonous speech). This approach supports both conventional fMRI analyses of semantic and prosodic processing within a 2 × 2 factorial design, and investigation of functional connectivity through psychophysiological interaction (PPI) analysis.

Given the proposed dissociation between IFGop and IFGtri in phonological/syntactic versus semantic processing, we aimed to examine whether these subregions exhibit differential functional connectivity profiles during processing of intonation and sentence semantics. By using seed-based PPI analysis centered on left and right IFGop and IFGtri, we tested whether functional specialization of IFG subregions extends beyond local activation patterns to their connectivity dynamics, depending on the type of linguistic information being processed. We use the term ’brain activation’ throughout this manuscript as a convention to refer to task-related BOLD signal changes, acknowledging that BOLD reflects hemodynamic responses that are indirect markers of underlying neuronal activity. We predicted that left IFGtri would exhibit stronger connectivity with temporal and inferior parietal semantic hubs during semantic processing, consistent with its role in lexico-semantic integration, whereas right IFGop may be involved in suprasegmental prosodic processing, particularly when intonation is not structurally informative, consistent with van der Burght et al. (2019). Behaviorally, we expected superior performance for connected (vs. scrambled) and intoned (vs. monotonous) speech consistent with prosody’s facilitatory role (Cutler et al., 1997).

Beyond these predictions, we aimed to provide a network-level perspective on how classical language areas (including IFG) dynamically engage with other cortical regions, particularly those implicated in the salience and/or default mode networks, in response to variations in linguistic structure and meaning.

2. Materials and methods

2.1. Participants

Functional neuroimaging data were acquired in 21 volunteers (8 males, 19 right-handed and 2 left-handed, mean age ± *S*.D.: 29 ± 8.8 years). All participants were healthy, with no history of neurological or psychiatric disorders. Six participants were native English speakers, while 15 were non-native, although proficient English speakers (their mean duration of active use of English ± *S*.D. was 20 ± 9.1 years). The non-native English speakers represented 3 different native languages: Hebrew (11), Russian (3), and German (1). The study was performed according to the 2013 Declaration of Helsinki and approved by the Sheba Medical Centre Research Authority. Written informed consent was obtained from all participants prior to the experiment.

2.2. Experimental conditions

In the present study, we manipulated semantic content and intonation contours of the auditorily presented fragments of continuous speech, which resulted in four experimental conditions:

1. Connected intoned speech
2. Connected monotonous speech
3. Scrambled intoned speech
4. Scrambled monotonous speech

2.3. Speech stimuli

The speech stimuli were created on the basis of the Michigan Corpus of Academic Spoken English (MICASE) which contains authentic speech recorded in university environments.

Connected speech stimuli were comprised of semantically coherent and grammatically well-formed extracts from monologic speech events typical of academic communication, such as lectures, seminars, and conference presentations (Example 1). All the stimuli were meaningful on their own outside the larger context of a speech event. The stimuli were reproduced by a trained speaker, who mimicked the original intonation patterns with high precision. Recordings took place in an acoustically shielded room.

Example 1. The stimulus comprises three excerpts (each shown in brackets) of the same monologic event.

[When we're looking at animal population cycles, we need long-term kinds of studies.] [If you look at how predation might change as the population grows, we might see cycles occurring.] [We have a high prey population, and the predators then eat a lot, and they are doing really well, so they reach a very high level. But then, as they eat up the prey, the prey start declining, and then the predators can't get as much food, and they start declining either through lower birth rates or higher death rates. And then, when the predators get rare, the prey can start increasing again.]

Scrambled speech stimuli represented meaningless, ungrammatical strings of words reproduced by the same speaker who was instructed to mimic intonations from the corresponding meaningful narratives (Example 2, Fig. 1). The word strings were created by shuffling the words from the corresponding meaningful narrative in such a way, that the number of syllables and positions of stressed syllables matched the number of syllables and positions of stressed syllables in prosodic phrases in the meaningful stimulus (Example 2). Therefore, lexical content of the stimuli remained constant across the experimental conditions, which ensured similar lexical retrieval and phonological processing.

Example 2. A. A full version of a scrambled stimulus corresponding to the stimulus given in Example 1. B. A fragment of the scrambled speech stimulus (in bold) versus a corresponding connected speech stimulus.

A. *Up we're through of animal get predation cycles, as eat long-term very studies. If and they the start population when prey as and they are prey the grows, we when at might as occurring. We either prey at population, and they predators reach get a lot, and rates an looking how well, so declining then the can't level. But need, you the start much the prey, then have might declining, and can the predators high look doing food, see the high*

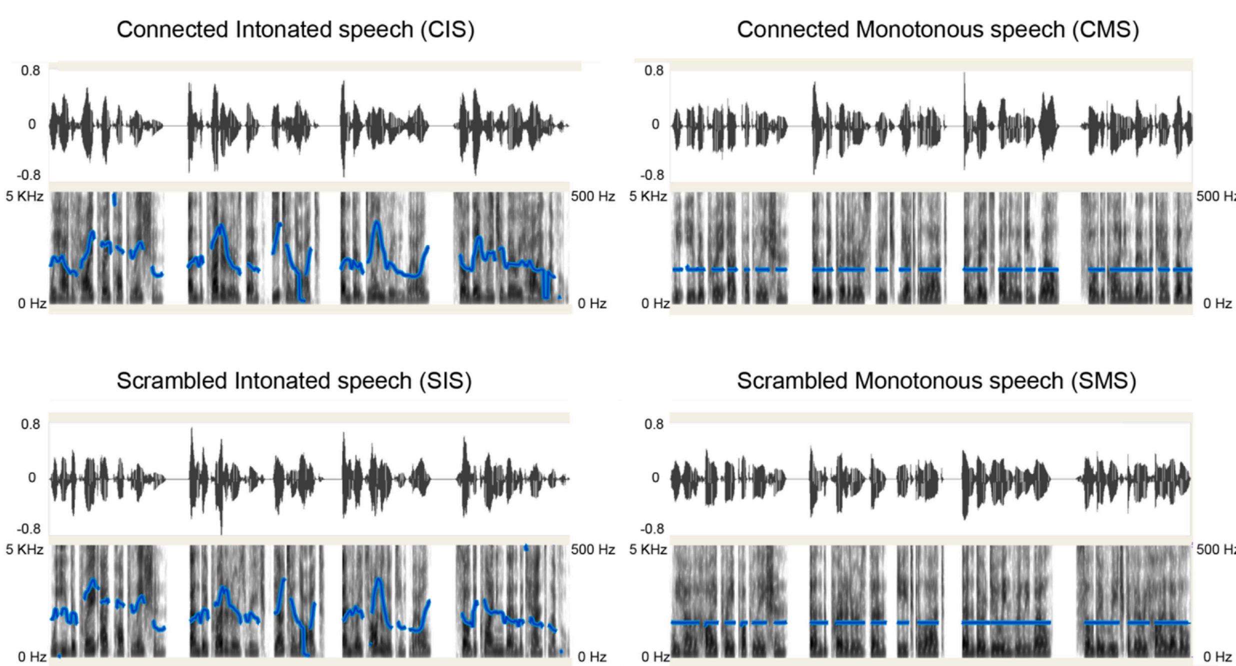

Fig. 1. Examples of the speech stimuli used in four experimental conditions. Upper panel shows the oscillogram of the stimulus, the lower panel shows the stimulus spectrogram and variation of the voice fundamental frequency (f_0) (blue line). The left frequency scale (0–5 kHz) corresponds to the spectrogram, while the right frequency scale (0–500 Hz) corresponds to the superimposed pitch (f_0) contour.

increasing really they higher start then or cycles eat rates. A change, lower predators death rare, then birth we kinds population again.

B.
we have a high prey population /
we either prey at population /
and the predators then eat a lot /
and they predators reach get a lot /
and they're doing really well /
and rates an looking how well /
so they reach a very high level //
so declining then the can't level //

In order to create *monotonous connected or scrambled stimuli*, we instructed the speaker to utter the same narratives as monotonously as possible, and then the residual variations of the fundamental frequency were removed using Praat 6.039 software (Boersma and van Heuven, 2001). This procedure prevented the occurrence of sound artifacts which would have been otherwise present if the fundamental frequency was removed from the original, intoned narratives.

All four types of the stimuli were synchronized to ensure similar speech rate and duration of phonological phrases and gaps. This was achieved by slightly adjusting the tempo of individual words or phrases and/or modifying the duration of pauses between them. Subsequently, the stimuli were intensity-normalized using the RMS (root mean square) function with Adobe Audition CC 2015 (match to total RMS, loudness – 26 dB).

2.4. Experimental design

During the MRI scanning session, subjects listened to the speech stimuli and answered comprehension questions after each stimulus. In a forced-choice paradigm, subjects categorized the "suggested topics" (Example 3) corresponding to each speech stimulus as either plausible or not, by pressing one of two buttons ("yes" or "no") on a response pad.

Example 3. Comprehension questions ("suggested topics") to the speech stimuli presented as Examples 1 and 2.

Scrambled intoned: *The lecture was about hormonal changes throughout the menstrual cycle* (No)

Scrambled monotonous: *The speaker talked about population cycles in a prey-predator system* (Yes)

Connected intoned: *The speaker described the relationship between animal population cycles and annual temperature fluctuations* (No)

Connected monotonous: *The lecture was about a dynamics of large-scale animal population cycles* (Yes)

During the experiment, the scrambled speech stimuli were always presented before the corresponding connected speech stimuli to prevent any semantic priming or familiarity effects that might arise from hearing the meaningful version first. Furthermore, the order of intoned and monotonous speech conditions was counterbalanced.

Presentation of the stimuli and collection of behavioral data were controlled by Presentation software (release 23.1, Neurobehavioral Systems, Inc., San Francisco, USA). The speech stimuli and the comprehension questions were presented binaurally through electrostatic MRI-compatible headphones (HP PI US; https://www.crsltd.com/). The onset of each condition (scrambled or connected speech) and the instruction to figure out the topic of the narrative were also announced through the headphones. The subjects were blindfolded throughout the whole MRI scanning session by wearing a sleep mask.

Acquisition of functional data was performed using a block-design paradigm. Each block included a single speech stimulus (duration 46–60 s), which was followed by a comprehension question and a motor response. Blocks were separated by instructions. There were four fMRI runs, each built of 10 blocks. Specifically, one run (12.3 min) included two scrambled intoned, two scrambled monotonous, two connected intoned, and two connected monotonous speech stimuli, and two resting periods of a comparable duration (46–56 s). Thus, each condition, including rest, was repeated 8 times during the experimental session.

Before the scanning session, the subjects completed a 7–10-minute practice session, during which the four versions of one stimulus (not included in the main experimental set) and the corresponding comprehension questions were presented.

2.5. Statistical evaluation of behavioral data

Analysis of behavioral data (percentage of correct responses during the scanning session) was performed using the Aligned Rank Transform (ART) method, which is a non-parametric approach to factorial ANOVA that enables analyzing the main effects and interaction. For the pairwise comparison across the experimental conditions we applied the Durbin-Conover test with Bonferroni correction for multiple comparisons.

2.6. Neuroimaging data acquisition

All neuroimaging data were acquired at the Ruth & Meir Rosental Imaging Center, Reichman University, Herzliya, Israel, using a 3-Tesla Siemens Prisma scanner equipped with a 64-channel head coil.

A high-resolution 3D T1-weighted MPRAGE image (resolution 1 × 1 × 1 mm) was acquired for each subject (TR = 2000 ms, TE = 1.9 ms, TI = 920 ms, flip angle = 9°). The volume consisted of 176 sagittal slices (FOV = 256 × 256 mm, 256 × 256 matrix size).

The fMRI data were acquired using an echo-planar imaging sequence (TR = 2000 ms, TE = 30 ms, flip angle = 80°). The FOV was 248 × 248 mm with a 124 × 124 matrix size; slice thickness 2 mm, spacing between slices = 2.0 mm, resulting in an effective voxel resolution of 2 × 2 × 2 mm. For each run, 370 vol of 72 slices were acquired in a continuous sampling paradigm.

Field-map images were acquired to correct for distortions due to magnetic field inhomogeneities (TR = 648 ms, TE1 = 4.8 ms, TE2 = 7.3 ms, flip angle = 45°, slice thickness = 2 mm, spacing between slices = 2 mm, FOV = 128 × 128 mm, 64 × 64 matrix size).

2.7. Neuroimaging data pre-processing and analysis

The neuroimaging data were analyzed using the FMRIB's software library (FSL, version 6.0.6, Oxford Centre for Functional MRI of the Brain (FMRIB), UK: http://www.fmrib.ox.ac.uk/fsl/) (Jenkinson et al., 2012; Smith et al., 2004).

Prior to the fMRI data analysis, the original T1-weighted MRIs were extracted using BET. Then, field-maps for distortion correction of fMRI data were created using the following steps: brains were extracted from subjects' magnitude field-maps with BET, then the images were eroded to remove noisy edges, and then the field-maps were prepared using the fsl_prepare_fieldmap tool.

Analysis of the functional data was conducted using the FSL FEAT (FMRI Expert Analysis Tool) (Woolrich et al., 2004, 2001). The preprocessing of the functional data included motion correction using MCFLIRT (Jenkinson et al., 2002), temporal high-pass filtering (filter cutoff 120 s), spatial smoothing using a 5-mm Gaussian filter, and denoising the data using MELODIC Independent Component Analysis (ICA).

Registration to the individual's structural MRI and to the standard MNI template was carried out using FLIRT (Greve and Fischl, 2009; Jenkinson et al., 2002; Jenkinson and Smith, 2001).

A general linear model (GLM) was used to analyze the blood oxygen level–dependent (BOLD) responses of each subject as a function of the experimental condition. Regressors of interest consisted of boxcar functions convolved with Gamma function (lag 6 s).

The general linear model included 5 explanatory variables: connected intoned speech (CIS), connected monotonous speech (CMS), scrambled intoned speech (SIS), scrambled monotonous speech (SMS), and rest baseline. Volumes corresponding to the instructions, comprehension questions, and motor responses were excluded from the analysis. The effects of semantic content and intonation presence were tested using the following contrasts: (CIS+CMS) – (SIS+SMS) and (CIS+SIS) – (CMS+SMS), respectively. Similarly, the effects of semantic content and intonation absence were tested using the opposite contrasts: (SIS+SMS) – (CIS+CMS) and (CMS+SMS) – (CIS+SIS). The interactions were tested using the contrasts: (CIS − SIS) − (CMS − SMS) and (SIS − CIS) − (SMS − CMS), which evaluate whether the effect of semantic content differs depending on the presence or absence of intonation.

For the group-level statistical analysis we used mixed effects modeling in order to account for intersubject variability. Z statistic images were thresholded using clusters determined at $Z > 2.3$ at the cluster p threshold of 0.05. This threshold was applied to define functional regions of interest (ROIs). ROIs were labeled using Harward-Oxford Cortical Structural Atlas, Juelich Histological Atlas, and Talairach Daemon Labels.

2.8. Generalized psychophysiological interactions (gPPI)

Psychophysiological interaction analysis (PPI) was used to study changes in functional connectivity across connected and scrambled, intoned and monotonous conditions. It was implemented within the CONN toolbox (version 22a), which performs the first-level analyses in SPM and the second-level random-effects analyses in MATLAB. All participants were included in the gPPI analysis. For each participant, the mean BOLD time series were extracted from the defined ROIs (McLaren et al., 2012). Additionally, we confirmed that all participants exhibited stable BOLD time series in the seed regions and that the extracted signal was not hampered by motion or outliers. Based on prior hypothesis, seed regions included bilateral inferior frontal gyrus pars opercularis and bilateral pars triangularis. The seed regions were defined using the multi-label atlas.nii file from the CONN toolbox's default Harvard-Oxford parcellation. Separately for each pair of seed and revealed target areas, a generalized psychophysiological interaction model was defined with seed BOLD signals as physiological factors, boxcar signals characterizing each individual task condition, convolved with an SPM canonical hemodynamic response function as psychological factors, and the product of the two as psychophysiological interaction terms. Functional connectivity changes across conditions were characterized by the Fisher-transformed semi-partial correlation coefficient of the psychophysiological interaction terms in each model. Group-level analyses were performed using a GLM. For each individual voxel a separate GLM was estimated, with first-level connectivity measures at this voxel as dependent variables. Voxel-level hypotheses were evaluated using multivariate parametric statistics with random-effects across subjects and sample covariance estimation across multiple measurements. Cluster-level inferences were based on parametric statistics from the Gaussian Random Field theory. Results were thresholded using a cluster-forming threshold of $p < 0.001$ (uncorrected) at the voxel level, and a false discovery rate (FDR) corrected threshold of $p < 0.05$ at the cluster level. The applied contrasts were: a) (CIS+CMS) – (SIS+SMS) to evaluate the effect of semantic content, and b) (CIS+SIS) – (CMS+SMS) to evaluate the effect of intonation presence.

3. Results

3.1. Behavioral results

The accuracy of task performance was affected significantly by both main effects, semantic content and intonation presence ($F(1, 60) = 20.7$, $p < 0.001$, and $F(1, 60) = 16.2$, $p < 0.001$ respectively). No significant interaction was found ($F(1, 60) = 0.4$, $p > 0.5$).

Durbin-Conover test showed that task performance accuracy did not differ significantly between the connected intoned and connected monotonous conditions. However, the difference was significant between the intoned and monotonous scrambled conditions ($Z = 3.2$, $p < 0.01$, Bonferroni corrected). The difference was also significant between connected and corresponding scrambled conditions (connected intoned and scrambled intoned: $Z = 2.6$, $p < 0.05$; connected monotonous and scrambled monotonous: $Z = 3.8$, $p < 0.01$, Bonferroni corrected) (Fig. 2).

3.2. fMRI results

Processing of connected intoned speech (compared to rest) produced activation in the primary and non-primary auditory cortex in temporal lobes, bilateral inferior frontal cortex (pars opercularis and pars triangularis), and bilateral cerebellum (Fig. 3). Deactivations were observed within the areas attributed to the default mode network: anterior and posterior cingulate cortex, and bilateral occipito-temporal regions (angular gyrus (AG)/superior occipital cortex).

Connected speech conditions compared to scrambled speech elicited left-lateralized activation within speech-related areas: anterior superior and middle temporal gyri (aSTG/MTG) (left > right), left posterior STG (pSTG)/AG, and left inferior frontal gyrus (IFG, pars triangularis) (Fig. 4A and Table 1). The opposite contrast (scrambled > connected speech) revealed significantly higher amplitude of BOLD response in bilateral dorsolateral prefrontal cortices (DLPFC), right posterior parietal cortex (PPC), bilateral anterior insula (aIns), and anterior cingulate cortex (ACC) (both dorsal and ventral segments) (Fig. 5A).

Processing of intoned compared to monotonous narratives produced bilateral activation within the anterior and posterior temporal regions (Fig. 4B and Table 1). The contrast monotonous > intoned speech showed significant difference in the amplitude of BOLD response within the anterior and posterior cingulate cortex (Fig. 5B).

Furthermore, there was a significant interaction between the main effects of semantic content and intonation. This interaction was confined to a single significant cluster within the right temporo-parietal junction (TPJ), specifically in the supramarginal/angular gyrus region (SMG/AG), a part of the inferior parietal lobule (IPL) (Fig. 4C). This area was consistently deactivated in all conditions except one: scrambled monotonous speech. In the presence of intonation, scrambled speech led to stronger deactivation of the right IPL compared to connected speech, whereas in the monotonous condition, the opposite pattern emerged – the IPL was deactivated by connected, but not by scrambled speech.

3.3. gPPI results

Psychophysiological interaction analyses (gPPI) revealed significant condition-dependent changes in functional connectivity from bilateral inferior frontal gyrus pars opercularis and pars triangularis (IFGop,

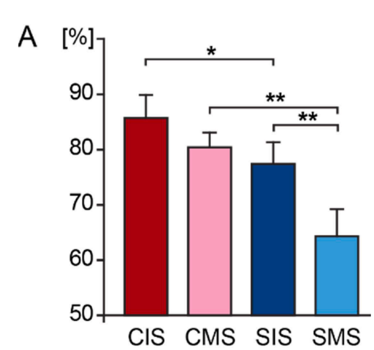

Fig. 2. A. Behavioral results. Accuracy of task performance across the experimental conditions. Color bars represent the mean percentage of correct responses (y-axes), vertical lines represent standard error of the mean (SEM). CIS = connected intoned speech, CMS = connected monotonous speech, SIS = scrambled intoned speech, SMS = scrambled monotonous speech. *$P < 0.05$, **$P < 0.01$.

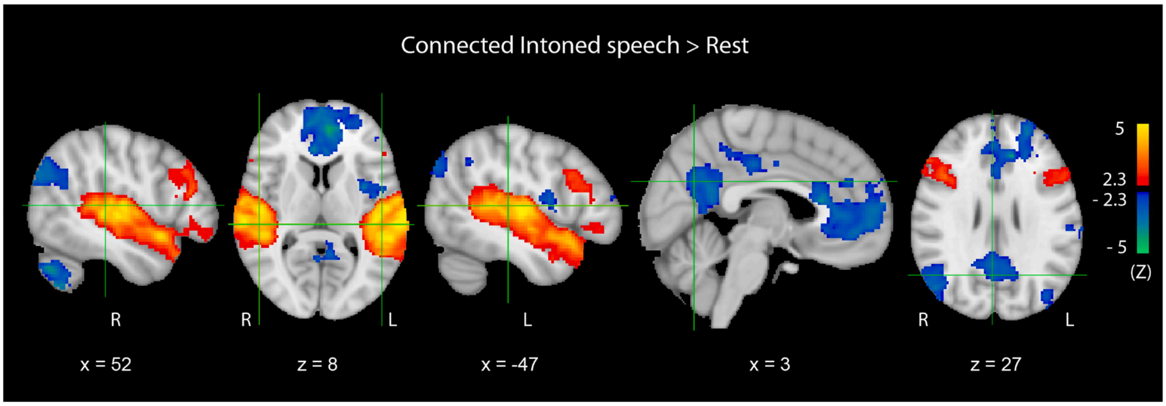

Fig. 3. Activation maps resulting from the contrast connected intoned speech > rest superimposed onto the MNI152 standard brain template. Brain regions activated during speech processing are shown in red-yellow, while deactivated in blue-green; $p < 0.05$, corrected with threshold-free cluster enhancement.

IFGtri) seed regions. The contrast comparing connected and scrambled conditions identified significant clusters showing increased connectivity with these seed regions. Specifically, bilateral IFGop, and right IFGtri showed increased connectivity with left-hemispheric clusters. For the left IFGop the target clusters were centered at MNI (x y z): –58 –58 18 and –46 –66 48, for the right IFGop the target cluster was centered at MNI: –42 –72 42, and for the right IFGtri the target cluster was centered at MNI: –42 –60 40. Furthermore, both left IFGop and left IFGtri showed increased FC with left posterior temporal cortex (pSTG). For the left IFGop the target cluster was centered at MNI: –54 –32 6, and for the left IFGtri at –46 –36 4 (Table 2).

The contrast comparing intoned versus non-intoned conditions revealed significant target clusters showing increased and decreased connectivity for IFGop. Specifically, the right IFGop showed increased FC with a cluster in left posterior temporal cortex (pMTG) (–70 –34 4), while the left IFGop showed decreased FC with right IPL/LOC (52 –68 18) (Fig. 6).

4. Discussion

In the present study we combined activation and functional connectivity analyses to provide a systems-level account of natural speech comprehension, revealing how large-scale brain networks dynamically reconfigure in response to variations in semantic coherence and prosodic structure. We employed a 2 × 2 factorial fMRI design to investigate the neural substrates underlying the processing of semantic content and intonation in continuous speech. This design enabled identification of distinct patterns of brain activation and functional connectivity associated with these linguistic features. Whole-brain fMRI analyses revealed: (1) engagement of left-hemispheric language networks during semantic processing, (2) bilateral activation in temporal lobes for intonation processing, and (3) recruitment of domain-general control networks (salience and executive networks) during degraded speech comprehension. Psychophysiological interaction (PPI) analyses demonstrated condition-dependent coupling between IFG subregions and posterior temporal/parietal areas. Specifically, bilateral IFGtri and IFGop showed enhanced connectivity with left temporal and inferior parietal regions during semantic processing, while right IFGop exhibited preferential connectivity with left temporal regions during intonation processing. Conversely, left IFGop displayed reduced connectivity with the right IPL during intoned speech processing. These findings reveal adaptive modulation of both localized activity and large-scale network interactions during speech perception.

4.1. Behavioral data

Analysis of the behavioral data revealed that task performance accuracy was affected by both the presence of semantic content and intonation (Fig. 1A). As expected, subjects answered comprehension questions more accurately for meaningful than for scrambled speech stimuli. Notably, participants performed above chance in the scrambled condition, even though the syntactic and semantic structure of the speech was disrupted. This suggests that listeners were able to extract informative cues from the auditory stream despite the lack of coherent narrative. A likely explanation is that participants relied on isolated content words, such as nouns and verbs, which remained intact during the scrambling process. These lexical cues, even when presented without proper grammatical structure, can convey partial semantic information that supports topic inference. Although performance in the scrambled conditions was significantly lower than in the corresponding connected conditions, the above-chance accuracy highlights the robustness of lexical-semantic processing and supports the notion that listeners can engage in meaningful interpretation based on sparse or degraded linguistic input.

Furthermore, participants performed significantly better in the scrambled intoned condition compared to the scrambled monotonous one. This suggests that prosodic cues may facilitate comprehension or topic inference even in the absence of the semantic and syntactic context that typically constrains prosody interpretation (Cole, 2015). One possible explanation is that natural intonation enhances the perceptual salience of key content words, making them more accessible and memorable. Prosody may also help listeners impose a pseudo-structure or grouping on an otherwise unstructured sequence, thereby supporting lexical retrieval, encoding in working memory, and hypothesis formation. The right hemisphere’s specialization for slower temporal modulations, essential for extracting prosodic information (Oderbolz et al., 2024), can provide a structural scaffold to guide meaning inference from degraded speech. Additionally, natural prosodic contours may enhance general attentional engagement with the stimulus (Fernald and Kuhl, 1987; Liu et al., 2022; Sullivan and Horowitz, 1983), leading to improved task performance. In contrast, the absence of a significant prosodic benefit in the connected speech conditions may reflect reduced reliance on prosody when rich semantic and syntactic information is already available.

4.2. Functional MRI data

Activation data revealed the classical left-lateralized language network in the processing of semantic content, and bilateral superior temporal areas in the processing of intonation.

Specifically, compared to rest, connected intoned speech engaged both auditory and language networks, which included primary and non-primary auditory cortex, bilateral inferior frontal cortex, and bilateral cerebellum, and induced deactivation of the default mode network

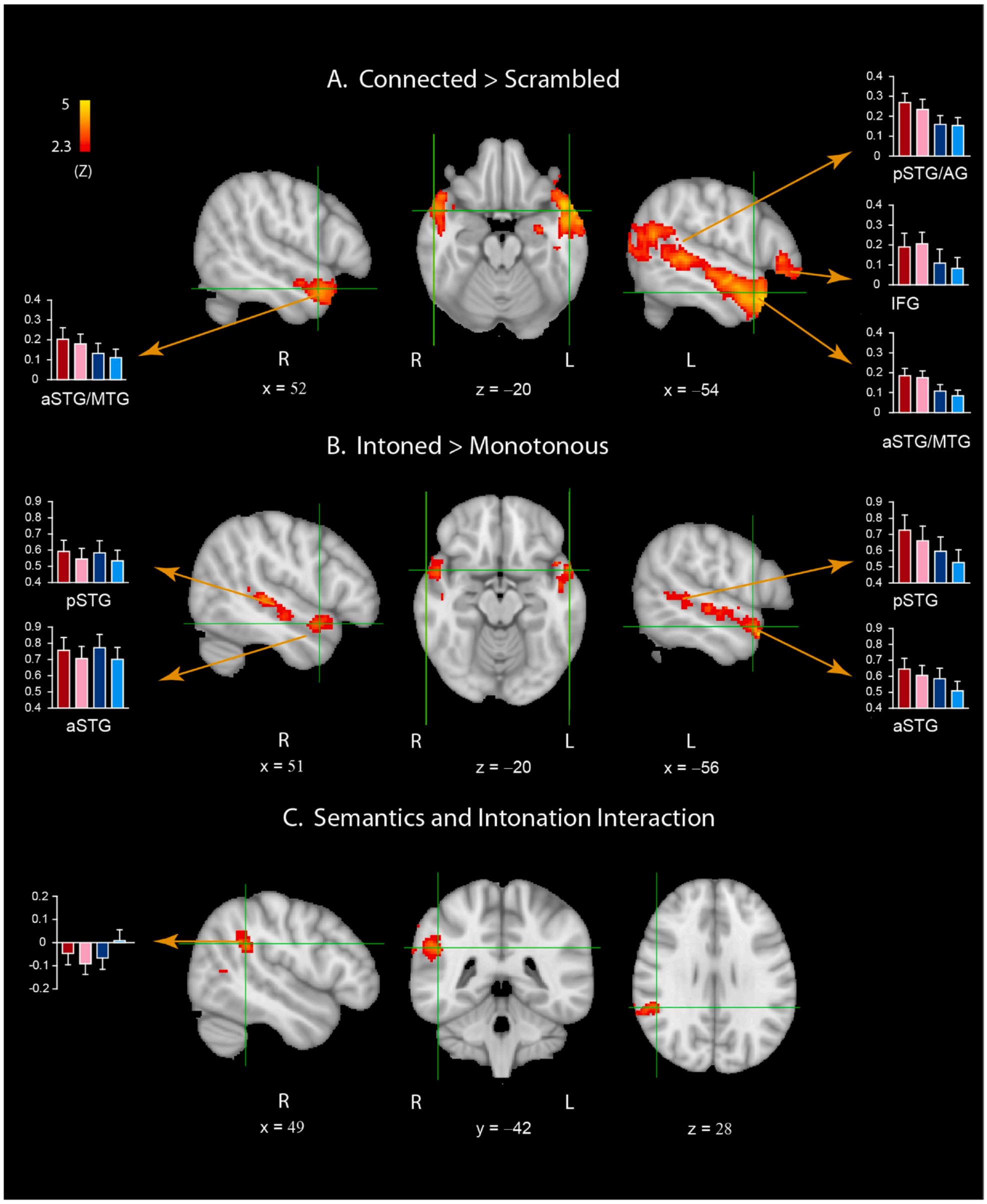

Fig. 4. The effects of semantic content and intonation presence (A, B) and semantics and intonation interaction (C). A. Activation maps resulting from the contrast connected > scrambled speech. B. Activation maps resulting from the contrast intoned > monotonous speech. C. Semantics and intonation interaction. The maps are superimposed onto the MNI152 standard brain template. Bar charts represent distributions of the BOLD percent signal change within the regions of interest across the experimental conditions (compared to Rest). The color codes are as in Fig. 2, vertical lines represent standard error of the mean (SEM). $P < 0.05$, corrected with threshold-free cluster enhancement.

(DMN) (Fig. 3).

Further comparisons between connected and scrambled speech, as well as between intoned and monotonous speech, allowed us to isolate semantic and intonation effects, while controlling for low-level auditory and lexical features. It has earlier been shown that amplitude of activation in language-related areas is higher during listening to sentences with greater intelligibility (Kyong et al., 2014), and with natural compared to flattened intonation contour (Zhao et al., 2008). In line with these studies, we found higher activation in language-related areas in both contrasts.

As expected from previous observations (Binder et al., 2009; Fedorenko et al., 2015; Price, 2000), the resulting activation pattern specific to semantic processing included bilateral anterior temporal association areas (STG/MTG) with left-hemispheric dominance, left posterior temporal cortex, left IPL, and the left inferior frontal gyrus (Fig. 4A).

At the same time, intoned compared to monotonous narratives involved bilateral anterior and posterior temporal regions (Fig. 4B). This finding is consistent with the results of another fMRI study specifically focused on sentence level prosody, where a functional localization approach was applied (Fedorenko et al., 2015). The opposite contrast (monotonous > intoned) (Fig. 5B) revealed weaker deactivation of the anterior and posterior cingulate cortex, a core part of the DMN, as may be expected for a less sensory-rich input (Smallwood et al., 2021), since the DMN is known to be deactivated by externally-oriented, i.e. stimulus-driven perceptual and cognitive tasks and activated by internally-oriented mental processes (see Menon, 2015 for review; Sestieri et al., 2011; Sormaz et al., 2018).

A more revealing pattern emerged when we contrasted activation elicited by connected speech with its scrambled counterpart. This contrast revealed an extensive set of cortical structures: bilateral DLPFC, right PPC, bilateral aIns, and both dorsal and ventral segments of the

Table 1
Brain regions showing significant effects of semantic content presence or absence, intonation presence or absence, and their interaction.

Contrast	Brain area	MNI coordinates (mm)			Z-value
		x	y	z	
Connected > Scrambled	aSTG/MTG, LH	−51	16	−25	5.1
	aSTG/MTG, RH	48	8	−24	4.5
	pSTG, LH	−52	−40	4	4.3
	AG, LH	−54	−60	20	3.9
	IFG, LH	−53	26	2	3.7
Intoned > Monotonous	aSTG, LH	−56	10	−16	3.9
	aSTG, RH	63	0	−4	4.3
	pSTG, LH	−60	−40	4	2.9
	pSTG, RH	54	−28	2	3.8
Scrambled > Connected	DLPFC, LH	−40	39	26	4.3
	DLPFC, RH	30	40	27	4.0
	aIns, LH	−38	14	−4	4.6
	aIns, RH	44	18	−3	4.3
	PPC, RH	54	−42	44	3.7
	ACC, dorsal, LH	−7	30	29	4.1
	ACC, dorsal, RH	4	22	35	4.3
	ACC, ventral, LH	−9	38	0	4.1
	ACC, ventral, RH	8	38	4	4.4
Monotonous > Intoned	ACC, ventral, LH	−2	34	7	3.4
	ACC, ventral, RH	4	40	0	3.7
	PCC, LH	−8	−56	22	3.6
	PCC, RH	6	−50	20	3.8
Semantics x Intonation Interaction	AG, RH	50	−42	26	3.9

aSTG and pSTG = anterior and posterior temporal gyrus respectively, MTG = middle temporal gyrus, AG = angular gyrus, IFG = inferior frontal gyrus, DLPFC = dorsolateral prefrontal cortex, aIns = anterior insula, ACCd and ACCv = dorsal and ventral anterior cingulate cortex respectively, PCC = posterior cingulate cortex, LH and RH = left and right hemisphere.

ACC (Fig. 5A). Besides the DMN, represented by its anterior node (ACCv), components of two other large-scale networks were distinguished: the salience network (SN) (bilateral aIns and dorsal ACC), and the central executive (fronto-parietal) network (CEN). Similarly, stronger activation of prefrontal cortex, aIns, and the deep frontal operculum were shown in response of different kinds of degraded compared to normal speech inputs (Meyer et al., 2004, 2002).

The salience network has been suggested to detect biologically and cognitively relevant events to guide flexible behavior (Menon, 2015; 2011; Menon and Uddin, 2010; Sridharan et al., 2008). Specifically, it aids in prioritizing and allocating attention to potentially important or unexpected stimuli to regulate behavioral adaptation to environmental changes by recruiting relevant networks for further processing. Scrambled speech used in our experiment represents a form of unexpected or incongruent stimulus, as it lacks the usual syntactic structure and semantic content present in normal speech. This incongruity makes it stand out against the background of typical linguistic inputs. Fig. 5(A) shows that bilateral aIns and the dorsal ACC are selectively activated by scrambled speech conditions, with stronger activation associated with the most distant from normal speech stimulus, the monotonous scrambled one, since in addition to syntactic and semantic structure, it is lacking intonation contour.

In previous studies involvement of the SN was shown to occur along with deactivation of the DMN and activation of the CEN. Furthermore, shorter peak latency of activation of the aIns and ACC allowed suggesting that the SN plays a critical and causal role in switching between the DMN and the CEN (Menon, 2015; Sridharan et al., 2008).

In the present study, deactivation of the DMN was observed in all experimental conditions (when compared to the baseline condition). Specifically, the anterior cluster (ACCv) was deactivated more strongly by connected compared to scrambled speech. Although in most neuroimaging studies the DMN is described as a homogenous network, some authors point out functional differentiation within the DMN (Sestieri et al., 2011; Uddin et al., 2009). Results of resting-state functional connectivity analysis indicated that the anterior and posterior nodes of the DMN exhibited different relationships with other networks (Uddin et al., 2009).

An important addition to the revealed pattern of activations comes from the interaction effect between semantics and prosody, which was confined to a single significant cluster in the right TPJ/IPL. This finding reflects how prosodic cues influence the processing of semantic content in speech, and/or vice versa. This region was consistently deactivated across all experimental conditions except one: scrambled monotonous speech, in which it remained at baseline. This pattern suggests the region's specific sensitivity to the combination of degraded syntax and lack of prosodic structure. Unlike the left IPL, which is often implicated in semantic and language integration (Kuhnke et al., 2023), the right IPL is considered a core hub of the DMN and has been linked to social cognition, perspective-taking, and internally guided information processing (Buckner et al., 2008; Mars et al., 2012). The lack of deactivation during scrambled monotonous speech may reflect the right IPL engagement in internally oriented processes in response to highly impoverished input.

Activation of the SN due to recognition of scrambled speech as unusual and potentially significant stimuli likely reflects the brain's effort to make sense of the scrambled input or to monitor and resolve the conflict created by the lack of coherence. This effort may be supported by the CEN through the activation of attention and working memory mechanisms, both necessary for detecting, reordering, and retaining content words from meaningless word streams to facilitate making a judgment regarding possible meaning. Indeed, right DLPFC and right PPC were found to be selectively activated by scrambled speech stimuli (Fig. 5A). Furthermore, trying to comprehend scrambled speech may involve an internally-oriented component, such as building and evaluating hypotheses on the basis of general knowledge retrieved from long-term memory, rejecting inappropriate ones, creating new, and finally making a decision regarding the meaning. The ACC is known to play a significant role in decision-making (Rangel et al., 2008), therefore, deactivation in the anterior node of the DMN was weaker during meaningless speech processing. At the same time, switching between these two information processing streams conducted by the CEN and DMN may be accomplished by the salience network.

4.3. Psychophysiological interaction (PPI) results

To investigate context-dependent functional connectivity of regions implicated in speech comprehension, we conducted a psychophysiological interaction (PPI) analysis using IFG subcomponents (pars triangularis and pars opercularis) as the seed regions. The seeds were selected based on their established role in semantic and prosodic processing, and integrative speech functions (Friederici and Alter, 2004; Kristensen et al., 2013; Sammler et al., 2015). We hypothesized that IFG connectivity would dynamically reconfigure depending on the availability of semantic and prosodic cues in speech stimuli, reflecting the neural adaptability underlying speech comprehension.

The PPI analysis revealed condition-dependent changes in functional connectivity between bilateral IFG subregions and other key cortical regions, further elucidating the neural dynamics underlying semantic and prosodic processing.

Compared to scrambled speech, connected speech (regardless of intonation) elicited stronger connectivity between bilateral IFG (including both IFGop and IFGtri) and left-hemispheric regions critical for semantic processing: the angular gyrus – a possible hub for semantic

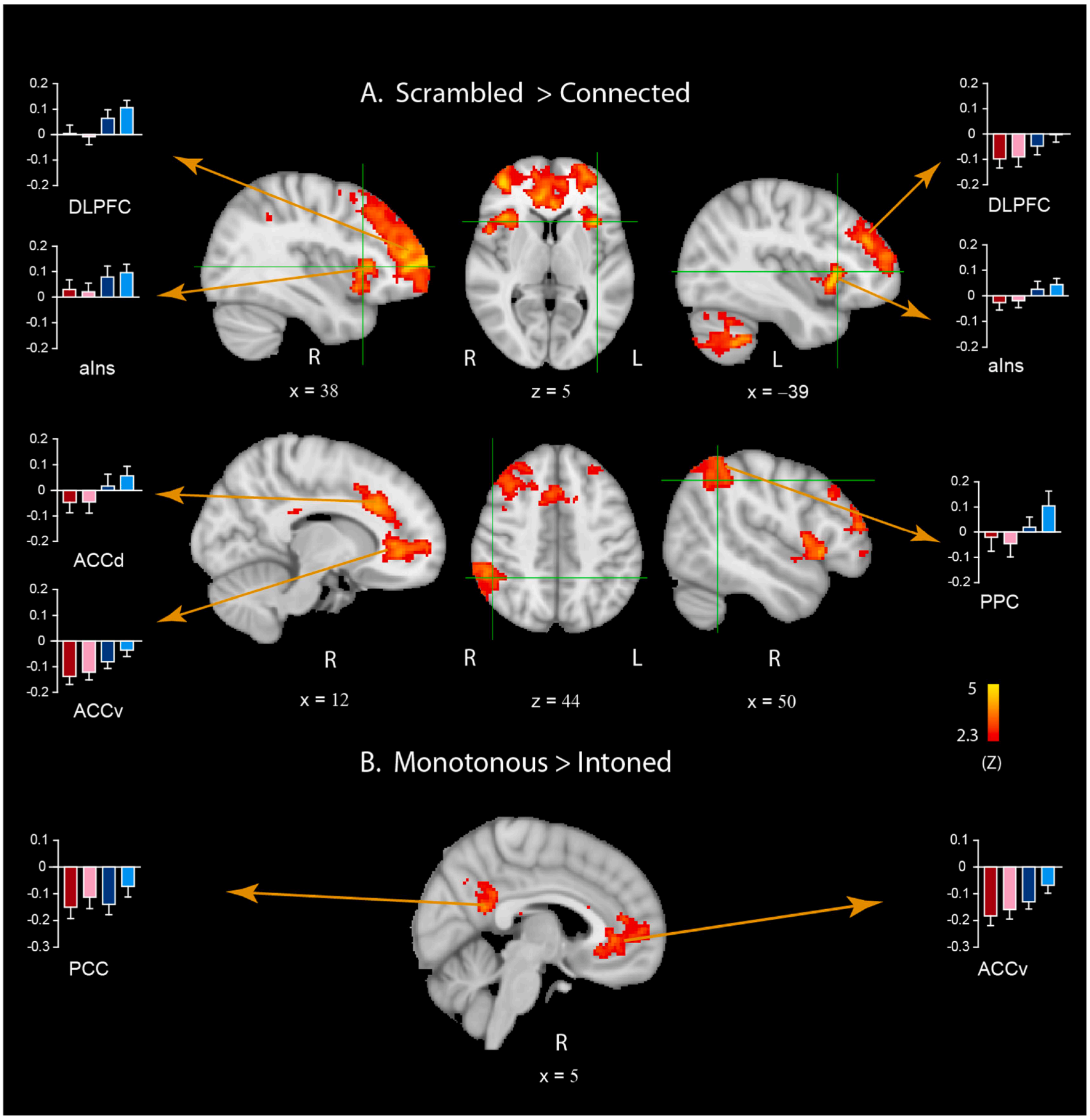

Fig. 5. The effects of semantic content and intonation absence. A. Activation maps resulting from the contrast scrambled > connected speech. B. Activation maps resulting from the contrast monotonous > intoned speech. All other designations are as in Fig. 4.

Table 2
Brain regions showing significant changes in functional connectivity with the seed regions during semantic content presence or absence, and intonation presence or absence.

Contrast	Seed	Brain area	MNI coordinates (mm)			Cluster size	p-FDR	Direction of change
			x	y	z			
Connected > Scrambled	IFGop, LH	IPL, LH	−58	−58	18	66	0.0063	Increase
		pSTG, LH	−54	−32	6	39	0.044	Increase
		IPL, LH	−46	−66	48	37	0.044	Increase
	IFGop, RH	IPL, LH	−42	−72	42	72	0.0025	Increase
	IFGtri, LH	pSTG/MTG, LH	−46	−36	4	59	0.0109	Increase
	IFGtri, RH	IPL, LH	−42	−60	40	76	0.0019	Increase
Intoned > Monoton	IFGop, LH	IPL/LOC, RH	52	−68	18	45	0.0453	Decrease
	IFGop, RH	pMTG, LH	−70	−34	−4	74	0.0016	Increase

IFGop = inferior frontal gyrus, pars opercularis, IFGtri = inferior frontal gyrus, pars triangularis, IPL = Inferior Parietal Lobule, LOC = Lateral Occipital Cortex. All other designations are as in Table 1.

integration (Binder et al., 2009) – and the posterior temporal cortex. This finding aligns with the neurocognitive model proposed by Jackson and colleagues (2019), who identified two distinct semantic subnetworks: a temporal component for representation, and a fronto-temporal "semantic control network" (including IFG and pMTG), which, together with the IPL, supports controlled retrieval and manipulation of semantic knowledge (Hodgson et al., 2021; Jackson et al., 2019; Noonan et al., 2013).

Our results indicate that coherent semantic content primarily enhances connectivity within this fronto-temporo-parietal control

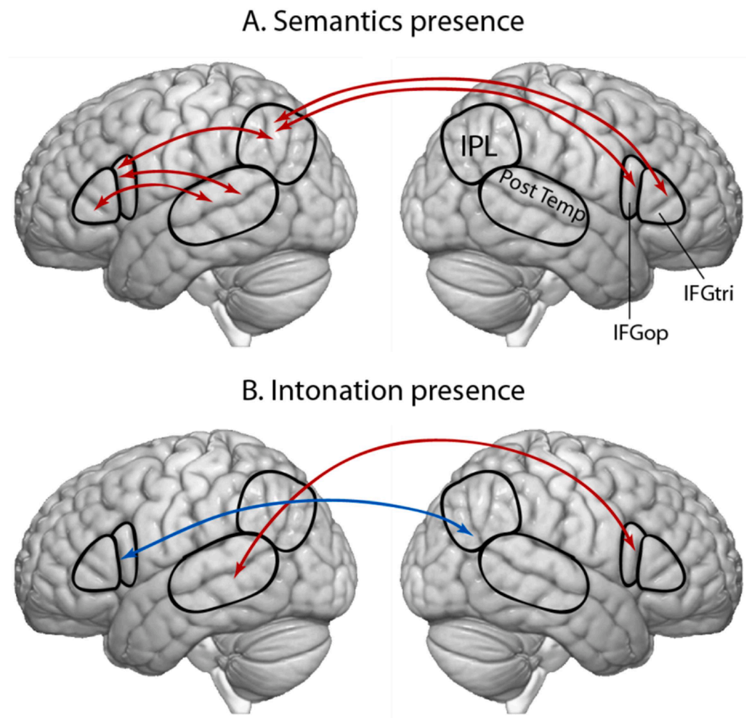

Fig. 6. Condition-dependent changes in functional connectivity from bilateral IFG seeds (pars triangularis and pars opercularis) as revealed by gPPI analysis. Connectivity changes associated with semantic content (A) and intonation presence (B) are shown with arrows (red – increased, blue – decreased connectivity). The right hemisphere is shown on the right.

subnetwork. While the temporal cortex serves as a principal site for storing conceptual knowledge (Binder et al., 2009; Jackson et al., 2019) and supports predictive coding (Service et al., 2007), lexico-semantic mapping, and concept retrieval, the observed IFG coupling likely reflects top-down modulation of left-hemispheric semantic hubs by control regions. This facilitation enables processes such as controlled retrieval, ambiguity resolution, and semantic unification (Hagoort, 2013; Noonan et al., 2013) – all of which are disrupted in scrambled speech. The finding that coherent semantic content enhances left IFGop (BA 44) connectivity with temporal and parietal regions is also in line with the view of this region as a hub for hierarchical structure-building (Friederici, 2020), suggesting this function is critical for integrating meaning during the comprehension of continuous narratives.

Notably, the bilateral involvement of the IFG aligns with growing evidence that language processing is not strictly left-lateralized. Neuroimaging studies increasingly report robust activation of right-hemispheric homologs of classical language structures during complex linguistic tasks, including metaphor interpretation, humor appreciation, and inference generation (Jung-Beeman, 2005; Noonan et al., 2013). Our connectivity results further emphasize the specific role of IFGtri in semantic (rather than prosodic) processing. Thus, connectivity changes induced by semantic content presence suggest that synchronized activity between bilateral IFG regions and left-hemispheric temporo-parietal cortex forms a core mechanism for comprehending semantically rich speech.

Intoned compared to monotonous speech (regardless of semantic coherence) enhanced functional connectivity between the right IFGop and left posterior temporal cortex, reflecting cross-hemispheric integration of prosodic and linguistic information, particularly in the presence of rich intonation contours that signal syntactic and pragmatic boundaries. This finding converges with a prior PPI study demonstrating that processing of linguistic intonation strengthens bilateral IFG-temporal connectivity, linking auditory perception to higher-order phonological processes, such as phonological labeling and categorization of intonation contours (Chien et al., 2021). Moreover, it further corroborates meta-analytic evidence highlighting specialized role of the pars opercularis in processing of linguistic intonation (Belyk and Brown, 2014).

In contrast, the left IFGop exhibited reduced connectivity with the right IPL in the presence of intonation contour (or, equivalently, increased connectivity in its absence). The right IPL was consistently deactivated across experimental conditions, suggesting its association with the default mode network. The observed advantage for intoned speech at the behavioral level, reflected in higher comprehension accuracy, suggests that prosodic cues facilitate comprehension by reducing inferential demands. When natural intonation is present, its acoustic markers support semantic integration by signaling syntactic boundaries and emphasizing salient content. The reduction in left IFGop–right IPL connectivity during naturally intoned speech, coupled with the right IPL's DMN-like deactivation profile, may reflect a context-dependent decoupling of language control regions from the DMN. This decoupling could potentially free cognitive resources by minimizing processes like mind-wandering, thereby facilitating the cross-hemispheric integration of prosody observed in the concurrent right IFGop–temporal connectivity increase. Conversely, in the absence of prosody, increased coupling between the right IPL and left IFGop may compensate for this deficit by recruiting internally-oriented resources when confronted with degraded input.

While three experimental conditions used in the present study are perceptually atypical, they are highly informative for delineating the neural architecture of speech comprehension. Studying the brain response to degraded input reveals the boundaries and compensatory mechanisms of the system. Specifically, it helps to dissociate a core language network that is sufficient for processing coherent speech from domain-general control systems that are recruited as compensatory resources when processing demands exceed the capacity of the core system. Therefore, these conditions reveal the full set of neural resources that support successful comprehension in suboptimal, real-world listening situations.

Taken together, our fMRI and PPI results reveal a dual-process mechanism for adaptive speech comprehension. The activation data demonstrate that a canonical language network supports the processing of coherent speech, while domain-general systems like the salience and executive networks may be additionally recruited to handle degraded, unpredictable input. The connectivity profiles extend this view by showing that the inferior frontal gyrus may support this adaptation through a functional division of labor: while bilateral IFGtri is associated with semantic network, the IFGop demonstrated functional flexibility, shifting its connectivity between a cross-hemispheric prosodic-integration state and a fronto-parietal configuration that is engaged during higher demanding processing of monotonous input. This network-level perspective elucidates how the brain dynamically reconfigures its resources to manage the varying cognitive challenges of varying speech conditions.

The dual-process mechanism is consistent with an evolutionary perspective. The functional dissociation we observed between a specialized language network for structured input and domain-general systems for degraded speech likely reflects a layered cognitive architecture. The robust recruitment of the salience and executive networks for scrambled speech suggests the re-engagement of phylogenetically older systems. These systems, which subserve fundamental functions like threat detection and conflict resolution, may be leveraged in humans to handle the cognitive challenge of degraded input. This capacity is supported by evolutionary precursors, including a conserved supramodal internal model for auditory-motor sequencing (Archakov et al., 2020). In humans, this general circuit may have been refined for vocal communication through rhythmic auditory-motor synchronization (Morillon et al., 2019; Poeppel and Assaneo, 2020), a process consistent with genetic and anatomical changes supporting increased brain complexity (Benítez-Burraco and Elvira-García, 2023; Friederici, 2017).

It is important, however, to note that while PPI reveals condition-dependent changes in functional coupling, it does not infer directional influence. Therefore, our conclusions about network reconfiguration should be considered correlational, highlighting the need for future research using causal modeling approaches to determine the directed interactions that underlie this adaptive flexibility.

5. Conclusion

The present results reveal task-dependent reorganization of large-scale neural networks during speech processing, modulated by the integrity of semantic content and prosodic structure. The fMRI data demonstrated a canonical left-lateralized language network for semantic processing and bilateral temporal involvement for intonation, with DMN suppression during externally focused speech comprehension. Crucially, scrambled speech engaged the salience network and executive control regions, possibly reflecting effortful meaning extraction. Notably, the right IPL's selective re-engagement during scrambled monotonous speech suggests a shift toward internally-guided processing when external input lacks both semantic and prosodic structure.

The PPI analysis extended these findings by showing that bilateral IFG flexibly modulates its connectivity based on linguistic context: stronger coupling with left posterior temporal cortex and IPL during connected speech reflects top-down semantic integration, while cross-hemispheric right IFG-left temporal connectivity supports prosodic processing. Together, these results highlight a functional dissociation in which specialized language networks and domain-general control systems are differentially engaged depending on the availability of semantic and prosodic structure. The IFG may play a context-dependent regulatory role by adapting its functional connections to optimize comprehension – engaging left posterior semantic regions for coherent speech while recruiting internally-oriented resources when confronted with degraded input. This network-based perspective offers a more thorough understanding of how the brain adapts to varying degrees of linguistic coherence and prosodic richness – a dimension still relatively underexplored in speech research.

Data availability statement

The data underlying the results in figures are available from the corresponding author upon reasonable request. The data was collected under provision of informed consent of the participants. Access to the data will be granted in line with that consent, subject to approval by the project ethics board and under a formal Data Sharing Agreement.

CRediT authorship contribution statement

Irina Anurova: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katarzyna Ciesla:** Writing – review & editing, Methodology, Investigation. **Amir Amedi:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Ms. Dalit Shlayn for her highly professional assistance with neuroimaging data acquisition. We would also like to express our gratitude to Professor Hanna Renvall for her insightful comments and constructive feedback on the revised version of this manuscript. This research was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 101017884 and the European Research Council under the European Union's Horizon 2020 research and innovation program grant agreement No 773121.

References

Archakov, D., DeWitt, I., Kuśmierek, P., Ortiz-Rios, M., Cameron, D., Cui, D., Morin, E.L., VanMeter, J.W., Sams, M., Jääskeläinen, I.P., Rauschecker, J.P., 2020. Auditory representation of learned sound sequences in motor regions of the macaque brain. Proc. Natl. Acad. Sci. USA 117, 15242–15252. https://doi.org/10.1073/pnas.1915610117.

Baum, S.R., Pell, M.D., 1999. The neural bases of prosody: insights from lesion studies and neuroimaging. Aphasiology. 13, 581–608. https://doi.org/10.1080/026870399401957.

Belyk, M., Brown, S., 2014. Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. Soc. Cogn. Affect. Neurosci. 9, 1395–1403. https://doi.org/10.1093/scan/nst124.

Benítez-Burraco, A., Elvira-García, W., 2023. Human self-domestication and the evolution of prosody. Lang. Linguist. Compass. 17, e12485. https://doi.org/10.1111/lnc3.12485.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex. 19, 2767–2796. https://doi.org/10.1093/cercor/bhp055.

Boersma, P., van Heuven, V., 2001. Speak and unSpeak with PRAAT. Glot International 5, 341–347. https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf.

Bornkessel-Schlesewsky, I., Schlesewsky, M., 2013. Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. Brain Lang. 125, 60–76. https://doi.org/10.1016/j.bandl.2013.01.010.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S., Frith, C.D., 1994. The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. Brain 117 (Pt 6), 1241–1253. https://doi.org/10.1093/brain/117.6.1241.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 1124, 1–38. https://doi.org/10.1196/annals.1440.011.

Chien, P.-J., Friederici, A.D., Hartwigsen, G., Sammler, D., 2021. Intonation processing increases task-specific fronto-temporal connectivity in tonal language speakers. Hum. Brain Mapp. 42, 161–174. https://doi.org/10.1002/hbm.25214.

Cole, J., 2015. Prosody in context: a review. Language, Cognition and Neuroscience 30, 1–31. https://doi.org/10.1080/23273798.2014.963130.

Cutler, A., Dahan, D., van Donselaar, W., 1997. Prosody in the comprehension of spoken language: a literature review. Lang. Speech. 40 (Pt 2), 141–201. https://doi.org/10.1177/002383099704000203.
DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. u S. a 109, E505–E514. https://doi.org/10.1073/pnas.1113427109.
Fedorenko, E., Behr, M.K., Kanwisher, N., 2011. Functional specificity for high-level linguistic processing in the human brain. Proc. Natl. Acad. Sci. u S. a 108, 16428–16433. https://doi.org/10.1073/pnas.1112937108.
Fedorenko, E., Hsieh, P.-J., Balewski, Z., 2015. A possible functional localizer for identifying brain regions sensitive to sentence-level prosody. Lang. Cogn. Neurosci. 30, 120–148. https://doi.org/10.1080/01690965.2013.861917.
Fernald, A., Kuhl, P., 1987. Acoustic determinants of infant preference for motherese speech. Infant. Behav. Dev. 10, 279–293. https://doi.org/10.1016/0163-6383(87)90017-8.
Flinker, A., Doyle, W.K., Mehta, A.D., Devinsky, O., Poeppel, D., 2019. Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. Nat. Hum. Behav. 3, 393–405. https://doi.org/10.1038/s41562-019-0548-z.
Friederici, A.D., 2020. Hierarchy processing in human neurobiology: how specific is it? Philos. Trans. R. Soc. Lond. B Biol. Sci. 375, 20180391. https://doi.org/10.1098/rstb.2018.0391.
Friederici, A.D., 2017. Evolution of the neural language network. Psychon. Bull. Rev. 24, 41–47. https://doi.org/10.3758/s13423-016-1090-x.
Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392. https://doi.org/10.1152/physrev.00006.2011.
Friederici, A.D., 2009. Pathways to language: fiber tracts in the human brain. Trends. Cogn. Sci. 13, 175–181. https://doi.org/10.1016/j.tics.2009.01.001.
Friederici, A.D., Alter, K., 2004. Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang. 89, 267–276. https://doi.org/10.1016/S0093-934X(03)00351-1.
Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proc. Natl. Acad. Sci. USA 103, 2458–2463. https://doi.org/10.1073/pnas.0509389103.
Frühholz, S., Gschwind, M., Grandjean, D., 2015. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. Neuroimage 109, 27–34. https://doi.org/10.1016/j.neuroimage.2015.01.016.
Giglio, L., Sharoh, D., Ostarek, M., Hagoort, P., 2024. Connectivity of Fronto-Temporal Regions in Syntactic Structure Building During Speaking and Listening. Neurobiol. Lang. (Camb) 5, 922–941. https://doi.org/10.1162/nol_a_00154.
Glasser, M.F., Rilling, J.K., 2008. DTI tractography of the human brain's language pathways. Cereb. Cortex. 18, 2471–2482. https://doi.org/10.1093/cercor/bhn011.
Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.
Hagoort, P., 2013. MUC (Memory, Unification, Control) and beyond. Front. Psychol. 4, 416. https://doi.org/10.3389/fpsyg.2013.00416.
Hagoort, P., Indefrey, P., 2014. The neurobiology of language beyond single words. Annu Rev. Neurosci. 37, 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847.
Hesling, I., Clément, S., Bordessoules, M., Allard, M., 2005a. Cerebral mechanisms of prosodic integration: evidence from connected speech. Neuroimage 24, 937–947. https://doi.org/10.1016/j.neuroimage.2004.11.003.
Hesling, I., Dilharreguy, B., Clément, S., Bordessoules, M., Allard, M., 2005b. Cerebral mechanisms of prosodic sensory integration using low-frequency bands of connected speech. Hum. Brain Mapp. 26, 157–169. https://doi.org/10.1002/hbm.20147.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. https://doi.org/10.1038/nrn2113.
Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99. https://doi.org/10.1016/j.cognition.2003.10.011.
Hodgson, V.J., Lambon Ralph, M.A., Jackson, R.L., 2021. Multiple dimensions underlying the functional organization of the language network. Neuroimage 241, 118444. https://doi.org/10.1016/j.neuroimage.2021.118444.
Jackson, R.L., Cloutman, L.L., Lambon Ralph, M.A., 2019. Exploring distinct default mode and semantic networks using a systematic ICA approach. Cortex 113, 279–297. https://doi.org/10.1016/j.cortex.2018.12.019.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6.
Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. Trends. Cogn. Sci. 9, 512–518. https://doi.org/10.1016/j.tics.2005.09.009.
Kandylaki, K.D., Henrich, K., Nagels, A., Kircher, T., Domahs, U., Schlesewsky, M., Bornkessel-Schlesewsky, I., Wiese, R., 2017. Where Is the Beat? The Neural Correlates of Lexical Stress and Rhythmical Well-formedness in Auditory Story Comprehension. J. Cogn. Neurosci. 29, 1119–1131. https://doi.org/10.1162/jocn_a_01122.
Kristensen, L.B., Wang, L., Petersson, K.M., Hagoort, P., 2013. The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. Cereb. Cortex. 23, 1836–1848. https://doi.org/10.1093/cercor/bhs164.
Kuhnke, P., Chapman, C.A., Cheung, V.K.M., Turker, S., Graessner, A., Martin, S., Williams, K.A., Hartwigsen, G., 2023. The role of the angular gyrus in semantic cognition: a synthesis of five functional neuroimaging studies. Brain Struct. Funct. 228, 273–291. https://doi.org/10.1007/s00429-022-02493-y.
Kyong, J.S., Scott, S.K., Rosen, S., Howe, T.B., Agnew, Z.K., McGettigan, C., 2014. Exploring the roles of spectral detail and intonation contour in speech intelligibility: an FMRI study. J. Cogn. Neurosci. 26, 1748–1763. https://doi.org/10.1162/jocn_a_00583.
Liem, F., Hurschler, M.A., Jäncke, L., Meyer, M., 2014. On the planum temporale lateralization in suprasegmental speech perception: evidence from a study investigating behavior, structure, and function. Hum. Brain Mapp. 35, 1779–1789. https://doi.org/10.1002/hbm.22291.
Liu, L., Götz, A., Lorette, P., Tyler, M.D., 2022. How Tone, Intonation and Emotion Shape the Development of Infants' Fundamental Frequency Perception. Front. Psychol. 13, 906848. https://doi.org/10.3389/fpsyg.2022.906848.
Mars, R.B., Neubert, F.-X., Noonan, M.P., Sallet, J., Toni, I., Rushworth, M.F.S., 2012. On the relationship between the "default mode network" and the "social brain. Front. Hum. Neurosci. 6, 189. https://doi.org/10.3389/fnhum.2012.00189.
McLaren, D.G., Ries, M.L., Xu, G., Johnson, S.C., 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage 61, 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068.
Menon, V., 2015. Salience Network. In: Toga, A.W. (Ed.), Brain Mapping: An Encyclopedic Reference. Academic Press: Elsevier, pp. 597–611. https://doi.org/10.1016/B978-0-12-397025-1.00052-X.
Menon, V., 2011. Large-scale brain networks and psychopathology: a unifying triple network model. Trends. Cogn. Sci. 15, 483–506. https://doi.org/10.1016/j.tics.2011.08.003.
Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214, 655–667. https://doi.org/10.1007/s00429-010-0262-0.
Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp. 17, 73–88. https://doi.org/10.1002/hbm.10042.
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 Lang. 89, 277–289. https://doi.org/10.1016/S0093-934X(03)00350-X.
Mišic, B., Betzel, R.F., Griffa, A., de Reus, M.A., He, Y., Zuo, X.-N., van den Heuvel, M.P., Hagmann, P., Sporns, O., Zatorre, R.J., 2018. Network-Based Asymmetry of the Human Auditory System. Cereb. Cortex. 28, 2655–2664. https://doi.org/10.1093/cercor/bhy101.
Morillon, B., Arnal, L.H., Schroeder, C.E., Keitel, A., 2019. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci. Biobehav. Rev. 107, 136–142. https://doi.org/10.1016/j.neubiorev.2019.09.012.
Noonan, K.A., Jefferies, E., Visser, M., Lambon Ralph, M.A., 2013. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. J. Cogn. Neurosci. 25, 1824–1850. https://doi.org/10.1162/jocn_a_00442.
Obleser, J., Eisner, F., Kotz, S.A., 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J. Neurosci. 28, 8116–8123. https://doi.org/10.1523/JNEUROSCI.1290-08.2008.
Oderbolz, C., Poeppel, D., Meyer, M., 2025. Asymmetric Sampling in Time: evidence and perspectives. Neurosci. Biobehav. Rev. 171, 106082. https://doi.org/10.1016/j.neubiorev.2025.106082.
Oderbolz, C., Stark, E., Sauppe, S., Meyer, M., 2024. Concurrent processing of the prosodic hierarchy is supported by cortical entrainment and phase-amplitude coupling. Cereb. Cortex. 34, bhae479. https://doi.org/10.1093/cercor/bhae479.
Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' Speech Communication. The Nature of Speech Perception 41, 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3.
Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. Nat. Rev. Neurosci. 21, 322–334. https://doi.org/10.1038/s41583-020-0304-4.
Pasquinelli, R., Tessier, A.M., Karas, Z., Hu, X., Kovelman, I., 2023. The Development of Left Hemisphere Lateralization for Sentence-Level Prosodic Processing. J. Speech. Lang. Hear. Res. 66, 1365–1377. https://doi.org/10.1044/2022_JSLHR-22-00103.
Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J. Anat. 197, 335–359. https://doi.org/10.1046/j.1469-7580.2000.19730335.x.
Pylkkänen, L., 2019. The neural basis of combinatory syntax and semantics. Science (1979) 366, 62–66. https://doi.org/10.1126/science.aax0050.
Rangel, A., Camerer, C., Montague, P.R., 2008. A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. 9, 545–556. https://doi.org/10.1038/nrn2357.
Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. https://doi.org/10.1038/nn.2331.
Sammler, D., Cunitz, K., Gierhan, S.M.E., Anwander, A., Adermann, J., Meixensberger, J., Friederici, A.D., 2018. White matter pathways for prosodic structure building: a case study. Brain Lang. 183, 1–10. https://doi.org/10.1016/j.bandl.2018.05.001.
Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P.E.G., Belin, P., 2015. Dorsal and Ventral Pathways for Prosody. Curr. Biol. 25, 3079–3085. https://doi.org/10.1016/j.cub.2015.10.009.

Sankaran, N., Leonard, M.K., Theunissen, F., Chang, E.F., 2024. Encoding of melody in the human auditory cortex. Sci. Adv. 10, eadk0010. https://doi.org/10.1126/sciadv.adk0010.

Saur, D., Kreher, B., Schnell, S., 2008. Ventral and dorsal pathways for language | PNAS. Proc. Natl. Acad. Sci. U.S.A. 105, 18035–18040. https://doi.org/10.1073/pnas.0805234105.

Service, E., Helenius, P., Maury, S., Salmelin, R., 2007. Localization of syntactic and semantic brain responses using magnetoencephalography. J. Cogn. Neurosci. 19, 1193–1205. https://doi.org/10.1162/jocn.2007.19.7.1193.

Sestieri, C., Corbetta, M., Romani, G.L., Shulman, G.L., 2011. Episodic memory retrieval, parietal cortex, and the default mode network: functional and topographic analyses. J. Neurosci. 31, 4407–4420. https://doi.org/10.1523/JNEUROSCI.3335-10.2011.

Sihvonen, A.J., Sammler, D., Ripollés, P., Leo, V., Rodríguez-Fornells, A., Soinila, S., Särkämö, T., 2022. Right ventral stream damage underlies both poststroke aprosodia and amusia. Eur. J. Neurol. 29, 873–882. https://doi.org/10.1111/ene.15148.

Smallwood, J., Bernhardt, B.C., Leech, R., Bzdok, D., Jefferies, E., Margulies, D.S., 2021. The default mode network in cognition: a topographical perspective. Nat. Rev. Neurosci. 22, 503–513. https://doi.org/10.1038/s41583-021-00474-4.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Sormaz, M., Murphy, C., Wang, H.-T., Hymers, M., Karapanagiotidis, T., Poerio, G., Margulies, D.S., Jefferies, E., Smallwood, J., 2018. Default mode network can support the level of detail in experience during active task states. Proc. Natl. Acad. Sci. USA 115, 9318–9323. https://doi.org/10.1073/pnas.1721259115.

Sridharan, D., Levitin, D.J., Menon, V., 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Natl. Acad. Sci. USA 105, 12569–12574. https://doi.org/10.1073/pnas.0800005105.

Sullivan, J.W., Horowitz, F.D., 1983. The effects of intonation on infant attention: the role of the rising intonation contour. J. Child Lang. 10, 521–534. https://doi.org/10.1017/s0305000900005341.

Tang, C., Hamilton, L.S., Chang, E.F., 2017. Intonational speech prosody encoding in the human auditory cortex. Science (1979) 357, 797–801. https://doi.org/10.1126/science.aam8577.

Themistocleous, C., 2025. Linguistic and emotional prosody: a systematic review and ALE meta-analysis. Neurosci. Biobehav. Rev. 175, 106210. https://doi.org/10.1016/j.neubiorev.2025.106210.

Turker, S., Kuhnke, P., Eickhoff, S.B., Caspers, S., Hartwigsen, G., 2023. Cortical, subcortical, and cerebellar contributions to language processing: a meta-analytic review of 403 neuroimaging experiments. Psychol. Bull. 149, 699–723. https://doi.org/10.1037/bul0000403.

Uddin, L.Q., Kelly, A.M., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2009. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. Hum. Brain Mapp. 30, 625–637. https://doi.org/10.1002/hbm.20531.

van der Burght, C.L., Goucha, T., Friederici, A.D., Kreitewolf, J., Hartwigsen, G., 2019. Intonation guides sentence processing in the left inferior frontal gyrus. Cortex 117, 122–134. https://doi.org/10.1016/j.cortex.2019.02.011.

van der Burght, C.L., Numssen, O., Schlaak, B., Goucha, T., Hartwigsen, G., 2023. Differential contributions of inferior frontal gyrus subregions to sentence processing guided by intonation. Hum. Brain Mapp. 44, 585–598. https://doi.org/10.1002/hbm.26086.

Van Lancker, D., Sidtis, J.J., 1992. The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: all errors are not created equal. J. Speech. Hear. Res. 35, 963–970. https://doi.org/10.1044/jshr.3505.963.

Van Lancker Sidtis, D., Pachana, N., Cummings, J.L., Sidtis, J.J., 2006. Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. Brain Lang. 97, 135–153. https://doi.org/10.1016/j.bandl.2005.09.001.

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2011. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. Neuroimage 54, 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036.

Wildgruber, D., 2009. A cerebral network model of speech prosody comprehension. Int. J. Speech. Lang. Pathol. 11, 277—281. https://doi.org/10.1080/17549500902943043.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb. Cortex. 14, 1384–1389. https://doi.org/10.1093/cercor/bhh099.

Wilson, S.M., Galantucci, S., Tartaglia, M.C., Rising, K., Patterson, D.K., Henry, M.L., Ogar, J.M., DeLeon, J., Miller, B.L., Gorno-Tempini, M.L., 2011. Syntactic processing depends on dorsal language tracts. Neuron 72, 397–403. https://doi.org/10.1016/j.neuron.2011.09.014.

Witteman, J., van Ijzendoorn, M.H., van de Velde, D., van Heuven, V.J.J.P., Schiller, N.O., 2011. The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. Neuropsychologia 49, 3722–3738. https://doi.org/10.1016/j.neuropsychologia.2011.09.028.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21, 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14, 1370–1386. https://doi.org/10.1006/nimg.2001.0931.

Xu, Y., 2019. Prosody, tone and intonation. In: Katz, W.F., Assmann, P.F. (Eds.), The Routledge Handbook of Phonetics. Routledge, New York, pp. 314–356. https://doi.org/10.4324/9780429056253.

Zaccarella, E., Meyer, L., Makuuchi, M., Friederici, A.D., 2017. Building by Syntax: the Neural Basis of Minimal Linguistic Structures. Cereb. Cortex. 27, 411–421. https://doi.org/10.1093/cercor/bhv234.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex. 11, 946–953. https://doi.org/10.1093/cercor/11.10.946.

Zatorre, R.J., Gandour, J.T., 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1087–1104. https://doi.org/10.1098/rstb.2007.2161.

Zhang, L., Shu, H., Zhou, F., Wang, X., Li, P., 2010. Common and distinct neural substrates for the perception of speech rhythm and intonation. Hum. Brain Mapp. 31, 1106–1116. https://doi.org/10.1002/hbm.20922.

Zhao, J., Shu, H., Zhang, L., Wang, X., Gong, Q., Li, P., 2008. Cortical competition during language discrimination. Neuroimage 43, 624–633. https://doi.org/10.1016/j.neuroimage.2008.07.025.