



From acoustic segmentation to language processing: evidence from optical imaging

Hellmuth Obrig^{1,2,3*}, Sonja Rossi^{1,3}, Silke Telkemeyer^{3,4,5} and Isabell Wartenburger^{3,4,5}

¹ Max-Planck-Institute for Cognitive and Brain Sciences, Leipzig, Germany

² Clinic for Cognitive Neurology, University Clinic Leipzig, University of Leipzig, Leipzig, Germany

³ Department of Neurology, Berlin NeuroImaging Center, Charité Universitätsmedizin Berlin, Berlin, Germany

⁴ Neurocognition of Language/Neurolinguistics, Department of Linguistics, University of Potsdam, Potsdam, Germany

⁵ Cluster Languages of Emotion, Freie Universität Berlin, Berlin, Germany

Edited by:

David Boas, Massachusetts General Hospital, USA; Massachusetts Institute of Technology, USA; Harvard Medical School, USA

Reviewed by:

Heather Bortfeld, University of Connecticut, USA
Sharon Fox, Massachusetts Institute of Technology, USA

*Correspondence:

Hellmuth Obrig, Max-Planck-Institute for Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. e-mail: obrig@cbs.mpg.de

During language acquisition in infancy and when learning a foreign language, the segmentation of the auditory stream into words and phrases is a complex process. Intuitively, learners use “anchors” to segment the acoustic speech stream into meaningful units like words and phrases. Regularities on a segmental (e.g., phonological) or suprasegmental (e.g., prosodic) level can provide such anchors. Regarding the neuronal processing of these two kinds of linguistic cues a left-hemispheric dominance for segmental and a right-hemispheric bias for suprasegmental information has been reported in adults. Though lateralization is common in a number of higher cognitive functions, its prominence in language may also be a key to understanding the rapid emergence of the language network in infants and the ease at which we master our language in adulthood. One question here is whether the hemispheric lateralization is driven by linguistic input *per se* or whether non-linguistic, especially acoustic factors, “guide” the lateralization process. Methodologically, functional magnetic resonance imaging provides unsurpassed anatomical detail for such an enquiry. However, instrumental noise, experimental constraints and interference with EEG assessment limit its applicability, pointedly in infants and also when investigating the link between auditory and linguistic processing. Optical methods have the potential to fill this gap. Here we review a number of recent studies using optical imaging to investigate hemispheric differences during segmentation and basic auditory feature analysis in language development.

Keywords: optical imaging, infants, language acquisition, acoustic segmentation, NIRS

INTRODUCTION: ACQUISITION OF LANGUAGE COMPETENCE

Language is considered a specific human faculty, which is universal with respect to features such as compositionality, arbitrariness of signs and recursion (Hauser et al., 2002). Beyond such *universal* features, common to all human languages, competence in a *specific* spoken language requires one to recognize phonological and prosodic features, to have knowledge about lexico-semantic representations, and to master specific syntactic rules (Cutler et al., 1983). In addition, languages are embedded into a social context, and their use is shaped by metalinguistic cultural conventions. Thus the acquisition of language implies that both: (i) humans are biologically endowed with the faculty of language and (ii) the acquisition of their specific native language(s) critically depends on social interaction (Bonatti et al., 2002; Kuhl et al., 2003; Kuhl, 2007). An important input for infants in social interaction is speech. Although language competence goes beyond speech perception, infants can use auditory regularities in speech to find anchors enabling the acquisition of language. Investigating the neuronal underpinnings of speech processing may hence delineate a framework in which both, universal – potentially inborn – prerequisites and specifically acquired regularities mutually shape language competence (Dehaene-Lambertz et al., 2006). Studies from a number of disciplines, including linguistics and psychology but also neuroimaging

have highlighted commonalities between the processing of complex auditory stimuli and features of speech (Kushnerenko et al., 2001; Benasich et al., 2006). More specifically, the well established lateralization of language processing has been discussed in the light of a potentially more basic specialization of the respective secondary auditory cortices (Zatorre et al., 2002; Poeppel et al., 2008). In this vein influential theories have been put forward, based on experimental findings which show double dissociations of auditory feature analysis, such as spectral vs. temporal complexity (e.g., Schonwiesner et al., 2005) or slow vs. fast modulations of the auditory input (e.g., Boemio et al., 2005). With regard to the question concerning the contribution of biologically endowed vs. use-dependent shaping of the brain’s language network, a close relation between the processing of complex auditory features and speech perception allows for an integrative view: acquisition of language builds on a neurobiological ability to process complex auditory stimuli with high precision while the specific language is environmentally shaped.

We proceed based on the fact that language and auditory processing are interrelated by speech and focus our review on the development of speech perception. Recent studies provide evidence that auditory feature analysis is already highly developed at birth (DeCasper and Prescott, 2009; Sambeth et al., 2009), but

is substantially modified by the specific language exposure within the first months of life. Additionally, there is ample evidence that the language acquisition process is bidirectional, in that language competence shapes the interpretation of ambiguous auditory signals, as has been shown in cases of conflicting visual–auditory input (Campbell, 2008; Bristow et al., 2009). Building on the research of the neuronal underpinnings of language acquisition, the scope of our contribution is to give a comprehensive account of recent work using optical imaging investigating speech processing in infants. Besides the methodological focus on this non-invasive tool we outline studies in the context of *segmentation* based on acoustic features. Generally segmentation denotes the ability to find structure in spoken language, which represents a more or less continuous auditory stream. Segmentation allows the listener to identify single words but also phrases and syntactic structure. While segmentation is relevant for the ease at which the language-competent listener decodes the auditory stream, it is a mandatory first step for the prelinguistic infant. Segmentation is initially based on specific “anchors” provided by auditory cues in the speech signal.

METHODOLOGICAL CONSIDERATIONS: WHY DO WE FOCUS ON OPTICAL IMAGING?

Traditionally neuro- and psycholinguistic research has been dominated by electrophysiological approaches. This stems from two properties of the EEG (and in part MEG) technique: (i) exquisite temporal resolution in the range of milliseconds and (ii) comparatively low experimental constraints (Friederici, 2004; Huotilainen et al., 2008). The former allows for the monitoring of temporally sequential sub-processes in language comprehension, and is therefore essential to construct models of hierarchical and parallel processing steps. The latter can be considered a prerequisite to extend neuro- and psycholinguistic research to developmental issues (Friederici, 2005). As an example, EEG recordings can provide reliable information on whether, a prelinguistic infant is sensitive to linguistically relevant features and can additionally test hypotheses that predict a temporal succession of steps affording linguistic operations (Benasich et al., 2002; Friederici, 2005; Kooijman et al., 2005; Oberecker et al., 2005; Mannel and Friederici, 2009). There are, however, substantial limitations. For instance, localization of the neuronal activity elicited by linguistic contrasts is difficult when relying on EEG recordings. Some components of the event-related potentials are typically seen over the central electrodes, such as the N400, essentially reflecting lexico-semantic access. This is counterintuitive given the broad evidence for a lateralization of the language network (Lau et al., 2008). Hence when using EEG the assignment of a language-specific component to a cortical area is somewhat arbitrary – in some instances even with respect to lateralization. Often topography serves to differentiate components rather than an anatomical ascription. Also while electrophysiological approaches are very powerful in detecting brief processes in the range of 10–1000 ms, the integration over a longer time frame requires less robust analytical approaches such as DC-EEG or the analysis of oscillatory activity in the recorded signal (Hald et al., 2006). Both a better localization of the activation and the integration over longer time frames are features of the vascular based techniques. Hence the advent of functional magnetic resonance imaging (fMRI) using BOLD-contrast has led to an enormous

boost in neurolinguistic research (Bookheimer, 2002; Gernsbacher and Kaschak, 2003). The technique allows us to investigate exact functional–anatomical relations in the language network and has successfully demonstrated that part of the ambiguity of EEG-based assumptions may resolve when assuming a much wider network to be activated by the language input (Pulvermuller et al., 2009). With respect to auditory presented tasks, however, instrumental noise of the scanner poses a very different challenge. Surprisingly, the comprehension of speech is quite robust. Competent listeners will extract linguistically relevant information even in very noisy environments and cochlear implant wearers extract almost the full linguistic content of spoken language from a very limited number of spectral bands transmitted to the auditory system by the device (Moore and Shannon, 2009). Nonetheless, interfering scanner noise may well be a crucial factor, especially with respect to studies in which the differentiation of subtle acoustic features such as phoneme discrimination is in the focus of interest. Also, concerning the emergence of language competence, fMRI approaches in infants are limited to a rather small number of research labs, in part due to ethical considerations (Dehaene-Lambertz et al., 2002). It therefore comes by no surprise that alternative methodologies such as non-invasive optical imaging have been used in a number of recent studies investigating language acquisition during infancy (Pena et al., 2003; Bortfeld et al., 2009; Hull et al., 2009).

Optical imaging, relying on the principles of near-infrared spectroscopy (NIRS), requires a comparatively simple set-up, which allows for measurements also in preterm infants, neonates, and infants (Aslin and Mehler, 2005; Lloyd-Fox et al., 2010). For an example see **Figure 2E**. The methodology is completely silent, permitting a near natural presentation of auditory stimuli by the lack of instrumental noise. Additionally, the simultaneous acquisition of EEG without any interference is of specific importance to the field (Koch et al., 2006, 2009; Telkemeyer et al., 2009), since many domains of language processing have been reliably mirrored in differential EEG components.

The assessment of neuronal activity by optical imaging is based on the vascular response much like PET and fMRI (Villringer and Dirnagl, 1995). Compared to fMRI it has a coarse spatial resolution in the range of centimeters (but see Zeff et al., 2007). This is comparable to the estimated lateral resolution of EEG. However, as opposed to EEG, the attribution of the signal to the underlying cortical area is reliable. Spatial depths resolution of optical imaging is limited by physical principles of light propagation in highly scattering media (Obrig and Villringer, 2003). Even novel approaches to enhance topographical resolution are restricted to the cortical surface and provide no information on subcortical signal changes or activations in mesial cortical areas, which are distant from the brain's surface. Optical imaging systems to supply a rough depth resolution using multiple distance measurements or time resolved spectroscopy have been used in infants (Hebden and Austin, 2007) and adults (Liebert et al., 2004, 2005), but have not been applied in cognitive tasks. Optimization of the technology is mandatory to allow for an even broader application in the field (Zeff et al., 2007; White et al., 2009).

For the application in language research, most groups have used monitors, which can detect changes in the concentrations of oxygenated and deoxygenated hemoglobin (ΔHbO , ΔHbR). The sum of both

changes results in changes in total hemoglobin concentration in the sampled volume ($\Delta HbT = \Delta HbO + \Delta HbR$). Due to their differential absorption spectra in the near-infrared spectrum (600–950 nm) the two hemoglobins can be differentiated when light attenuation is measured at two or more wavelengths. This can be perceived as an extension of the well-known fact that fully oxygenated (arterial), and partially deoxygenated (venous) blood differ in color. By solving a simple equation-system the changes in attenuation at the different wavelengths are transformed into hemoglobin concentration changes based on the Beer-Lambert approach (Cope and Delpy, 1988). Methodologically this is a rather conservative approach (see **Figure 1**).

The translation of the measured quantities into a physiologically meaningful signal needs to be mentioned (see also Lloyd-Fox et al., 2010). In brief, the vascular imaging community has adopted the term “activation” for an increase in regional cerebral blood flow (rCBF). The underlying assumption is that an increase in neuronal activity generates an increased metabolic demand, which is met by the increase in blood flow. It is important to note that this tight neurovascular coupling has been demonstrated in a large number of studies using various techniques both invasively in animals and also non-invasively in adult humans (Lauritzen and Gold, 2003; Iadecola, 2004; Logothetis and Wandell, 2004). For optical imaging, the increase in blood flow is reflected by an increase in oxygenated and total hemoglobin ($HbO\uparrow$, $HbT\uparrow$), the latter corresponding to blood volume and a decrease in deoxygenated hemoglobin ($HbR\downarrow$), resulting from an increase in blood flow velocity. The decrease in HbR is less intuitive, since the consumption of oxygen in an activated area intuitively would yield an increase in HbR. However,

it has been demonstrated that in most instances the increase in blood flow overcompensates the oxygen demand (Fox and Raichle, 1986). A focal hyperoxygenation, including a decrease in HbR, is therefore expected in an “activated” area.

There is an on-going discussion regarding whether HbO, HbT, or HbR changes will better reflect “activation” in an optical imaging study. This discussion becomes even more heated when it comes to interpreting the findings in infants and neonates, since a number of reports have suggested that in early infancy the vascular response may follow substantially different rules due to a different metabolic rate of oxygen in non-myelinated brain tissue (Kusaka et al., 2004; Marcar et al., 2004; but Colonnese et al., 2008). Additionally, the fact that many studies have shown increases in HbO without corresponding decreases in HbR has been used to advocate the methodology’s sensitivity to activations which would be “silent” in an fMRI-study (but see **Figures 2A–C**). The comparative study simultaneously assessing optical imaging and BOLD-contrast in infants of different ages is missing. Hence, it seems appropriate to report both compounds irrespective of whether both or only one compound yields the hypothesized effect. In our review we consider studies reporting increases in HbO and HbT and/or decreases in HbR as an indicator of an activation in the underlying cerebral tissue (see **Figure 1**). We mark the parameter on which statistics were performed in brackets. When linking results of optical imaging studies to the much larger body of vascular imaging based on BOLD-contrast fMRI, it should be acknowledged that an area in which HbR does not decrease does not easily correspond to an area showing an activation (increase

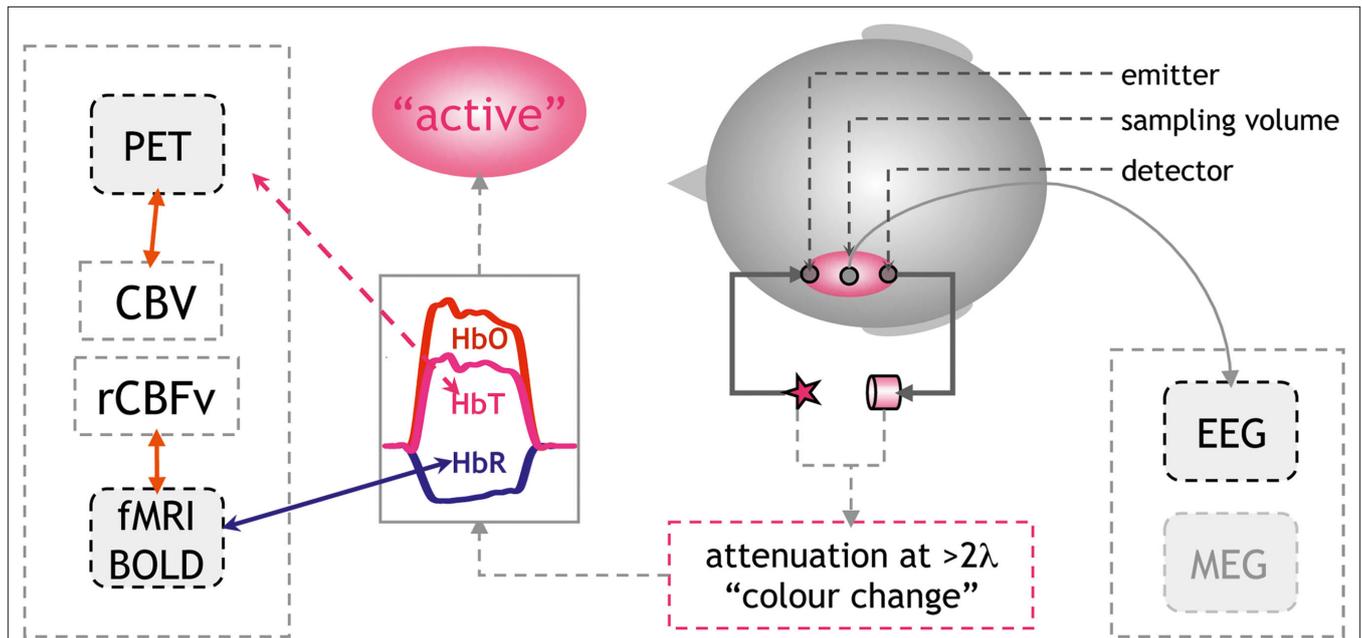
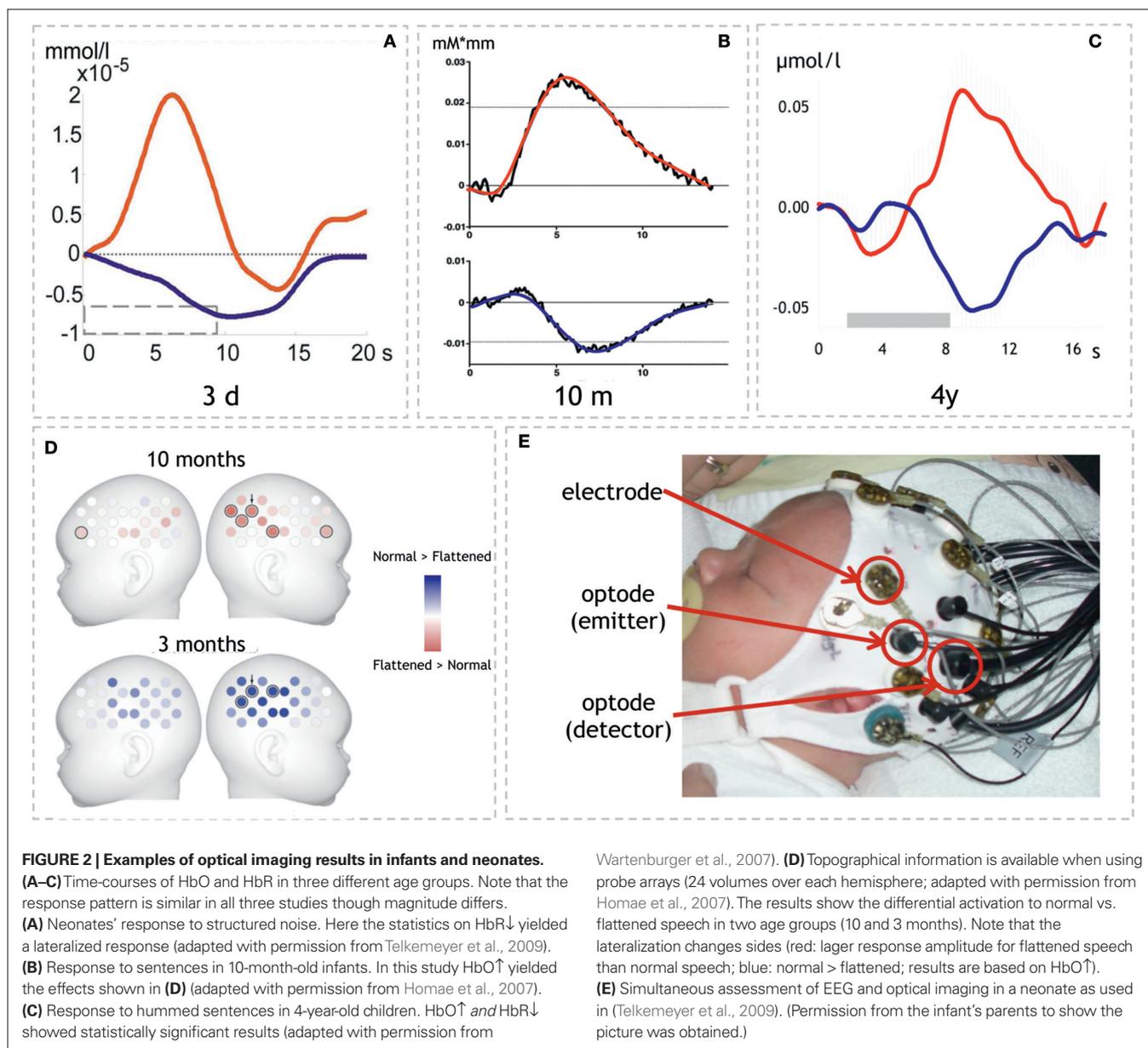


FIGURE 1 | Principles of non-invasive optical imaging: Optical probes are fixed to the head defining the sampling volume (pink ellipse). Measuring changes in attenuation at two or more wavelengths allows to calculate changes in the concentration of HbO, HbR. An increase in HbO and/or a decrease in HbR indicate cerebral activation. Physiologically this pattern results from an increase in regional cerebral blood flow (rCBF) consisting of an increase in blood flow velocity (rCBFv) and volume (rCBV). The increase in rCBF overcompensates the

demand in oxygen (Fox and Raichle, 1986). An increased “washout” of HbR results from increased rCBFv. This decrease in HbR is the major source of fMRI BOLD-contrast increases (Steinbrink et al., 2006). The change in total hemoglobin concentration ($HbT = HbR + HbO$) correlates with blood volume (CBV), which can be assessed by positron emission tomography (PET). Notably the assessment of electrophysiological markers of activation by EEG or MEG (right side of graph) is undemanding as demonstrated in a neonate in **Figure 2E**.



in BOLD-contrast; Kleinschmidt et al., 1996; Steinbrink et al., 2006)¹. If HbR increases and HbO decreases this would in most

¹BOLD-contrast stems from two major effects. The larger effect is the extra- to intravascular field gradient largely caused by concentration changes of paramagnetic HbR in the vessel. The second effect is due micro-inhomogeneities in the vessel itself. The latter depends on the ratio of HbR and HbO in the vessel, or saturation. Thus, theoretically, an isolated increase in blood volume could yield an increase in BOLD-contrast (i.e., inflow of HbO into the voxel while concentration of HbR remains constant). Such scenarios cannot be easily explained by any of the proposed models of rCBF changes in an activated area. These models describe the complex interplay between blood flow velocity, blood volume, oxygenation, and the BOLD-contrast (Mandeville et al., 1999; Buxton et al., 2004; Stephan et al., 2004). An increase in HbR could only elicit a positive BOLD-contrast in scenarios of an extreme increase in blood volume, which is even less compatible with any of these models. In most cases an increase in deoxy-Hb will correspond to a decrease in BOLD, signaling "deactivation" (Wenzel et al., 2000). For a detailed discussion of models to explain the interdependence of BOLD, blood flow, and the local hemoglobin concentrations please refer to Steinbrink et al. (2006).

instances yield a decrease in a BOLD-contrast study and would thus be termed "deactivated" or "negative BOLD-contrast" (Wenzel et al., 2000). Increases in both HbO and HbR in response to a stimulus will require a new model for neurovascular coupling. In other words, such a scenario means that a blood-volume based methodology will term the area "activated", while a corresponding BOLD-contrast fMRI study may yield a "deactivation" in the very same area.

The complexity and potential ambiguity of the different response modalities including the vascular response, the electrophysiological response and even the behavioral measure of looking times calls for multimodal approaches. Such approaches may broaden our knowledge on the relation between neuronal, vascular, and behavioral response parameters and will definitely strengthen our confidence in the response seen in either of the methodologies.

SEGMENTATION AND ITS RELEVANCE FOR LANGUAGE ACQUISITION

Apart from the methodological focus, why do we focus our review of studies on language acquisition in the context of acoustic segmentation? When we listen to speech we are confronted with a *continuous* auditory stream. On the contrary, all languages follow *hierarchical* principles, which are conventionalized in each language by specific rules. To decode the meaning of speech, it is mandatory to identify single words, phrases, and sentences. During language acquisition this process is initially guided by acoustic cues aiding the segmentation into smaller units. We here differentiate between two levels of such acoustic cues. These two levels differ temporally with respect to the length on which the auditory feature acts (see **Figure 3**). Linguistically, they largely convey cues relevant either for suprasegmental or segmental information (see **Figure 5**).

- (i) Sentential prosody – the melodic pitch and durational variation at the sentential level – supplies acoustic cues to chunk the auditory input into sentences and phrases. Since prosodic features act across a number of segments, the linguistic information conveyed is also termed “suprasegmental.” Such prosodic grouping of lexical items in a sentence is relevant also for the language-competent adult. However, language-competent listeners will also strongly rely on syntactic rules². On the contrary, during language acquisition in infancy, prosody plays a critical role for segmentation. It also aids acquisition of syntax as highlighted by the “prosodic bootstrapping account”

- (ii) While acoustic cues supplied by prosody segment the speech stream into larger units, the differentiation of single words is also aided by acoustic cues acting on a much shorter time scale. These cues can be based on the specific phonemic inventory of a given language, since it provides constraints of how segments within the auditory stream are ordered to constitute single lexico-semantic entries (i.e., words). Phonemes are the smallest units that distinguish meaning (e.g., /pit/ vs. /bit/ differ only in the initial distinctive phoneme). The linguistic information conveyed by such language-specific regularities is also termed “segmental.” As is illustrated in **Figure 3** the duration of such cues is much shorter than those provided by sentential prosody and can reach down to ~10 ms (see **Figure 3**). Studies that deal with such sublexical cues and their relevance for speech processing are discussed in the paragraph “Optical Imaging Studies on the Neural Correlates of Segmental Processing³.”

²As an example the meaning of the sentence “The teacher said the pupil is stupid” depends on how the lexical items are syntactically related (“The teacher said // the pupil is stupid.” vs. “The teacher// said the pupil //is stupid.”). Typically the ambiguity of this sentence will not occur in spoken language, because sentential prosody will clearly mark boundaries (“//” in the above example). The hierarchical structure of the sentence can also be unambiguously coded by means of syntax (e.g., “The teacher said, *that* the pupil is stupid.”).

³It should be noted that any acoustic cue that relies on the relation between several phonemes could be considered “suprasegmental.” Thereby strictly speaking durational contrasts, pitch accents, and tonal cues could be considered suprasegmental. However, in some languages the relative length of a vowel defines a phonemic boundary, providing segmental information in a specific spoken language (e.g., Minagawa-Kawai et al., 2002). Thus, the terminology may be somewhat misleading in these instances. We discuss here studies on such cues in the context of segmental information as opposed to the studies on sentential prosody, which unambiguously act on a suprasegmental level.

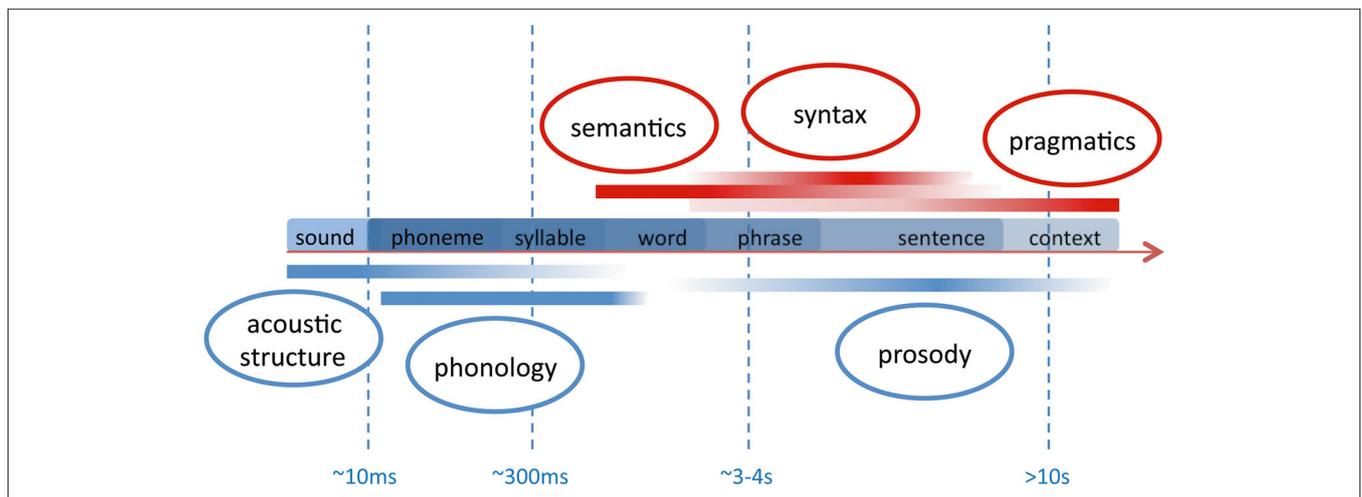


FIGURE 3 | Sketch of temporal frames relevant for speech perception. The smallest *linguistically* relevant segmentation deals with the differentiation of phonemes which rely on transitional differences in the range of 10 ms. Syllables and words form the nucleus of a lexico-semantic decoding of speech. Syntax and prosody act on the level of phrases and sentences which may have a duration of several seconds. The speaker’s intention (pragmatics)

evolves over largely different temporal frames ranging from single vocalizations to tales and stories. Note that the components below the time line (blue) are available also to the prelinguistic infant due to their acoustic prominence. The other components (red) are fully available only when a certain language competence is reached. In the language-competent adult, all levels strongly interact.

Having sketched some principal issues of segmentation, we next address the question why segmentation issues are essential for language acquisition in infancy. When newborn infants are confronted with the continuous auditory stream of their environment, they must detect relevant lexical and syntactic units without prior knowledge. Because infants have no or only poorly developed lexical representations and have not yet acquired the syntactic rules, they have to rely on the acoustic structure of the signal in a bottom-up fashion. Infants use auditory markers resulting from phonological regularities and prosodic cues to segment the input (Gervain and Mehler, 2010), because in natural speech the acoustic signal does not contain reliable pauses to indicate word boundaries (Cutler, 1994). The prosodic bootstrapping account assumes that the infant initially recognizes the prosodically marked units of the input to learn the syntactic structure of the language (Gleitman and Wanner, 1982). Thus, prelinguistic infants extract acoustically detectable “anchors” allowing them to segment the input into smaller entities to which meaning will be assigned at a later stage of their development. A number of behavioral findings support an eminent role of auditory feature sensitivity by demonstrating that auditory discrimination starts even before the infant is born. It has been shown that prosodic features and specific speech frequencies are perceived by the fetus (Draganova et al., 2007). This may constitute the fact that newborns can discriminate between the speech of their mother to speech belonging to different rhythmical classes (Mehler et al., 1988). Additionally, they prefer their mother’s voice over other female voices (Mehler et al., 1978; DeCasper and Fifer, 1980).

In summary, segmentation is a crucial task to be tackled by the infant even during the early stage of language acquisition. Clearly, for auditory language comprehension adults also have to segment the speech stream. However, while adults can also use the stored lexical and semantic information as well as their pragmatic and contextual knowledge to improve the recognition processes in a top-down fashion, there is converging evidence that infants initially rely on acoustic cues to “tune in” to the language and identify segmental and suprasegmental features of the language, to which they are exposed.

OPTICAL IMAGING STUDIES ON THE NEURAL CORRELATES OF SUPRASEGMENTAL PROCESSING

A number of optical imaging studies have addressed the cerebral oxygenation response to stimuli which differ in suprasegmental properties. Most studies use materials in which prosodic features are varied on a sentential level. On the sentential level prosody segments the speech stream into phrases (Cutler et al., 1997). The prosodic phrase boundary is characterized by three cues: a pitch rise, a pre-boundary lengthening, and a pause. The perception of such a boundary elicits a specific event-related potential in adults (Steinhauer et al., 1999) even if there is no pause between two phrases. This has been discussed with respect to the hypothesis that prosodic properties of a sentence have a primacy for decoding the syntactic structure of a spoken sentence. As pointed out above, the detection of such prosodic cues is especially important in infant language acquisition. We here review optical imaging studies in infants that have addressed two essential questions: (i) whether and at what age prosodic information is processed by an infant and (ii) whether this processing is lateralized.

A recent study addresses the first question in neonates. In their study, Saito et al. (2007b) used two versions of a synthesized reading of a fairy tale. One version included pitch variation at the sentential level while the other did not. The pitch-modulated version elicited a larger activation ($\text{HbO}\uparrow$) over frontal areas, however the expected lateralization to the right hemisphere was not supported by the data. The study is remarkable in showing a very early differentiation between two auditory streams, different only with respect to pitch modulation, during the first days of life. However, the fact that only one (frontal) channel was recorded over both hemispheres and rapid habituation to the modulated speech was seen within 10 s of the stimulation block, may indicate that neonates are sensitive to the different acoustic features rather than to their potential linguistic significance. Interestingly, the same group reported differences between the responses to the mother’s and the nurse’s voice in premature neonates (Saito et al., 2009). They found an activation ($\text{HbO}\uparrow$) over the left frontal channel in response to both voices. However, only the nurse’s voice also elicited an increase over the right frontal area interrogated by the optical probe. Though limited by similar methodological aspects as the earlier study, this difference is interpreted based on the assumption of a stronger familiarization to the nurse’s voice, who can be considered the primary care-giver in the highly isolated premature neonate. A stronger activation ($\text{HbO}\uparrow$ bilateral frontal) for infant- vs. adult-directed speech was also seen in a study in healthy full term neonates (Saito et al., 2007a).

The data provides evidence that pitch modulation is detected already from birth. However, the weak lateralization and the very coarse ascription to the frontal lobe raise the question of the relevance to specifically language development. Using a much larger probe array the lateralization of prosodic processing was investigated by Homae et al. (2006). They acoustically presented normal sentences and flattened sentences to 3-month-old infants. For the flattened sentences the pitch contour was digitally replaced by the mean value of the pitch in the normal sentences. Thus the two stimuli differed with respect to the presence of prosodic information. However, they also differed with respect to the “naturalness” of the material as will be discussed below. The results of this optical imaging study using an array of 24 channels over each hemisphere confirm that prosodic information is specifically processed by the infant’s brain (Figure 2D). While there was a widespread activation ($\text{HbO}\uparrow$ and $\text{HbR}\downarrow$) for both conditions, a direct comparison of normal vs. flattened speech yielded a stronger activation ($\text{HbO}\uparrow$) over right temporo-parietal areas. The data thus indicate that a rightward lateralization of suprasegmental, prosodic processing is present already at the age of 3 months. In a follow-up study the same group examined the identical paradigm in 10-month-old infants (Homae et al., 2007). Again, activations over bilateral temporo-parietal areas were seen in response to both stimulus modalities. A direct comparison between flattened vs. normal speech, however, yielded a stronger activation ($\text{HbO}\uparrow$) over the right hemisphere for the flattened material. The authors interpret this – at first glance contradicting – finding by postulating that the processing in 10-month-olds might not only reflect pitch processing *per se* as in 3-month-olds, but that additional processes might play a role here (Figure 2D). This is supported by the additional increase in frontal activation in response to the flattened material

in the 10-month-old infants. In other words, 10-month-old infants are assumed to recognize the linguistic input also for the flattened condition, thus no longer solely relying on prosodic cues. The larger activation to the unnaturally aprosodic linguistic input indicates the sensitivity to deviations from normal speech.

A different paradigm was chosen in our optical imaging study investigating lateralization effects in 4-year-old children (Wartenburger et al., 2007). Normally spoken sentences were contrasted with hummed sentences containing solely the prosodic information. Importantly, the hummed condition was naturally performed by the same actor as the normal speech condition. Thus the unnaturalness of digitally modified material was avoided. The results showed a stronger right-hemispheric activation (HbO \uparrow and HbR \downarrow , see **Figure 2C**) for the hummed in contrast to normal sentences, whereas a stronger left-hemispheric processing was present for normal compared to hummed sentences. Thus, normal sentences containing the full linguistic information and fully accessible to these language-competent children, elicit a lateralization to the language-dominant left hemisphere. In contrast, when all segmental features are omitted, resulting in sentences containing pure suprasegmental prosodic information, the activation is lateralized to right-hemispheric regions. The study explored a third stimulus: a digitally flattened version of the hummed sentences. The naturally hummed material elicited stronger right-hemispheric activations than the flattened version. The finding is in line with the account that no prosodic processing can be expected in the flattened material. Using a different kind of flattened material, Homae et al. (2007) found a stronger rightward lateralization in 10-month-old infants (see above and **Figure 2D**). Notably, while Homae et al. (2006, 2007) used flattened normal sentences whose meaning could be comprehended – at least by adults – in our material the flattened stimuli did not contain segmental features (hummed stimuli were flattened, thus only rhythmic features were preserved). We suggest that in our study the older, language-competent children discarded the flattened material without segmental content as non-linguistic material, thus, not eliciting a sufficient activation of the language network. A rightward lateralization of a prosodic contrast (/itta/ vs. /itta?/) was also reported by Sato et al. (2003) as quoted in Minagawa-Kawai et al. (2008) using a habituation/dishabituation paradigm in different age groups (7 months to 5 years). Interestingly the lateralization was only seen in infants older than 11–12 months of age. In infants aged 7–10 months no lateralization was present. This is in contrast to the report on a much earlier sensitivity to prosodic pitch modulation in 3-month-old infants (Homae et al., 2006), most likely due to the difference in stimulus material using pitch alteration on a syllabic vs. sentential level respectively.

In summary, the optical imaging studies have demonstrated that suprasegmental information at a sentential level is processed by infants starting at a very early age. The different studies also suggest that the processing is lateralized. Variable results concerning the role of the right hemisphere have been interpreted to indicate either a stronger response to an unexpected lack in prosodic information or by the actual processing of the suprasegmental information. Similar diverging lateralizations have been reported in adults in BOLD-contrast fMRI studies using flattened and purely prosodic material (Meyer et al., 2002, 2004). To further clarify the issue it will be relevant to perform a longitudinal study, based on identical material across all

age groups. Yielding conflicting results for digitally flattened material, the studies also show that a compromise between control of acoustic/linguistic features and “naturalness” of the material may be especially difficult in infants. With respect to methodological issues, it should be noted that vascular techniques like optical imaging are advantageous when investigating the neural correlates of suprasegmental processing on a sentential level. Ideally the assessment of the more slowly evolving features at the sentential level can be complemented by the electrophysiological markers at the phrase boundaries (e.g., closure positive shift, CPS; Mannel and Friederici, 2009).

It should be noted, that suprasegmental cues also act within single words. Indeed one prominent feature of prosody is the stress pattern that characterizes the intonation of syllables within words (Jusczyk, 1999). Additionally prosodic content, like pitch variation, is relevant in tonal languages (Gandour et al., 2003) and serves lexico-semantic differentiation. Since lexico-semantic tasks predominantly activate the left-hemispheric language areas, such suprasegmental contrasts may lateralize to the left hemisphere. An optical imaging study addressing pitch accent in Japanese infants (Sato et al., 2009) highlights this issue and is discussed below in the framework of different models put forward to explain lateralization of speech processing (paragraph How is Segmentation Guided by Auditory Analysis).

OPTICAL IMAGING STUDIES ON THE NEURAL CORRELATES OF SEGMENTAL PROCESSING

On a smaller temporal scale, phonemes and their combinatorial rules define the inventory from which syllables and words are built. Phonemes and their respective combinatorial rules are also specific to a given language. Hence, they also constitute regularities that allow for segmentation based on the auditory features. With respect to the processing of phonological information, the principle has long been established that at birth infants are sensitive to virtually every phonemic contrast in any existing language (Streeter, 1976; Werker and Tees, 1999; Kuhl and Rivera-Gaxiola, 2008). When infants acquire their native language(s) they lose this ability for contrasts that are not relevant in the given native language(s) (Naatanen et al., 1997; Cheour et al., 1998; Winkler et al., 1999). While this may be a hurdle for foreign language acquisition in adult life, the adaptation to the environmental needs can be considered an essential step to establish the high efficiency by which we decode speech signals in our own language(s). Some optical studies reviewed below have addressed this very issue.

PROCESSING OF PHONOTACTIC CUES

Specific languages allow specific phoneme-combinations. The combinatorial rules of different phonemes in a given language are called phonotactic rules. They are not only specific for a language but also with respect to the on- and offset of a word. For example, /fl/ is a possible combination at the onset of an English or German word (such as /flight/; German: /Flug/), whereas /tl/ is not. Also /ft/ is not present in English and German for the onset of a word, while it occurs at the coda-position (English: /loft/; German: /Luft/ (air)). Evidently such phonotactic regularities are important for the segmentation of speech into smaller units such as words.

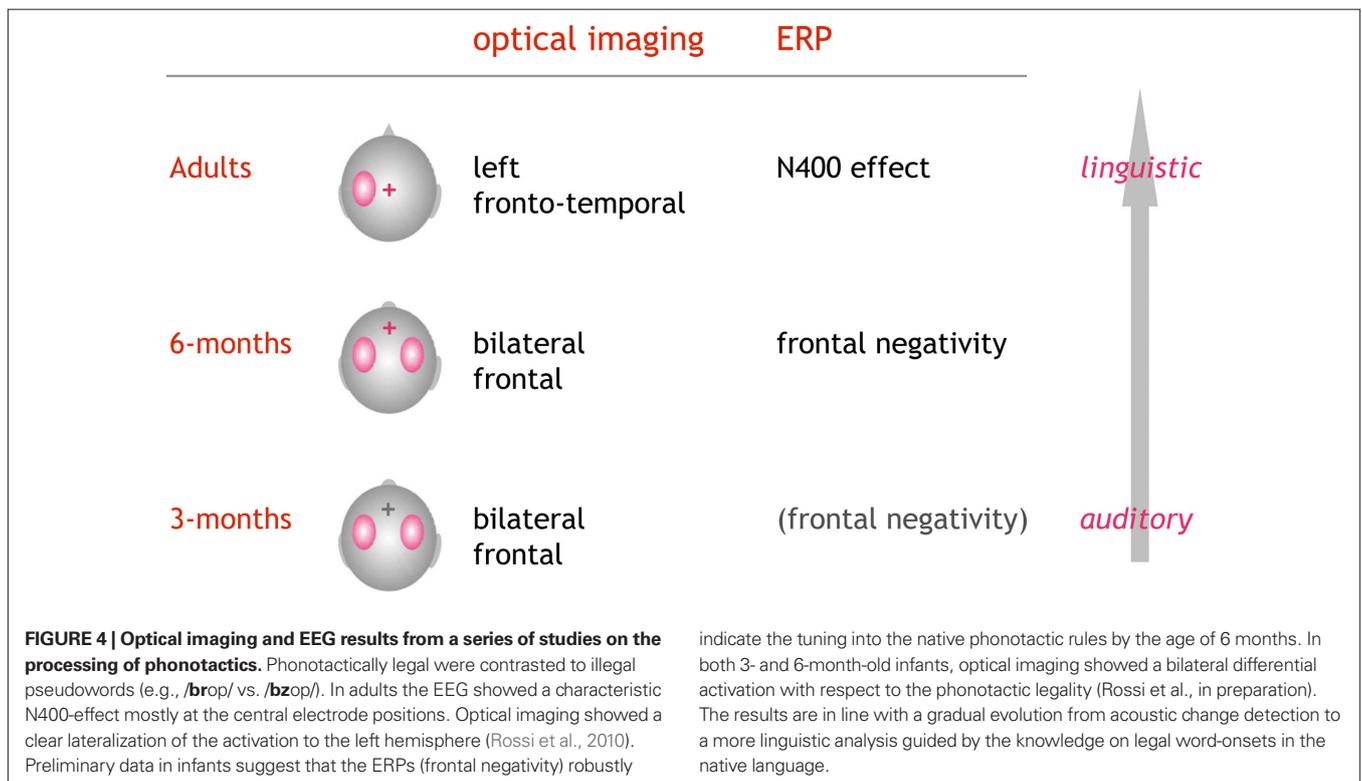
In a series of optical imaging studies, we investigated how phonotactic rules are processed. To this end, pseudowords, legal or illegal with respect to the native phonotactic rules, were acoustically

presented to 3- and 6-month-old infants and adults. In adults we find a stronger activation (HbR↓) for the legal compared to the illegal pseudowords with a clear lateralization to the left hemisphere (Rossi et al., 2010). Preliminary results in the infant studies suggest a differential activation pattern for the two kinds of phonotactic rules at both 3 and 6 months of age, which can be interpreted as a preference for native phonotactics. However, in these age groups the observed effect did not show a clear lateralization. Since we simultaneously recorded EEG in all three studies, we also analyzed the respective electrophysiological markers. In the adult group a larger N400 was seen in response to the legal pseudowords. This further supports the hypothesis that these phonotactically legal pseudowords qualify as potential candidates for lexico-semantic analysis, while illegal items are “discarded” at an earlier, prelexical stage. In infants a similar effect for a frontal negativity was seen at 6 months, while at 3 months this ERP-effect was not robust. Neither in infants nor in adults were the ERP-components lateralized. In summary this experimental series suggests that neural correlates of phonotactic processing show an evolution beginning in early infancy. Conceptually the ERPs confirm the differentiation between native and non-native rules at the age of 6 months. Beyond the fact that a differential processing was already seen at the age of 3 months in the optical imaging data, the methodology also demonstrates the evolving lateralization of the processing from a bilateral auditory to a left-lateralized more linguistic analysis of the material (see **Figure 4**).

SENSITIVITY TO PHONEMIC BOUNDARIES

A similar developmental change has been reported in a sequence of optical imaging studies on the discrimination of phonemes specific to Japanese (Minagawa-Kawai et al., 2002, 2004, 2005, 2007). In

Japanese the relative duration of a vowel can determine the meaning of a word. Vowels of different lengths with respect to a prior vowel constitute different phonemes. A phonemic boundary is the discrete transition between two phonemes. In the example used by Minagawa-Kawai et al. (2005), the phonemic boundary for the second vowel in the pseudoword /ma-ma/ lies between 184 and 217 ms when the first vowel is kept constant at 110 ms. In adult native Japanese speakers, the durational vowel contrast showed a stronger response (HbO↑) over the left hemisphere only for across-boundary presentations, supporting the notion that the durational contrast is processed as linguistically relevant (Minagawa-Kawai et al., 2002). Interestingly non-native high-proficiency adult speakers did not show this lateralization (Minagawa-Kawai et al., 2004, 2005). Most notable for our focus on developmental issues the material was used in infants of five different age groups (Minagawa-Kawai et al., 2007). In the youngest age group (3–4 months) the authors found identical response amplitudes for across-boundary and within-category comparisons. However, at 6–7 months stronger responses were present for across-category comparisons. Interestingly, this effect disappeared at 10–11 months to be stable for both age groups older than 12 months. Beyond the evidence for the phoneme differentiation, the older infants (>12 months) also showed a left-hemispheric dominance of the effect. The authors interpret these findings as a developmental change from a more auditory mechanism for the discrimination of the different vowel lengths at 6–7 months to a more linguistic processing mechanism after 12 months. A similar evolution of leftward lateralization at ~12 months is reported in Sato et al. (2003) as quoted in Minagawa-Kawai et al. (2008). Beyond the fact that this work is an impressive demonstration of how optical imaging can be applied in longitudinal developmental studies in

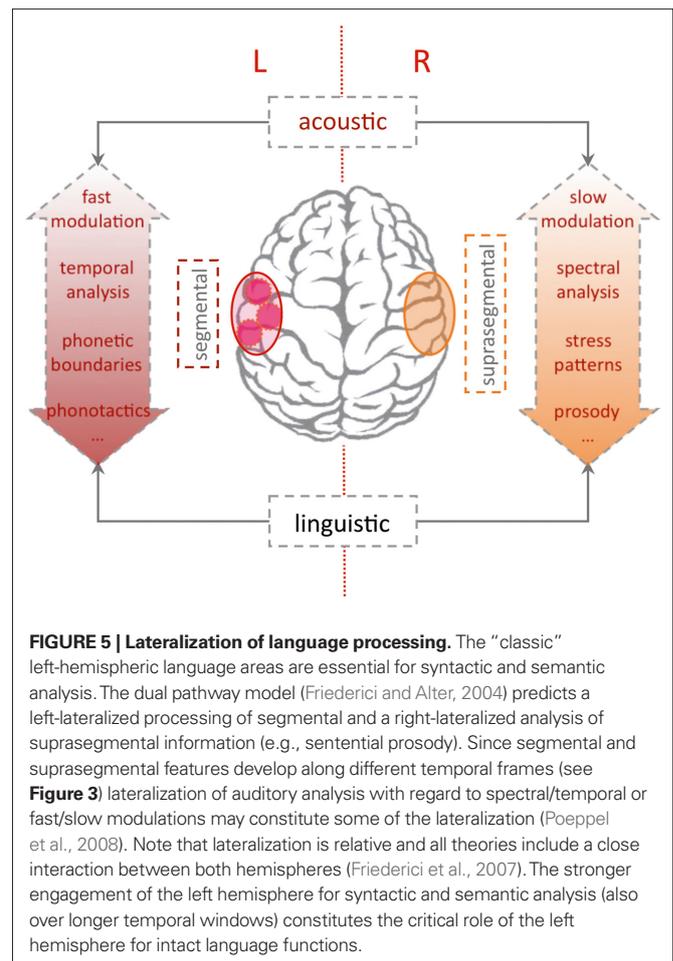


infants, the elegant design needs mentioning. Using a habituation/dishabituation paradigm, the subjects were first exposed to a period in which no contrast was presented. The actual test phase included a mixture of two durational contrasts, either across or within the phoneme category. Such a design is well suited to identify the “blocked” vascular response supplying spatial information and an integration over all neuronal processing steps involved. By using a combination with an electrophysiological measure of change detection (e.g., mismatch negativity), the response to the individual exemplar of the stimuli could be simultaneously assessed.

HOW IS SEGMENTATION GUIDED BY AUDITORY ANALYSIS?

So far, we can conclude that optical imaging studies have confirmed that segmentation cues at both a supra- and a segmental level lead to specific brain responses at an early age. Additionally an asymmetry for the processing has been consistently shown to evolve during infancy. How is this related to the more basic principles on auditory analysis? Based on morphological analyses, it has been long established that the planum temporale is larger on the left when compared to the right hemisphere. This has been discussed as a potential reason for the left auditory cortex to have a better capacity to process speech. Thus, when assuming an auditory bottom-up contribution to the evolution of language competence, asymmetry of the auditory core areas may indicate a structural-anatomical disposition partially leading to the strong lateralization of language (for a review of planum temporale asymmetry see for instance Shapleske et al., 1999). Converging evidence from numerous fields including aphasiology and basic research on the auditory system, lead to three partially overlapping theories (see Figure 5). Exploring the commonalities and differences between music and language, Zatorre et al. (2002) proceed from a very basic acoustic principle: to precisely temporally analyze rapid modulations within an acoustic object, the sampling must be coarse with respect to the spectral analysis. Conversely precise spectral information of an auditory object necessitates integration over a longer temporal time frame (Zatorre et al., 2002). The differential relevance of precise temporal resolution for language and the necessity to perceive subtle pitch (spectral) differentiation in music, explains the respective lateralization for these otherwise similarly structured and complex auditory inputs. However, the right-hemispheric mastery of spectral analysis and a left-hemispheric specialization for fine temporal decoding is also consistent with the differential lateralization of transitional phonemic contrasts (/bin/ vs. /pin/) vs. the much slower prosodic cues.

The dual pathway model proceeds from a more linguistic perspective (Friederici and Alter, 2004). Assuming parallel pathways for left-lateralized-segmental and more right-lateralized-suprasegmental processing, the model predicts a relevant interaction. With respect to suprasegmental information, the processing will elicit a more rightward activation when presented in isolation. When suprasegmental cues gain more linguistic salience, activation will shift to the left hemisphere. This specification introduces a primacy of the left hemisphere for linguistically guided analysis of the auditory stream. The prediction also accommodates findings in a recent optical imaging study, which investigates pitch accent in Japanese in a habituation design similar to the above studies on the durational contrast (Sato et al., 2009). Investigating 4- and 10-month-old



infants, the study reports that both age groups behaviorally differentiated the lexical pitch accent. With respect to lateralization, the lexical pitch accent elicited bilateral activation ($HbO\uparrow$) at 4 months. On the contrary at 10 months, the infants showed a left-lateralized response ($HbO\uparrow$) to the lexical pitch contrast, as was described in an earlier study in adults (Sato et al., 2007). Pure tones exhibiting identical pitch variations as the linguistic material, showed a weaker bilateral activation in all age groups. The studies support the model’s prediction that with increasing linguistic salience the suprasegmental pitch contrast shifts the activation to the left hemisphere.

A third model assumes an intermediate position. The “multi-time-resolution” hypothesis (Poeppel et al., 2008) postulates at least two temporal integration windows relevant for the processing of speech input. A bilateral activation of auditory cortices is predicted for an integration at a relatively fast rate (20–30 ms window), which is appropriate to decode segment-level information. Slower rate sampling (150–300 ms), which is more relevant for suprasegmental feature analysis, is predicted to elicit a more right-hemispheric activation. The prediction was confirmed by an fMRI study in adults using noise segments that were modulated at different rates according to the predicted windows (Boemio et al., 2005). However in adults, a lateralization may also stem from the primacy of language with respect to

auditory analysis. A recent optical imaging study by our group used the same stimuli in newborns in order to explore whether this lateralization based on the acoustic properties is already present at birth (Telkemeyer et al., 2009). Four different temporal modulations were presented. The four conditions contained segments varying every 12, 25, 160, and 300 ms; the first two kinds of stimuli belong to the fast varying category, whereas the last two stimuli are more slowly modulated. The optical imaging results for 2–6 days old newborns revealed a bilateral activation for the 25 ms fast-modulated stimulus and a right-hemispheric dominance for both slow modulation stimuli (HbR↓). These findings suggest that differences in the acoustic and not the linguistic properties of the auditory input may drive lateralization of speech processing during the first days of life.

BEYOND AUDITORY FEATURES

We have highlighted that analysis of the auditory features in speech is a prerequisite for language acquisition in infancy, since lexico-semantic and syntactic knowledge are not yet developed. Clearly, auditory analysis may segment the speech stream, but it does not allow for the extraction of regularities unless more general cognitive abilities are developed in parallel. This has been the focus of extensive research on artificial grammars, since the demonstration of statistical learning abilities in 8-month-old infants (Saffran et al., 1996). An optical imaging study has recently shown that a precursor of such regularity extraction may be present at birth (Gervain et al., 2008). The authors exposed neonates to different artificial grammars governing the structure of syllable triplets. Adjacent (ABB: e.g., “mubaba”) and non-adjacent (ABA: e.g., “bamuba”) repetitions were contrasted to non-structured triplets (ABC: e.g., “mubage”). The authors find evidence for a discrimination of ABB, which they interpret as an indicator of a “repetition detector,” which may form a nucleus for further structural analysis. Their conclusion rests on the fact that a stronger activation (HbO↑) was seen in frontal and temporal areas for the ABB when compared to ABC structures. Interestingly the activation over left frontal channels showed an increase over the exposure period of ~20 min. Thus, it is evident that very early on, structural, initially “Gestalt-like” features of the auditory stream may aid segmentation. In adults a recent optical imaging study demonstrated that non-linguistic tone sequences which could be segmented due to transitional probabilities, yield a stronger left inferior frontal activation (HbO↑)

REFERENCES

- Abla, D., and Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia* 46, 2787–2795.
- Aslin, R. N., and Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *J. Biomed. Opt.* 10, 11009.
- Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., and Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia* 44, 396–411.
- Benasich, A. A., Thomas, J. J., Choudhury, N., and Leppanen, P. H. (2002). The importance of rapid auditory processing abilities to early language development: evidence from converging methodologies. *Dev. Psychobiol.* 40, 278–292.
- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395.
- Bonatti, L., Frot, E., Zangl, R., and Mehler, J. (2002). The human first hypothesis: identification of conspecifics and individuation of objects in the young infant. *Cogn. Psychol.* 44, 388–426.
- Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.
- Bortfeld, H., Fava, E., and Boas, D. A. (2009). Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. *Dev. Neuropsychol.* 34, 52–65.
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychol. Sci.* 16, 298–304.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., and Mangin, J. F. (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *J. Cogn. Neurosci.* 21, 905–921.
- Buxton, R. B., Uluda, K., Dubowitz, D. J., and Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage* 23(Suppl. 1):S220–S233.
- Campbell, R. (2008). The processing of audio-visual speech: empirical and

when compared to random sequences (Abla and Okanoya, 2008). Thus the development from a nucleus of a repetition detector to statistical learning may well be accessible to developmental studies using optical imaging.

In addition, the assumption that no lexico-semantic entries are present until later language development may hold only for very young infants. Bortfeld et al. (2005) were able to show that some very basic lexico-semantic entries may be present already at 6 months (“Mommy and me”). Such lexico-semantic “anchors” may well aid segmentation as discussed in a recent review (Swingley, 2009). It will be interesting to see whether these anchors can also be identified in the activation patterns when relying on optical imaging (Bortfeld et al., 2009).

CONCLUSIONS AND PERSPECTIVE

Non-invasive optical imaging has proven to be especially useful for the research on the neuronal underpinnings of speech perception during language acquisition. This research area greatly profits from the method’s clear advantages to supply a silent non-invasive and low-constraint alternative to fMRI. Longitudinal studies tracking the emerging and pre-existing lateralization of auditory and language functions can be considered a solid basis to investigate the rapid unfolding of the cornucopia of linguistic functions during the first years of life. Simultaneous EEG-assessments can converge and extend knowledge along different modalities and time scales. Though fascinating in its broad applicability, it is necessary to also enquire into the somewhat cumbersome topics of technical advances and – not least – questions concerning the physiology of the hemodynamic response. Combined fMRI experiments and careful designs respecting the specific limitations of transcranial optical spectroscopy will decide on the scientific future of this new tool in infant and language research.

ACKNOWLEDGMENTS

Financial support of the EU (NEST 012778, EFRE 20002006 2/6, nEUROpt 201076), BMBF (BNIC, Bernstein Center for Computational Neuroscience, German-Polish cooperation FK: 01GZ0710), and DFG (WA2155) are gratefully acknowledged. Isabell Wartenburger is supported by the Stifterverband für die Deutsche Wissenschaft (Claussen-Simon-Stiftung). We thank Stephen Lapaz for helping to improve the English of the manuscript.

- neural bases. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 1001–1010.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., and Naatanen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* 1, 351–353.
- Colonnese, M. T., Phillips, M. A., Constantine-Paton, M., Kaila, K., and Jasanoff, A. (2008). Development of hemodynamic responses and functional connectivity in rat somatosensory cortex. *Nat. Neurosci.* 11, 72–79.
- Cope, M., and Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med. Biol. Eng. Comput.* 26, 289–294.
- Cutler, A. (1994). The perception of rhythm in language. *Cognition* 50, 79–81.
- Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40 (Pt 2), 141–201.
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1983). A language-specific comprehension strategy. *Nature* 304, 159–160.
- DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176.
- DeCasper, A. J., and Prescott, P. (2009). Lateralized processes constrain auditory reinforcement in human newborns. *Hear. Res.* 255, 135–141.
- Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., and Dubois, J. (2006). Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends Neurosci.* 29, 367–373.
- Draganova, R., Eswaran, H., Murphy, P., Lowery, C., and Preissl, H. (2007). Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Hum. Dev.* 83, 199–207.
- Fox, P. T., and Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. U.S.A.* 83, 1140–1144.
- Friederici, A. D. (2004). Event-related brain potential studies in language. *Curr. Neurol. Neurosci. Rep.* 4, 466–470.
- Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends Cogn. Sci.* 9, 481–488.
- Friederici, A. D., and Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang.* 89, 267–276.
- Friederici, A. D., von Cramon, D. Y., and Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron* 53, 135–145.
- Gandour, J., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Sathannuwong, N., and Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang.* 84, 318–336.
- Gernsbacher, M. A., and Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol.* 54, 91–114.
- Gervain, J., Macagno, F., Cogoi, S., Pena, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14222–14227.
- Gervain, J., and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu. Rev. Psychol.* 61, 191–218.
- Gleitman, L., and Wanner, E. (1982). "Language acquisition: the state of the art," in *Language Acquisition: The State of the Art*, eds E. Wanner and L. Gleitman (Cambridge, MA: Cambridge University Press), 3–48.
- Hald, L. A., Bastiaansen, M. C., and Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Lang.* 96, 90–105.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hebden, J. C., and Austin, T. (2007). Optical tomography of the neonatal brain. *Eur. Radiol.* 17, 2926–2933.
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., and Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci. Res.* 54, 276–280.
- Homae, F., Watanabe, H., Nakano, T., and Taga, G. (2007). Prosodic processing in the developing brain. *Neurosci. Res.* 59, 29–39.
- Hull, R., Bortfeld, H., and Koons, S. (2009). Near-infrared spectroscopy and cortical responses to speech production. *Open. Neuroimag. J.* 3, 26–30.
- Huotilainen, M., Shestakova, A., and Hukki, J. (2008). Using magnetoencephalography in assessing auditory skills in infants and children. *Int. J. Psychophysiol.* 68, 123–129.
- Iadecola, C. (2004). Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nat. Rev. Neurosci.* 5, 347–360.
- Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci.* 3, 323–328.
- Kleinschmidt, A., Obrig, H., Requardt, M., Merboldt, K. D., Dirnagl, U., Villringer, A., and Frahm, J. (1996). Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy. *J. Cereb. Blood Flow Metab.* 16, 817–826.
- Koch, S. P., Steinbrink, J., Villringer, A., and Obrig, H. (2006). Synchronization between background activity and visually evoked potential is not mirrored by focal hyperoxygenation: implications for the interpretation of vascular brain imaging. *J. Neurosci.* 26, 4940–4948.
- Koch, S. P., Werner, P., Steinbrink, J., Fries, P., and Obrig, H. (2009). Stimulus-induced and state-dependent sustained gamma activity is tightly coupled to the hemodynamic response in humans. *J. Neurosci.* 29, 13962–13970.
- Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Brain Res. Cogn. Brain Res.* 24, 109–116.
- Kuhl, P., and Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu. Rev. Neurosci.* 31, 511–534.
- Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120.
- Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9096–9101.
- Kusaka, T., Kawada, K., Okubo, K., Nagano, K., Namba, M., Okada, H., Imai, T., Isobe, K., and Itoh, S. (2004). Noninvasive optical imaging in the visual cortex in young infants. *Hum. Brain Mapp.* 22, 122–132.
- Kushnerenko, E., Cheour, M., Ceponiene, R., Fellman, V., Renlund, M., Soininen, K., Alku, P., Koskinen, M., Sainio, K., and Naatanen, R. (2001). Central auditory processing of durational changes in complex speech patterns by newborns: an event-related brain potential study. *Dev. Neuropsychol.* 19, 83–97.
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.
- Lauritzen, M., and Gold, L. (2003). Brain function and neurophysiological correlates of signals used in functional neuroimaging. *J. Neurosci.* 23, 3972–3980.
- Liebert, A., Wabnitz, H., Steinbrink, J., Moller, M., Macdonald, R., Rinneberg, H., Villringer, A., and Obrig, H. (2005). Bed-side assessment of cerebral perfusion in stroke patients based on optical monitoring of a dye bolus by time-resolved diffuse reflectance. *Neuroimage* 24, 426–435.
- Liebert, A., Wabnitz, H., Steinbrink, J., Obrig, H., Moller, M., Macdonald, R., Villringer, A., and Rinneberg, H. (2004). Time-resolved multidistance near-infrared spectroscopy of the adult head: intracerebral and extracerebral absorption changes from moments of distribution of times of flight of photons. *Appl. Opt.* 43, 3037–3047.
- Lloyd-Fox, S., Blasi, A., and Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284.
- Logothetis, N. K., and Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66, 735–769.
- Mandeville, J. B., Marota, J. J., Ayata, C., Zaharchuk, G., Moskowitz, M. A., Rosen, B. R., and Weisskoff, R. M. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J. Cereb. Blood Flow Metab.* 19, 679–689.
- Mannel, C., and Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *J. Cogn. Neurosci.* 21, 1988–2006.
- Marcar, V. L., Strassle, A. E., Loenneker, T., Schwarz, U., and Martin, E. (2004). The influence of cortical maturation on the BOLD response: an fMRI study of visual cortex in children. *Pediatr. Res.* 56, 967–974.
- Mehler, J., Bertoncini, J., and Barriere, M. (1978). Infant recognition of mother's voice. *Perception* 7, 491–497.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29, 143–178.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., and von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* 17, 73–88.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., and von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang.* 89, 277–289.
- Minagawa-Kawai, Y., Mori, K., Furuya, I., Hayashi, R., and Sato, Y. (2002). Assessing cerebral representations of short and long vowel categories by NIRS. *Neuroreport* 13, 581–584.

- Minagawa-Kawai, Y., Mori, K., Hebden, J. C., and Dupoux, E. (2008). Optical imaging of infants' neurocognitive development: recent advances and perspectives. *Dev. Neurobiol.* 68, 712–728.
- Minagawa-Kawai, Y., Mori, K., Naoi, N., and Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *J. Neurosci.* 27, 315–321.
- Minagawa-Kawai, Y., Mori, K., and Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *J. Cogn. Neurosci.* 17, 1376–1385.
- Minagawa-Kawai, Y., Mori, K., Sato, Y., and Koizumi, T. (2004). Differential cortical responses in second language learners to different vowel contrasts. *Neuroreport* 15, 899–903.
- Moore, D. R., and Shannon, R. V. (2009). Beyond cochlear implants: awakening the deafened brain. *Nat. Neurosci.* 12, 686–691.
- Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.
- Oberecker, R., Friedrich, M., and Friederici, A. D. (2005). Neural correlates of syntactic processing in two-year-olds. *J. Cogn. Neurosci.* 17, 1667–1678.
- Obrig, H., and Villringer, A. (2003). Beyond the visible – imaging the human brain with light. *J. Cereb. Blood Flow Metab.* 23, 1–18.
- Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, E., and Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11702–11705.
- Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 1071–1086.
- Pulvermuller, F., Kherif, F., Hauk, O., Mohr, B., and Nimmo-Smith, I. (2009). Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. *Hum. Brain Mapp.* 30, 3837–3850.
- Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., and Obrig, H. (2010). Implicit processing of phonotactic cues: evidence from electrophysiological and vascular responses. *J. Cogn. Neurosci.* (in press).
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., Kobayashi, M., and Toshima, T. (2007a). Frontal cerebral blood flow change associated with infant-directed speech. *Arch. Dis. Child. Fetal Neonatal Ed.* 92, F113–F116.
- Saito, Y., Kondo, T., Aoyama, S., Fukumoto, R., Konishi, N., Nakamura, K., Kobayashi, M., and Toshima, T. (2007b). The function of the frontal lobe in neonates for response to a prosodic voice. *Early Hum. Dev.* 83, 225–230.
- Saito, Y., Fukuhara, R., Aoyama, S., and Toshima, T. (2009). Frontal brain activation in premature infants' response to auditory stimuli in neonatal intensive care unit. *Early Hum. Dev.* 85, 471–474.
- Sambeth, A., Pakarinen, S., Ruohio, K., Fellman, V., van Zuijen, T. L., and Huottilainen, M. (2009). Change detection in newborns using a multiple deviant paradigm: a study using magnetoencephalography. *Clin. Neurophysiol.* 120, 530–538.
- Sato, Y., Mori, K., Furuya, I., Hayashi, R., Minagawa-Kawai, Y., and Koizumi, T. (2003). Developmental changes in cerebral lateralization to spoken language in infants: Measured by near-infrared spectroscopy. *Jpn. J. Logoped. Phoniatr.* 44, 165–171.
- Sato, Y., Sogabe, Y., and Mazuka, R. (2007). Brain responses in the processing of lexical pitch-accent by Japanese speakers. *Neuroreport* 18, 2001–2004.
- Sato, Y., Sogabe, Y., and Mazuka, R. (2009). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *J. Cogn. Neurosci.* 22, 2503–2513.
- Schonwiesner, M., Rubsamen, R., and von Cramon, D. Y. (2005). Spectral and temporal processing in the human auditory cortex – revisited. *Ann. N. Y. Acad. Sci.* 1060, 89–92.
- Shapleske, J., Rossell, S. L., Woodruff, P. W., and David, A. S. (1999). The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Res. Brain Res. Rev.* 29, 26–49.
- Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., and Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI-fNIRS studies. *Magn. Reson. Imaging* 24, 495–505.
- Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196.
- Stephan, K. E., Harrison, L. M., Penny, W. D., and Friston, K. J. (2004). Biophysical models of fMRI responses. *Curr. Opin. Neurobiol.* 14, 629–635.
- Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259, 39–41.
- Swingle, D. (2009). Contributions of infant word learning to language development. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 3617–3632.
- Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., and Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733.
- Villringer, A., and Dirnagl, U. (1995). Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. *Cerebrovasc. Brain Metab. Rev.* 7, 240–276.
- Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A. D., and Obrig, H. (2007). The processing of prosody: evidence of interhemispheric specialization at the age of four. *Neuroimage* 34, 416–425.
- Wenzel, R., Wobst, P., Heekeren, H. H., Kwong, K. K., Brandt, S. A., Kohl, M., Obrig, H., Dirnagl, U., and Villringer, A. (2000). Saccadic suppression induces focal hypooxygenation in the occipital cortex. *J. Cereb. Blood Flow Metab.* 20, 1103–1110.
- Werker, J. F., and Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* 50, 509–535.
- White, B. R., Snyder, A. Z., Cohen, A. L., Petersen, S. E., Raichle, M. E., Schlaggar, B. L., and Culver, J. P. (2009). Mapping the human brain at rest with diffuse optical tomography. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 4070–4072.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R. J., and Naatanen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642.
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
- Zeff, B. W., White, B. R., Dehghani, H., Schlaggar, B. L., and Culver, J. P. (2007). Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12169–12174.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2010; paper pending published: 31 March 2010; accepted: 27 May 2010; published online: 23 June 2010.

Citation: Obrig H, Rossi S, Telkemeyer S and Wartenburger I (2010) From acoustic segmentation to language processing: evidence from optical imaging. *Front. Neuroenerg.* 2:13. doi: 10.3389/fneuro.2010.00013
Copyright © 2010 Obrig, Rossi, Telkemeyer and Wartenburger. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.