Obleser, J., and Weisz, N. (**2010**). "Differential influences of spectral and temporal features of speech on human oscillatory brain dynamics" Program No. 837.6. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010. Online.

Obleser, J., Wise, R. J. S., Alex Dresner, M., and Scott, S. K. (**2007**). "Functional integration across brain regions improves speech perception under adverse listening conditions" The Journal of Neuroscience: J. Neurosci., **27**, 2283-2289.

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., Carlyon, R. P., and Griffiths, T. D. (**2007**). "An information theoretic characterisation of auditory encoding" PLoS Biol., **5**, e288.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise" J. Acoust. Soc. Am., **97**, 593-608.

Pichora-Fuller, M. K., and Singh, G. (**2006**). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation" Trends Amplif., **10**, 29-29.

Rauschecker, J. P., and Scott, S. K. (**2009**). "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing" Nat. Neurosci., **12**, 718-724.

Scherg, M., Vajsar, J., and Picton, T. W. (**1989**). "A source analysis of the late human auditory evoked potentials" J. Cogn. Neurosci., **1**, 336-355.

Schreiner, C. E., and Winer, J. A. (**2007**). "Auditory cortex mapmaking: principles, projections, and plasticity." Neuron, **56**, 356-365.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (**2000**). "Identification of a pathway for intelligible speech in the left temporal lobe" Brain., **123**, 2400-2406.

Shannon, R. V., Fu, Q.-J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation" Acta Otol. Suppl., **552**, 50-54.

Sharp, D. J., Scott, S. K., Mehta, M. A., and Wise, R. J. S. (**2006**). "The neural correlates of declining performance with age: evidence for age-related changes in cognitive control" Cereb. Cortex, **16**, 1739-1749.

Shinn-Cunningham, B. G., and Best, V. (**2008**). "Selective attention in normal and impaired hearing" Trends Amplif., **12**, 283-299.

Stickney, G. S., and Assmann, P. F. (**2001**). "Acoustic and linguistic factors in the perception of bandpass-filtered speech" J. Acoust. Soc. Am., **109**, 1157-1165.

Weisz, N., Hartmann, T., Müller, N., and Obleser, J. (**2011**). "Alpha Rhythms in Audition : Cognitive and Clinical Perspectives" Front. Psychol., **2**, 73.

Dynamic and task-dependent encoding of speech and voice in the auditory cortex

MILENE BONTE [1,2] AND ELIA FORMISANO [1,2]

[1] Maastricht Brain Imaging Center and [2]Dept. of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

Speech is at the core of verbal communication and social interaction. It conveys linguistic content and speaker-specific vocal information that listeners exploit for identification. Cortical processing of speech relies on the formation of abstract representations that are invariant to highly variable acoustic input signals and critically depends on behavioral demands. In a series of EEG and fMRI studies we have recently investigated temporal as well as spatial neural coding mechanisms for forming such abstract representations. We focused on categorical and task-dependent neuronal responses to natural speech sounds (vowels /a/, /i/, /u/) spoken by different speakers. Brain activity was measured during passive listening (fMRI, EEG) and during performance of behavioural tasks on vowel or speaker identity (EEG). Our EEG results show that dynamic changes of sound-evoked responses and phase patterns of cortical oscillations in the alpha band (8-12 Hz) closely reflect the abstraction and analysis of the sounds along the task-relevant dimension. Our fMRI results show that spatially distributed activation patterns in early and higher level auditory cortex encode vowel-invariant representations of speaker identity and speaker-invariant representations of vowel identity. Both the transient and task-dependent realignment of neuronal responses (EEG) and the spatially distributed cortical fingerprints (fMRI) provide robust cortical coding mechanisms for forming abstract representations of auditory (speech) signals.

We are suprisingly efficient in understanding who is speaking and what is being said from highly variable speech signals. Furthermore, dependent on the current behavioural goal, we may choose to focus our attention on either speaker identity or speech content and ignore the other dimension. Such adaptive behaviour requires computational mechanisms that enable different (abstract) representations of the same acoustic input.

Cognitive and connectionist models suggest that speech recognition involves the formation of intermediate entities such as phonemes that are invariant to changes in the acoustic input (due to speaker variability, noise or signal distortion) and that can be used for further linguistic processing (McClelland and Elman, 1986; Norris and McQueen, 2008). Similarly, speaker recognition may involve the formation of

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

abstract representations of a person's voice based on a unique combination of characteristics such as fundamental frequency, timbre and breathiness (Murry and Singh, 1980; Klatt and Klatt, 1990). In this paper we discuss recent work showing first evidence for spectrotemporal (Bonte *et al.*, 2009) and spatial (Formisano *et al.*, 2008a) neural mechanisms that enable this abstraction.

CORTICAL RESPONSES TO SPEECH SOUNDS AND VOICES

Speech - as any other acoustically complex sound - evokes extensive and bilateral activation in the superior temporal cortex, including the primary areas on Heschl's gyrus, the planum temporale, the superior temporal gyus (STG) and superior temporal sulcus (STS) (Binder *et al.*, 2000) .

Over the last decade, a key issue in neuroimaging research has been the investigation of specialized cortical areas/pathways for the processing auditory categories. Together with lesion and electrophysiology studies, this research has lead to the formulation of several models for the cortical representation of human voices and speech. These models suggest a functional segregation of areas specialized for the processing of 1) speech sounds in the mid-posterior (Hickok and Poeppel, 2007) or anterior (Rauschecker and Scott, 2009) superior temporal cortex, and of 2) vocal sounds along the STG/STS, with strong voice selectivity especially in the right anterior STS (Belin *et al.*, 2004).

Research using methods with excellent temporal resolution, such as electroencephalography (EEG) and magneto-encephalography (MEG), has indicated that basic speech features modulate activity in non-primary auditory cortex from 50-100 ms onwards (Poeppel *et al.*, 1996; Obleser *et al.*, 2004b; Parviainen *et al.*, 2005). The analysis of more complex speech features continues after 100–200 ms (Näätänen *et al.*, 1997; Bonte *et al.*, 2005). Pre-attentive discrimination of voices has been estimated at 100-200 ms (Titova and Näätänen, 2001; Beauchemin *et al.*, 2006). In a subsequent time-window, i.e. 280-420 ms, attended voices have been reported to elicit stronger activity than voice-like instrumental sounds (Levy *et al.*, 2003). Furthermore, the few studies that simultaneously examined speech and voice processing suggest that early EEG/MEG responses to phonemes reflect processing of specific acoustic features (formants) independently of pitch (Poeppel *et al.*, 1997) or speaker (Obleser *et al.*, 2004a; Shestakova *et al.*, 2004), whereas no speech-invariant responses to speakers have been reported.

SPECTRO-TEMPORAL CODING MECHANISMS

To better understand how our auditory cortex facilitates selective attention to speech content versus voice information, a crucial factor that needs to be addressed is the way in which neural responses to speech signals are modulated by task demands. We therefore examined neurophysiological activity underlying the task-dependent representation of speech sounds and speakers' voices (Bonte *et al.*, 2009). Using EEG, we recorded and analyzed cortical responses to three vowels: /a/, /i/ and /u/, spoken by three speakers (Fig. 1). The behaviorally relevant stimulus dimension was manipulated by presenting these stimuli in the context of different tasks: (1) a one-back task on speaker identity (Speaker task), (2) a one-back task on vowel identity (Vowel task) and (3) passive listening (Passive task). Correct performance of the one-back tasks requires abstraction of the behaviorally relevant percept from the acoustic input - speaker or vowel identity, respectively - and its maintenance in working memory for comparison with the following sound.

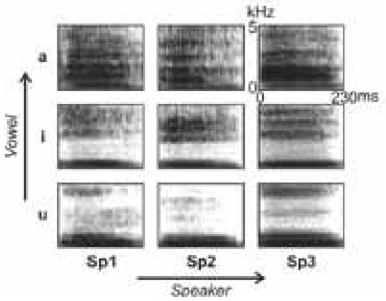

Fig. 1: Stimuli and Design. Examples of spectrograms of three vowels (/a/, /i/, /u/) pronounced by three speakers (speaker 1, 2, 3). Adapted from Formisano *et al.*, 2008a.

Time-course of stimulus-driven and categorical analysis

In a first analysis we focused on event-related potentials (ERPs). In both active tasks, speech stimuli elicited a comparable sequence of typical auditory responses, including an N1 (90-130 ms), and P2 (170-230 ms) response, followed by a positive ERP around 340 ms (Fig. 2). Amplitude changes in these responses delineate a stage of largely stimulus-driven analysis followed by task-specific processing of behaviorally relevant stimulus categories. Stimulus-driven analysis was indicated in the first 200 ms by stimulus-dependent N1-P2 amplitude differences that were unaffected by our task manipulations. This bottom-up analysis is most likely driven by characteristic acoustic-phonetic features such as fundamental frequency, timbre or breathiness for speaker discrimination (Murry and Singh, 1980; Klatt and Klatt, 1990), and the first and second formant frequencies for vowel discrimination (Obleser *et al.*, 2004b; Shestakova *et al.*, 2004).

Task-dependent processing of speech stimuli occurred around 310-370 ms, as indexed by the selective enhancement of ERP amplitude differences for either vowels (Vowel task) or speakers (Speaker task). This amplification of task-relevant stimulus differences may reflect cortical processing and formation of abstract vowel or voice representations based on the outcome of a perceptual analysis of acoustic-phonetic stimulus characteristics. Similar processes may be utilized during the allocation of attention to different classes of auditory objects (including speech sounds - Bonte *et al.*, 2006, and voices - Levy *et al.*, 2003) such as encountered in complex auditory scenes during everyday life.

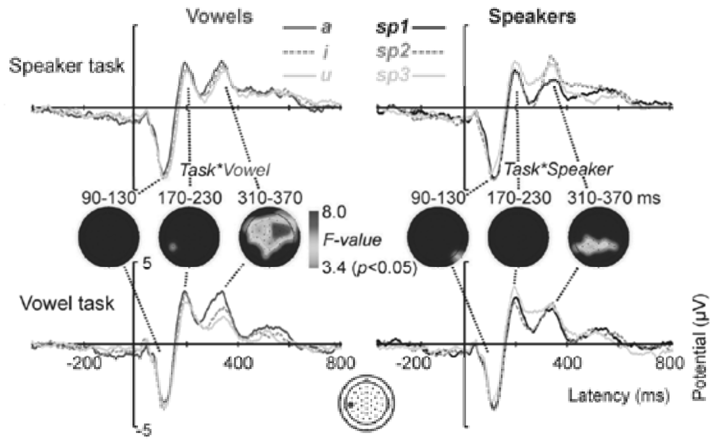

Fig. 2: ERPs recorded at a left temporal electrode during the Speaker and Vowel tasks. ERP activity elicited by all 9 speech sounds was grouped across vowels (left panel) or speakers (right panel). Statistical topographic maps show Task * Stimulus interactions (thresholded at p<0.05). Adapted from Bonte *et al.*, 2009.

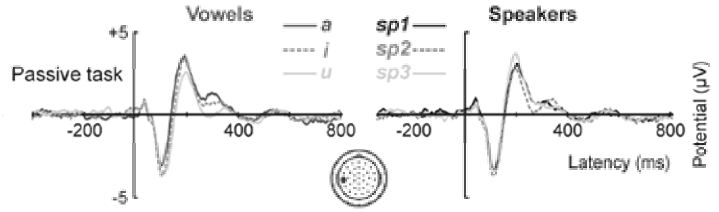

Fig. 3: 'Default' categorization of vowels during passive listening. Grand-average ERPs recorded at a left temporal electrode during passive listening. Adapted from Bonte *et al.*, 2009.

Our ERP results also show that the linguistic dimension (vowels) represents the 'default' processing mode for speech. That is, during passive listening ERP responses around 340 ms showed a selective amplification of vowel, but not of speaker, differences comparable to the Vowel task (Fig. 3). This enhanced saliency of vowels as compared to speakers during passive listening was also found in a subsequent analysis of cortical oscillations (see Bonte *et al.,* 2009).

Cortical coding of vowel/speaker invariance

To investigate the role of neuronal oscillations in coding abstract properties of speech stimuli, we further focused on inter-trial phase coherence (ITC) and examined how performance of the Vowel versus Speaker task influenced the synchronization between EEG responses. We focused on phase information, rather than oscillatory power estimates, because it directly relates to the timing of neural activity (Engel *et al.*, 2001; Salinas and Sejnowski, 2001). More specifically, we hypothesized that the analysis of abstract properties of speech (vowel/speaker identity) is encoded by a task-dependent realignment of oscillatory activity to different acoustic events (Fig. 4).

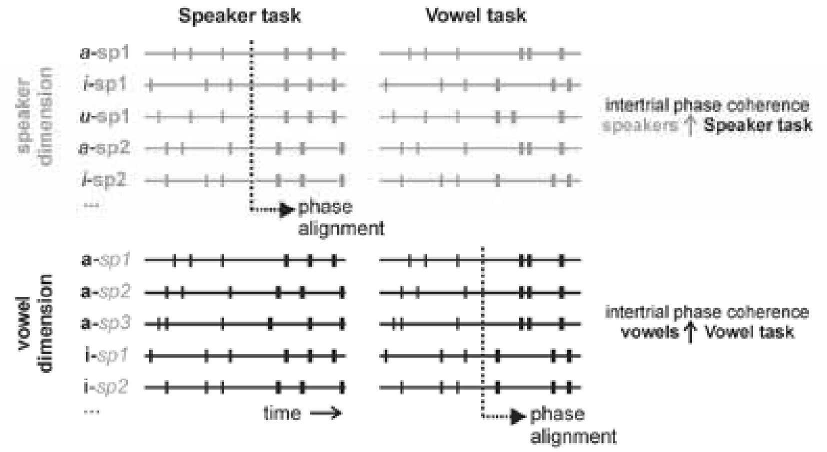

Fig. 4: Hypothesized neural coding scheme. Subsequent to a stimulus driven analysis, during which each speech stimulus elicits a unique pattern of neural responses (indicated by tick marks), top-down task demands lead to a phase alignment of stimulus-locked oscillatory responses. Thus, the Speaker task aligns phases for each of the speakers, independent of the vowel that was pronounced (left part, upper row), whereas the Vowel task aligns phases for each of the vowels, independent of who was speaking (right part, lower row). Adapted from Bonte *et al.*, 2009.

The predicted phase alignment of oscillatory responses along the behaviourally relevant stimulus dimension started around 250 ms after stimulus onset (Fig. 5). Thus, at temporal electrodes alpha (8-12 Hz) oscillations showed significantly increased phase synchronization between speakers as compared to vowels during the Speaker task, and the opposite pattern of increased synchronization between vowels as compared to speakers in the Vowel task. This alpha phase alignment may organize neurophysiological responses such that the behaviorally relevant stimulus dimension is maintained for further processing (von Stein *et al.*, 2000) Our findings highlight the importance of the precise timing of alpha oscillations for cortical information processing. Furthermore, they demonstrate a specific role of alpha phase alignment in the adaptive tuning of neural activity enabling abstract and task-dependent analysis of sensory input.

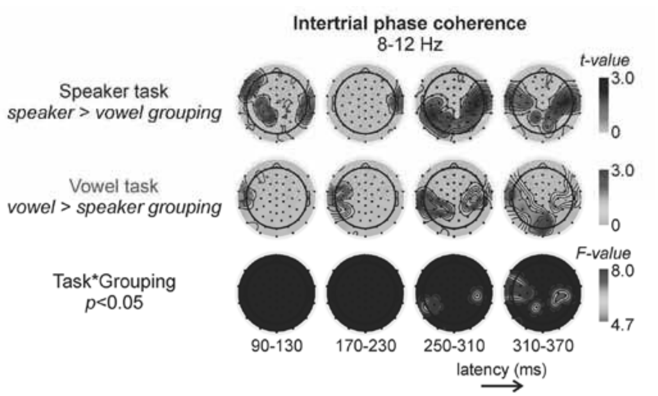

Fig. 5: Task-dependent phase reorganization of stimulus-locked alpha oscillations. Statistical topographic maps in different time windows illustrating areas of increased alpha intertrial phase coherene across speakers in the Speaker task (upper row), across vowels in the Vowel task (middle row), and electrodes showing significant Task * Stimulus interactions (lower row, thresholded at $p<0.05$). Adapted from Bonte *et al.*, 2009.

Interestingly, phase reorganization and ERP amplitude modulations occurred in the same time window. This raises the possibility that alpha phase alignment contributes to the task-dependent enhancement of ERP amplitude differences (Fig. 2), which is consistent with the suggestion that ERP waveforms are modulated by the precise timing of alpha oscillations (Makeig *et al.*, 2002; Klimesch *et al.*, 2007; Mazaheri and Jensen, 2008).

As both speaker-independent vowel, and vowel-independent speaker analysis demonstrated comparable time-courses and oscillatory patterning they most likely relied on similar computational mechanisms. The different spatial distribution of the effects, and in particular a left hemispheric bias for the Vowel but not the Speaker task (Bonte *et al,* 2009), indicates the involvement of distinct networks of brain areas. Based on this observation and the putative role of alpha oscillations in modulating functional connectivity between brain areas (von Stein *et al.*, 2000; Kujala *et al.*, 2007) we hypothesize that the observed phase reorganization operates at an interregional level. In particular, this reorganization may mediate temporal binding of distributed neural activity in distinct (auditory) cortical areas.

AUDITORY CORTICAL FINGERPRINTS

Auditory cortical processing of speech and voice has been investigated in numerous functional magentic resonance imaging (fMRI) studies. By comparing speech or vocal sounds to control sounds, several cortical regions in the superior temporal cortex have been characterized in terms of their “selectivity” or “specialization” for individual features that are relevant to the “speechness” or “voiceness” of the stimuli. A proposed functional hierarchy within these areas for the processing of speech sounds (Hickock and Poeppel, 2007; Rauschecker and Scott, 2009) and voices (Belin *et al*., 2004) was also found in studies using fMRI designs that do not rely on “voice vs non-voice” (adaptation design, Belin and Zatorre, 2003) or “speech vs non-speech” comparisons (parametric design, Davis and Johnsrude, 2003). However, each of these fMRI designs relies on between category differences, which only allow for partial and indirect inferences on what distinguishes the auditory cortical representations of individual sounds. Furthermore, it remains unclear how the implicated brain areas contribute to the transformation of a speech sound into the more abstract entity of “phoneme” or “speaker” identity.

To investigate speaker-independent representations of individual phonemes and phoneme-independent representation of individual voices, in a recent study (Formisano *et al*., 2008a), we measured single-trial fMRI responses while participants passively listened to the vowels /a/, /i/ and /u/ spoken by three speakers (Fig. 1). Consistent with previous studies, sounds evoked significant fMRI responses in a wide expanse of the superior temporal cortex, including early auditory areas (Heschl’s gyrus) and multiple regions in the planum temporale (PT), along the superior temporal gyrus (STG), the superior temporal sulcus (STS), and the middle temporal gyrus. Univariate statistical contrasts, however, yielded only weak response differences (below significance) or no differences between conditions.

By applying multivariate statistical pattern recognition techniques (De Martino *et al.*, 2008; Formisano *et al.*, 2008b) to the single-trial fMRI responses we were able to estimate and decode the distinct activation patterns elicited by the vowels and speakers and directly assess the invariance of the estimated neural representations. In a first step, a machine learning algorithm was trained to associate labeled fMRI activation patterns with the different vowels /a/, /i/, and /u/, independent of who pronounced the vowel (“vowel learning”) or with the different speakers sp1, sp2 and sp3, independent of which vowel was pronounced (“speaker learning”). In a second step, we tested whether our algorithm would correctly classify unlabeled fMRI activation patterns. Furthermore training and testing were either performed on all stimuli: *discrimination analysis*, or vowel (speaker) learning was performed on one of the speakers (vowels) and testing was performed on the remaining speakers (vowels): *generalization analysis*. In the generalization analysis, stimuli used for training and testing differ in many acoustical dimensions and accurate decoding of activation patterns would thus indicate that the associated neuronal populations entail abstract information of that particular vowel (or speaker) beyond the contingent mapping of its acoustic properties.

In all subjects and in both the discrimination and the generalization analysis, the algorithm successfully learned the functional relation between sounds and corresponding spatial patterns and classified correctly the unlabeled sound-evoked patterns (see Formisano *et al.*, 2008a). Cortical locations that contributed most to the discrimination of vowels (Fig. 6, left panel) were widely distributed bilaterally in the superior temporal cortex, including classical speech related areas in the PT (mainly in the left hemisphere), extended portions of the STS/STG (both hemispheres) but also regions in early auditory cortex. Discriminative patterns for speakers (Fig. 6, right panel) were more confined and right-lateralized and included early auditory areas adjacent to the region obtained for vowels and three clustered regions along the anterior-posterior axis of the right STS, also interspersed with vowel regions. Although sparser, the discriminative maps resulting from the generalization analysis included a subset of these locations for vowel and speaker discrimination (outlined regions in Fig. 6).

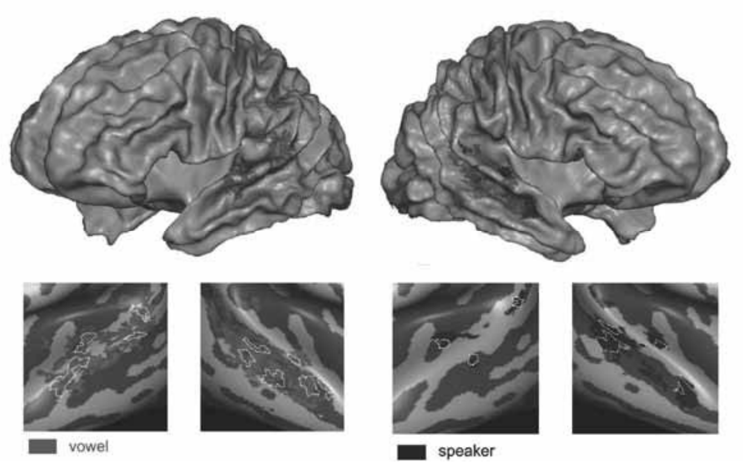

Fig. 6: Cortical discriminative maps for decoding of vowels and speakers. Maps are visualized on the folded cortex and on an inflated representation of the auditory cortex (light gray, gyri, dark gray, sulci). A location was color-coded (vowels, red; speakers, blue) if it was present on the individual maps of at least four of the seven subjects. Outlined regions on the inflated auditory cortex representations indicate cortical regions that were also included in the group maps obtained with the generalization analysis. Adapted from Formisano *et al.*, 2008a.

Spatial proximity of discriminative activation patterns for the individual vowels and speakers was visualized using self organizing maps (SOMs), which convert complex relations between high-dimensional items into simple geometric relations (Fig. 7). The spatial proximity and grouping of the conditions in the SOM based two-dimensional display thus reflects the level of abstraction and categorical information entailed in the discriminative patterns of vowels and speakers. To investigate which acoustic features drive this neural abstraction, we examined the relative distance between the brain based representations of the stimuli and their description in terms of typical acoustic features. We found that the distances between the cortical representations of the sounds correlated best with a description of the stimulus based on the first two formants (F1, F2) in the case of vowels (Fig. 8, left panel) and on the fundamental frequency (F0) in the case of speakers (Fig. 8, right panel). These results provide empirical support for cognitive models of speech and voice processing postulating the existence of abstract intermediate representations resulting from the transformation of relevant acoustic features (the (F1, F2) pair for vowels and (F0) for speakers) and the suppression of the irrelevant ones.

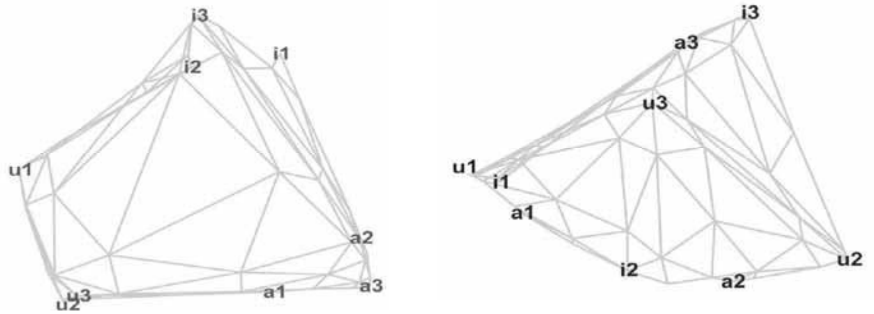

Fig. 7: SOM-based display of the discriminative patterns in the group of seven subjects for vowel (left panel, in red) and speaker learning (right panel, in blue). Adapted from Formisano *et al.*, 2008a.

In conclusion, this fMRI study demonstrated the feasibility of decoding speech content and speaker identity from auditory cortical activation patterns. Furthermore, our findings prompt for a revision of models on speech and voice abstraction which assume that a hierarchy of processing steps is "mapped" into a functional hierarchy of specialized neural modules. Instead our findings suggest that an abstract representation of a vowel or speaker emerges from the joint encoding of information occurring not only in specialized higher-level regions but also in auditory regions, which—because of their anatomical connectivity and response properties— have been associated with early stages of sound processing. These findings have been further corroborated in a subsequent fMRI study where cross-modal recalibration was employed to bias the perceptual categorization of acoustically identical speech sounds into distinct auditory phonemes (Kilian-Hütten *et al.,* 2011). In this study, classification of fMRI responses was based on purely perceptual labels, thus controlling for the acoustic similarity among stimuli (e.g. among the same vowels even when uttered by distinct speakers). Consistent with the hypothesis that constructive perceptual processes take place in relatively early cortical auditory networks, the results showed that it is possible to retrieve the perceptual interpretation of ambiguous phonemes-information from brain activity in auditory areas in the superior temporal cortex, most prominently on the posterior bank of the left Heschl's gyrus and sulcus and in the adjoining left planum temporale (Kilian-Hütten *et al*., 2011).

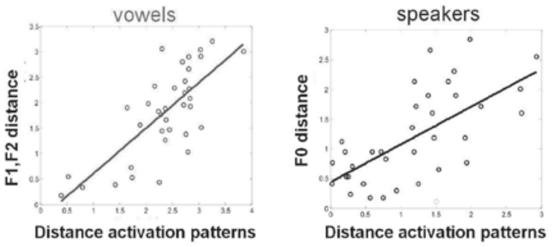

Fig. 8: Relation between normalized distances of the multidimensional auditory cortical activation patterns and normalized distances of the vowels in the (F1, F2) space of formants (left panel, in red) and of the speakers in the space of fundamental frequency (F0) (right panel, in blue). Adapted from Formisano *et al.*, 2008a.

CONCLUSION

Our results provide a detailed empirical demonstration of how the human brain forms computationally efficient representations required for speech comprehension and speaker identification. In particular, they lead to the hypothesis that abstract encoding of vowel and speaker identity relies on temporal binding across distinct and distributed neuronal populations throughout the superior temporal cortex (Formisano *et al.*, 2008a) by means of transient and goal-dependent alignment of their firing patterns (Bonte *et al.*, 2009).

The presented studies measured brain responses in healthy adults to three vowels and three speakers that were presented in isolation in order to obtain distinct neural activation patterns. Extension of these results to identify a word or concatenation of words in streams of longer speech segments and in varying acoustic conditions (e.g. noisy environments), provides a compelling challenge and will contribute to create a general brain-based decoder of sounds in the context of real-life situations. Furthermore, extension to different subject populations may reveal relevant aspects of learning and plasticity in auditory cortical representations during normal and anomalous development and hearing impairment.

REFERENCES

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., and Lassonde, M. (**2006**). "Electrophysiological markers of voice familiarity" Eur. J. Neurosci. **23**, 3081-3086.

Belin, P., Fecteau, S., and Bedard, C. (**2004**). "Thinking the voice: neural correlates of voice perception" Trends Cogn Sci **8**, 129-135.

Belin, P., and Zatorre, R. J. (**2003**). "Adaptation to speaker's voice in right anterior temporal lobe" Neuroreport **14**, 2105-2109.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (**2000**). "Human temporal lobe activation by speech and nonspeech sounds" Cereb. Cortex **10**, 512-528.

Bonte, M., Parviainen, T., Hytonen, K., and Salmelin, R. (**2006**). "Time course of top-down and bottom-up influences on syllable processing in the auditory cortex" Cereb. Cortex **16**, 115-123.

Bonte, M., Valente, G., and Formisano, E. (**2009**). "Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations" J. Neurosci. **29**, 1699-1706.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., and Blomert, L. (**2005**). "Auditory cortical tuning to statistical regularities in phonology" Clin. Neurophysiol. **116**, 2765-2774.

Davis, M. H., and Johnsrude, I. S. (**2003**). "Hierarchical processing in spoken language comprehension" J. Neurosci. **23**, 3423-3431.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., and Formisano, E. (**2008**). "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns" Neuroimage **43**, 44-58.

Engel, A. K., Fries, P., and Singer, W. (**2001**). "Dynamic predictions: oscillations and synchrony in top-down processing" Nat. Rev. Neurosci. **2**, 704-716.

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (**2008a**). ""Who" is saying "what"? Brain-based decoding of human voice and speech" Science **322**, 970-973.

Formisano, E., De Martino, F., and Valente, G. (**2008b**). "Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning" Magn. Reson. Imaging **26**, 921-934.

Hickok, G., and Poeppel, D. (**2007**). "The cortical organization of speech processing" Nat. Rev. Neurosci. **8**, 393-402.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers" J. Acoust. Soc. Am. **87**, 820-857.

Kilian-Hutten, N., Valente, G., Vroomen, J., and Formisano, E. (**2011**). "Auditory cortex encodes the perceptual interpretation of ambiguous sound" J. Neurosci. **31**, 1715-1720.

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., and Freunberger, R. (**2007**). "Event-related phase reorganization may explain evoked neural dynamics" Neurosci. Biobehav. Rev. **31**, 1003-1016.

Kujala, J., Pammer, K., Cornelissen, P., Roebroeck, A., Formisano, E., and Salmelin, R. (**2007**). "Phase coupling in a cerebro-cerebellar network at 8-13 Hz during reading" Cereb. Cortex **17**, 1476-1485.

Levy, D. A., Granot, R., and Bentin, S. (**2003**). "Neural sensitivity to human voices: ERP evidence of task and attentional influences" Psychophysiology **40**, 291-305.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., and Sejnowski, T. J. (**2002**). "Dynamic brain sources of visual evoked responses" Science **295**, 690-694.

Mazaheri, A., and Jensen, O. (**2008**). "Asymmetric amplitude modulations of brain oscillations generate slow evoked responses" J. Neurosci. **28**, 7781-7787.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception" Cognit. Psychol. **18**, 1-86.

Murry, T., and Singh, S. (**1980**). "Multidimensional analysis of male and female voices" J. Acoust. Soc. Am. **68**, 1294-1300.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, 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.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: a Bayesian model of continuous speech recognition" Psychol. Rev. **115**, 357-395.

Obleser, J., Elbert, T., and Eulitz, C. (**2004a**). "Attentional influences on functional mapping of speech sounds in human auditory cortex" BMC Neurosci. **5**, 24.

Obleser, J., Lahiri, A., and Eulitz, C. (**2004b**). "Magnetic brain response mirrors extraction of phonological features from spoken vowels" J. Cogn. Neurosci. **16**, 31-39.

Parviainen, T., Helenius, P., and Salmelin, R. (**2005**). "Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia" Cereb. Cortex **15**, 1054-1063.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P., and Marantz, A. (**1997**). "Processing of vowels in supratemporal auditory cortex" Neurosci. Lett. **221**, 145-148.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., and Marantz, A. (**1996**). "Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds" Brain Res. Cogn. Brain. Res. **4**, 231-242.

Rauschecker, J. P., and Scott, S. K. (**2009**). "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing" Nat. Neurosci. **12**, 718-724.

Salinas, E., and Sejnowski, T. J. (**2001**). "Correlated neuronal activity and the flow of neural information" Nat. Rev. Neurosci. **2**, 539-550.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., and Huotilainen, M. (**2004**). "Orderly cortical representation of vowel categories presented by multiple exemplars" Brain Res. Cogn. Brain Res. **21**, 342-350.

Titova, N., and Näätänen, R. (**2001**). "Preattentive voice discrimination by the human brain as indexed by the mismatch negativity" Neurosci. Lett. **308**, 63-65.

von Stein, A., Chiang, C., and Konig, P. (**2000**). "Top-down processing mediated by interareal synchronization" Proc. Natl. Acad. Sci. U S A **97**, 14748-14753.

Low-frequency versus high-frequency synchronisation in chirp-evoked auditory brainstem responses

FILIP MUNCH RØNNE AND KRISTIAN GØTSCHE-RASMUSSEN

Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark

This study investigates the frequency specific contribution to the auditory brainstem response (ABR) of chirp stimuli. Frequency rising chirps were designed to compensate for the cochlear traveling wave delay, and lead to larger wave-V amplitudes than for click stimuli as more auditory nerve fibres fire synchronously. Traditional click stimuli were believed to only excite high-frequency fibres synchronously. It is still currently unclear whether the broad-band chirp stimulus leads to increased synchronisation of both low- and high-frequency fibres. It is also unclear if both these groups of fibres contribute significantly to the overall wave-V amplitude. In the present study, ABRs were recorded from 10 normal-hearing listeners using low- and high-frequency band-limited chirps and clicks (0.1 – 1.5 kHz and 1.5 - 10 kHz) presented at a level of 40 dB HL. The results showed significantly larger wave-V amplitudes for both low and high-frequency band-limited chirps than for the filtered clicks. This demonstrates that the synchronisation of nerve fibres occurs across the entire frequency range at this presentation level, and this leads to significant increases in wave-V amplitudes. The increase for the low-frequency chirp was found to be clearly larger than that obtained at the higher frequencies.

INTRODUCTION

ABRs in response to transient sound stimuli represent the summed electric potential from many remotely located neurons, recorded via scalp electrodes. The click evoked ABR has 7 distinct waves, where wave-V is the most prominent. One key feature of the ABR wave-V is the peak latency which is dependent on both stimulus frequency (Neely *et al*., 1988) and level (Dau, 2003). The frequency dependence is due to the tonotopic mapping on the basilar membrane (BM) with high-frequency at base and low-frequency at apex (Greenwood, 1990). Each frequency component of a stimulus is associated with a certain delay, and a click stimulus will thus elicit responses over a relatively large time span. This limits the synchronicity of the response, and thereby reduces the ABR amplitude evoked by such a stimulus (Elberling *et al*., 2007). Frequency rising chirps have been designed to compensate for the cochlear travelling wave delay. The use of chirp stimulus leads to larger wave-V amplitudes than for click stimuli as more auditory nerve fibres fire synchronously (see Elberling *et al*., 2007, for review). The increase in synchronicity has traditionally been argued to occur mainly at low frequencies, where the peaks of

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.