International Journal of Psychophysiology 95 (2015) 175–183

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

# The role of temporal structure in the investigation of sensory memory, auditory scene analysis, and speech perception: A healthy-aging perspective

Johanna Maria Rimmele [a,*], Elyse Sussman [b], David Poeppel [c,d]

[a] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[b] Albert Einstein College of Medicine, Dominick P. Purpura Department of Neuroscience, Bronx, NY, United States
[c] Department of Psychology and Center for Neural Science, New York University, New York, NY, United States
[d] Max-Planck Institute for Empirical Aesthetics, Frankfurt, Germany

ARTICLE INFO

*Article history:*
Received 16 October 2013
Received in revised form 13 June 2014
Accepted 15 June 2014
Available online 20 June 2014

*Keywords:*
Auditory cortex
Oscillations
MMN
Prediction
Time

ABSTRACT

Listening situations with multiple talkers or background noise are common in everyday communication and are particularly demanding for older adults. Here we review current research on auditory perception in aging individuals in order to gain insights into the challenges of listening under noisy conditions.

Informationally rich temporal structure in auditory signals – over a range of time scales from milliseconds to seconds – renders temporal processing central to perception in the auditory domain. We discuss the role of temporal structure in auditory processing, in particular from a perspective relevant for hearing in background noise, and focusing on sensory memory, auditory scene analysis, and speech perception.

Interestingly, these auditory processes, usually studied in an independent manner, show considerable overlap of processing time scales, even though each has its own 'privileged' temporal regimes. By integrating perspectives on temporal structure processing in these three areas of investigation, we aim to highlight similarities typically not recognized.



## 1. Introduction

In this review we discuss evidence for age-related deficits in temporal structure processing at relatively slow (~100 ms to several seconds) time scales. An important characteristic of any extended acoustic signal is that it evolves over time, rendering auditory processing inherently temporal. Auditory perception requires the recognition of temporal structure, that is, a patterned organization of the stimulation over time. By containing repeating patterns, a temporally structured stimulus can be, in a broader sense, 'regular,' in contrast, to a temporally unstructured stimulus that contains no systematic regularities repeated over time. Temporal information occurs at multiple time scales, associated with different perceptual phenomena (Fig. 1). For example, in the case of speech, the most compelling temporal regularity is seen in the modulation spectrum, which demonstrates (across languages) an amplitude modulation with a peak at ~5 Hz, corresponding to the mean syllabic rate. Phonemic (sub-syllabic) information is, by necessity, associated with a higher modulation rate; analogously, intonation contours at the phrasal or sentence level reflect slow regularities (typically variation of the fundamental frequency) over hundreds to thousands of milliseconds. Other signals, too, contain more 'local,' short time scale variation and more 'global,' longer scale temporal structure. Age-related deficits in temporal processing are certainly not constrained to relatively slow time scales (100 ms to several seconds). However, this review will focus on these longer time scales as they are fundamentally involved in processing temporal structure at the level of auditory cortex, including speech perception, and auditory scene analysis (ASA) (Fig. 2). These time scales are relevant for processing speech (Fig. 2b) and other natural sounds (e.g., such as a dog barking) that involve auditory sensory memory (ASM; sounds typically used to research ASM, Fig. 2a). When we listen to one out of multiple simultaneous speakers (or other sound sources), temporal characteristics of separate signals overlap (Fig. 2c). We discuss research that suggests temporal characteristics of the sound input, can aid the segregation of multiple speakers, or the process of auditory stream segregation.

We focus on age-related deficits in speech perception and ASA that may arise from deficits of temporal structure processing in ASM.

*Abbreviations:* ASA, auditory scene analysis; ASM, auditory sensory memory; ECoG, electrocorticography; EEG, electroencephalography; ERP, event-related potential; fMRI, functional magnetic resonance imaging; ITPC, inter-trial phase coherence; MEG, magnetoencephalography; MMN, mismatch negativity; SPIN test, Speech Perception in Noise test; STG, superior temporal gyrus; TFS, temporal fine structure.

* Corresponding author at: Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, D-20246 Hamburg, Germany. Tel.: +49 40 7410 0.
*E-mail address:* j.rimmele@uke.de (J.M. Rimmele).

http://dx.doi.org/10.1016/j.ijpsycho.2014.06.010
0167-8760/

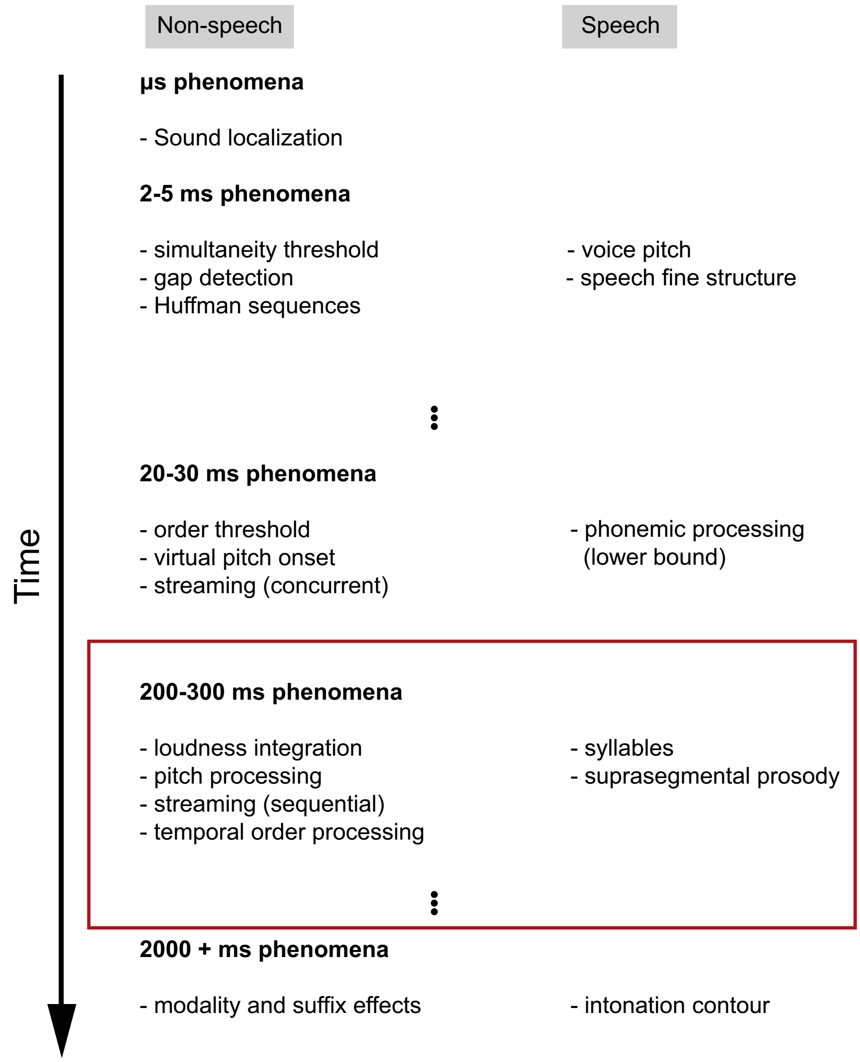


**Fig. 1.** Sound contains temporal structure at multiple time scales. Non-speech perceptual phenomena at multiple time scales are displayed in the left column; speech perceptual phenomena are displayed in the right column (time bar: short duration phenomena are displayed at the top and longer duration at the bottom). The review discusses temporal processing at a time scale from about 200 ms to several seconds (red box).

## 2. Temporal structure processing in auditory sensory memory

ASM, the retention of sound information after the physical input has ceased, plays an important role in the ability to integrate successive sounds over time into meaningful auditory events (Cowan, 1995; Neisser, 1967). The transient storage of auditory information can last up to 30 s (Sams et al., 1993; Winkler et al., 2002), depending upon how ASM is defined. However, theories differ in how they propose that temporal information is accessed at different time scales. A classical approach to the study of ASM (Cowan, 1988, 1984; Massaro, 1972; Näätänen, 1992, 1990; Näätänen et al., 1978) proposes two different types of sensory memory, one for short storage (called "short auditory store/storage") and another for longer storage (called "long auditory store/storage") (Cowan, 1988, 1984; Massaro, 1972). The short auditory store refers to processes occurring at a faster time scale (up to several hundreds of milliseconds), and the long auditory store to those at a slower time scale (up to several seconds). The two sensory stores have been related to different phenomena that suggest that the temporal processing mechanisms in both stores differ (Cowan, 1988; Massaro, 1972). For example, the short auditory store is associated with perceptual integration, including phenomena such as loudness summation (Zwislocki, 1969) and backward masking (Massaro, 1975). Whereas the short auditory store holds representations of sound features that are integrated over time, the long auditory store is argued to represent information about the temporal order of sound segments. In an extension of the classical approach, a "predictive processing" approach of ASM provides a description of how temporal structure is accessed (Schröger, 1997; Schröger et al., 2013; Winkler and Czigler, 2012; Winkler, 2007; Winkler et al., 2009). Some form of "chunking" of single element representations that keeps the temporal order of elements, as already proposed by the classical approach, is now explained in more detail.[1] The predictive approach emphasizes the role of regularity extraction and predictive processing in ASM ("regularity-violation" explanation). Regularities are extracted from the auditory stimulation and represented in ASM based on representations of the relations of successive sounds (Schröger, 1997). Regularity representations are used to predict the incoming auditory stimulation. Incoming stimulations that match the prediction are integrated into a predictive model of a tone sequence or an otherwise temporally evolving auditory object

[1] Note that Schröger and colleagues avoid the term "auditory sensory memory". Instead they propose a so-called "Auditory Event Representation System (AERS)" and use the term "auditory stimulus event representations" that corresponds to the classical term "auditory sensory memory representation" (Schröger et al., 2013).