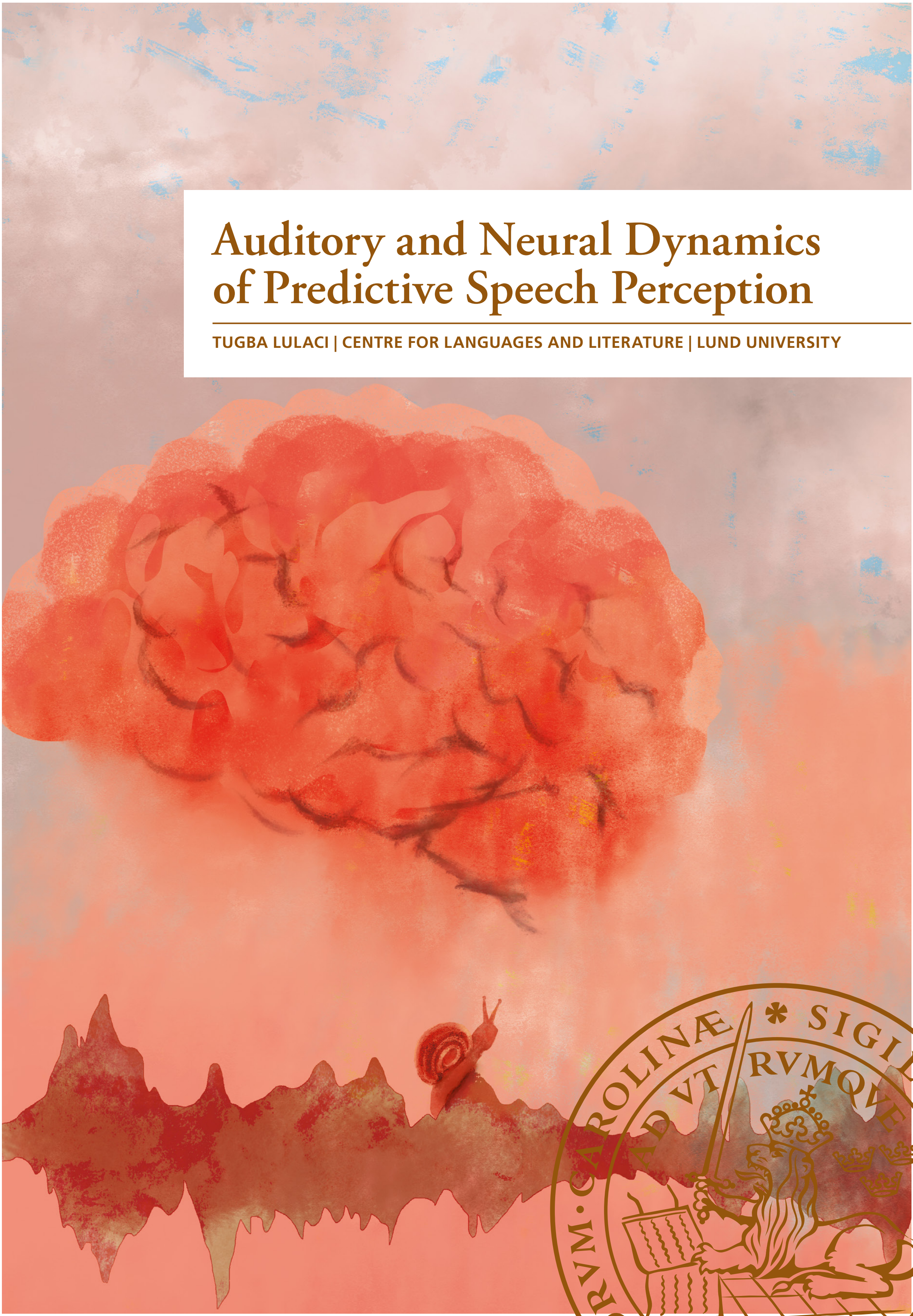
Auditory and Neural Dynamics of Predictive Speech Perception

Lulaci, Tugba

2026

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Auditory and Neural Dynamics of Predictive Speech Perception

TUGBA LULACI | CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY

Auditory and Neural Dynamics of Predictive Speech Perception

Auditory and Neural Dynamics of Predictive Speech Perception

Tugba Lulaci

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Joint Faculties of Humanities and Theology at Lund University to be publicly defended on the 27th of March 2026 at 10.00 in LUX:C121.

Faculty opponent

Dr. Matthew H. Davis

Medical Research Council (MRC) Cognition and Brain Sciences Unit (CBU)
University of Cambridge

Organization: LUND UNIVERSITY

Document name: DOCTORAL DISSERTATION

Date of issue 2026-03-27

Author(s): Tugba Lulaci

Sponsoring organization: -

Title and subtitle:

Auditory and Neural Dynamics of Predictive Speech Perception

Abstract:

In everyday listening, speech perception occurs under conditions where acoustic information may be incomplete or ambiguous, including background noise and variability in signal quality. Listeners must navigate ambiguous, masked or rapidly unfolding speech signal in order to comprehend spoken language. While prediction has been widely discussed and acknowledged in speech perception, less is known about how listeners predict upcoming information when acoustic cues are limited or ambiguous particularly at the early points of the speech signal. This thesis investigated how listeners anticipate upcoming sounds and update predictions during speech perception, focusing on the fine-grained acoustic cues in the rapidly unfolding speech signal. It also explored how these processes are influenced by individual extended high-frequency hearing thresholds, background noise and spectrotemporal dynamics in the signal. By combining behavioral tasks, electroencephalography and audiological assessments, the thesis traced speech processing from acoustic detail to auditory perception and neural activity in the cortex. Taken together, across studies, the findings suggested that listeners used fine-grained acoustic cues to anticipate upcoming speech sounds and that the differences in these predictions were affected by what listeners were able to access from the signal. By integrating behavioral, audiological, and neural evidence across paradigms, this thesis offers insights into how speech perception unfolds dynamically over time through the interaction of early acoustic information and listeners' expectations and individual differences in hearing.

Key words: speech perception; cognitive neuroscience; hearing; auditory system; event-related potentials; neurophysiology; phonetic cues; acoustics; spoken word recognition; prediction; EEG; coarticulation

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

Number of pages: 144

ISSN and key title:

ISBN:

978-91-90055-70-0 (print)

978-91-90055-71-7 (pdf)

Recipient's notes

Price

Security classification

Signature

Date 2026-02-05

Auditory and Neural Dynamics of Predictive Speech Perception

Tugba Lulaci

LUND
UNIVERSITY

Cover art by Cansu Damat

The Joint Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN: **978-91-90055-70-0 (print)**
ISBN: **978-91-90055-71-7 (pdf)**

Printed in Sweden by Media-Tryck, Lund University
Lund 2026

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

Dedicated to my mother, father and brother

Table of Contents

Acknowledgements

I would like to thank my supervisors Mikael Roll and Pelle Söderström for their support and mentorship throughout this journey. Without their unwavering support and guidance this thesis would never have been written. I am grateful for their feedback and scientific discussions throughout the years that shaped this thesis and helped me grow as a researcher. I am deeply grateful for my main supervisor Mikael Roll for his kindness, support and encouragement throughout the years. I am also grateful to my co-supervisor Pelle Söderström for his kindness and steady support throughout my PhD journey. I could not have asked for better supervisors; they guided me in science and showed me what supportive mentorship looks like. I would also like to express my sincere gratitude to Merle Horne for her kindness and encouragement. Her support and presence meant more to me than I can express.

A special thank you to Johan Frid for the valuable feedback and insightful discussion on an earlier version of this thesis. I would like to thank the Lund Neurolinguistics Lab and my lab mates and friends Renata, Claudia, Anna, Jinhee, Sabine and Pei-Ju for valuable discussions, support and all the fun we shared over the years. I have been very fortunate to be a part of a lab full of amazing researchers. Thank you for the good times and moments we shared. I would also like to thank Arthur Holmer for his support and kindness and the colleagues and collaborators at the Centre for Languages and Literature, and a special thank you to Kajsa and Filip for making this journey more supportive and enjoyable. I also thank all the participants who took part in the studies; their time and effort made this research possible. My sincere thanks to Burcu Ilkay Karaman for her support and encouragement throughout my academic journey.

My family has been the biggest cheerleader throughout my journey; this thesis is as much yours as it is mine. I am grateful to my mother and my father who supported me unconditionally, listened to me, encouraged me and always believed in me. I thank my brother for teaching me to be curious about science and always supporting me. I will always be grateful for your support; I could not have done this without you.

I also thank my friends Cansu and Civan for their friendship and support over the years. Special thanks to Cansu for creating the artwork for the cover of this thesis. It is difficult to put into words how much this meant to me.

Alexandre, I could not have done this without your immense support. Thank you for being by my side throughout this journey. Thank you for always believing in me and encouraging me and thank you for being always there. Your constant support helped me keep going.

Finally, thank you, Lumen, for being the light during the darkest nights.

In the end, beyond all struggles, what remains are the voices, kindness and memories we leave behind. I am deeply grateful for all the kind people I have met during my PhD journey.

Abstract

In everyday listening, speech perception occurs under conditions where acoustic information may be incomplete or ambiguous, including background noise and variability in signal quality. Listeners must navigate ambiguous, masked or rapidly unfolding speech signal in order to comprehend spoken language. While prediction has been widely discussed and acknowledged in speech perception, less is known about how listeners predict upcoming information when acoustic cues are limited or ambiguous particularly at the early points of the speech signal. This thesis investigated how listeners anticipate upcoming sounds and update predictions during speech perception, focusing on the fine-grained acoustic cues in the rapidly unfolding speech signal. It also explored how these processes are influenced by individual extended high-frequency hearing thresholds, background noise and spectrotemporal dynamics in the signal. By combining behavioral tasks, electroencephalography and audiological assessments, the thesis traced speech processing from acoustic detail to auditory perception and neural activity in the cortex. Taken together, across studies, the findings suggested that listeners used fine-grained acoustic cues to anticipate upcoming speech sounds and that the differences in these predictions were affected by what listeners were able to access from the signal. By integrating behavioral, audiological, and neural evidence across paradigms, this thesis offers insights into how speech perception unfolds dynamically over time through the interaction of early acoustic information and listeners' expectations and individual differences in hearing.

Abbreviations

2AFC	Two-alternative forced choice
ABR	Auditory brainstem response
ASA	Auditory scene analysis
CoG	Centre of gravity
CVC	Consonant-vowel-consonant
dB	Decibel
dB HL	Decibel hearing level
dB SPL	Decibel sound pressure level
ECoG	Electrocorticography
EEG	Electroencephalography
EHF	Extended high-frequency
ERP	Event-related potentials
FFR	Frequency-following response
Hz	Hertz
IC	Inferior colliculus
kHz	Kilohertz
MMN	Mismatch negativity
ms	Millisecond
NP	Noun phrase
PMN	Phonological mapping negativity
PP	Prepositional phrase
PrAN	Pre-activation negativity
PTA	Pure tone average
RMS	Root mean square
RT	Response time
SNR	Signal-to-noise ratio
STG	Superior temporal gyrus
µV	Microvolt

1.Introduction

Speech perception involves multiple levels of processing in the human brain. It begins with the ear capturing fine-grained acoustic cues and ends with the brain making sense of it as meaningful pieces of information. This process often happens so swiftly that we hardly notice the complexity. Speech is an acoustically and informatively rich signal composed of adjacent phonemes. It is rapid, dynamic and coarticulated, and every sound carries traces from neighboring sounds. Speech perception is, therefore, a complex yet fascinating process in which auditory and cognitive systems work together to make sense of a dynamic and often ambiguous or masked auditory signal. Further, speech typically occurs in conditions of background noise, competing speech, and the need to extract relevant information and respond in real time. Despite the variability of the speech signal and the frequent presence of background noise and ambiguity, however, the human brain can process speech rapidly and accurately. In everyday communication, this complex processing happens rapidly, while the comprehension of speech often requires resolving ambiguity; for instance, the signal clarity can be affected by background noise or reduced hearing sensitivity, increasing the need to process early fine-grained acoustic cues available at word onset to sustain rapid perception. In these listening situations, people can still usually successfully comprehend what is said. This could be explained by listeners actively anticipating what comes next rather than passively waiting.

The brain is proposed to work predictively, continuously generating predictions about upcoming input to minimize surprise and updating predictions when the input does not match the expectation (Friston et al., 2012). Considering the rapid nature of speech and frequent adverse listening conditions, successfully processing speech requires anticipating upcoming speech sounds, or “filling in the gaps.” This would be one way to explain the brain’s resilience in perceiving speech even under adverse listening conditions. Studying how the human brain manages these conditions can reveal how speech perception operates under suboptimal conditions and how it adapts when speech input is partial, ambiguous, or overlapped with noise.

This thesis investigates how listeners perceive and process speech predictively under varying conditions, using behavioral, audiological, and electrophysiological measures. A central focus is the time course of how rapidly unfolding acoustic cues affect spoken-word recognition and support predictive processing. Across four studies, it assesses speech perception from several perspectives: the predictive use of fine-grained acoustic cues at word onset, the influence of hearing sensitivity on these predictions, how prosodic cues in noise can assist predictive speech perception, and the neural responses when these predictions are violated.

1.1. Aim and research questions

The overall aim of this thesis is to assess the temporal dynamics of how listeners use fine-grained acoustic information to anticipate upcoming speech and guide spoken-word recognition as the signal unfolds. The thesis also examines how this process is shaped by acoustic signal properties and individual differences in hearing profiles. The research question is thus how different constraints and listening settings shape listeners' ability to anticipate speech using rapidly unfolding acoustic cues in real time. Speech perception depends on both acoustic encoding and linguistic processes. Therefore, the studies in the thesis manipulate different types of constraints to investigate how listeners use partial acoustic information for spoken-word recognition as the signal unfolds. In the present thesis, speech perception is investigated through the predictive use of specific acoustic cues one of which is coarticulation. Speech consists of overlapping sounds (Kent & Minifie, 1977). For instance, word-onset speech sounds already carry acoustic traces from upcoming sounds. Another cue investigated here is prosody, where suprasegmental cues guide speech perception beyond the segmental level. In Swedish, word tones known as "word accents" signal the upcoming suffix, giving clues to the listener about how the word will unfold (Roll et al., 2015; Söderström, Horne, & Roll, 2017).

Predictive speech perception was tested when the signal was only partially available (gating), when individual hearing sensitivity affected the access to subtle acoustic details (extended high-frequencies), when the speech signal was degraded by speech-shaped noise (prosodic cues masked by noise), and when the early cues were misleading (fine-grained coarticulatory cues mismatched using cross-splicing). Together, these studies provide an ear-to-cortex view of prediction, linking the accessibility of fine-grained acoustic cues, hearing acuity, and neural correlates of predictive speech perception.

Study	**Focus**
Study 1	Behavioral investigation of how listeners rapidly use fine-grained acoustic information during spoken-word recognition, using a gating paradigm.
Study 2	Combined audiological assessment and behavioral responses to fine-grained acoustic cues, focusing on individual differences in standard (0.25–8 kHz) and extended high-frequency (up to 16 kHz) hearing thresholds.
Study 3	EEG study of speech perception in noise, using prosodic patterns as cues, speech-shaped noise masking, and cross-splicing manipulations to assess neural correlates of predictive speech processing in challenging listening conditions.
Study 4	EEG study employing a cross-splicing paradigm to investigate neural responses to fine-grained acoustic congruency and the temporal dynamics of anticipatory and onset-following dynamics.

Table 1. Studies included in the thesis and the individual focus of each study

Four central questions guided this thesis. First, the thesis asks how early in the spoken-word signal predictive information becomes available, investigating whether the first few milliseconds of word onsets carry enough acoustic cues to facilitate word recognition for words in isolation. The second question was whether individual differences in hearing sensitivity, particularly extended high-frequency hearing thresholds, support access to fine-grained acoustic cues and affect the predictive use of these cues. Third, it was investigated whether prosodic cues continue to serve as reliable predictive cues when the speech signal is degraded by noise masking and how the brain adapts when prediction needs to rely on prosody under adverse listening conditions. Finally, it was scrutinized how rapidly the brain extracts and evaluates fine-grained acoustic cues and how predictive updating is reflected in electrophysiological responses when those cues do not align with the upcoming information within the unfolding signal. At a broader level, these questions asked how the brain generates and updates predictions when the speech signal is uncertain and which acoustic and auditory factors affect this process as the signal unfolds and becomes ambiguous. Finally, this thesis examined how the brain operates under conditions of uncertainty in the speech signal. The following chapters address these questions through four studies.

The four studies in this thesis address these mechanisms from different perspectives, combining audiological assessment, behavioral evidence and neural electroencephalographic (EEG) measures across studies to investigate how hearing sensitivity, fine-grained acoustic cues, and prosodic cues support prediction and how the brain responds when expectations are violated.

2.Background

2.1. Auditory processing

The human auditory system is capable of processing a range of frequencies from 20 to 20,000 Hz (Schnupp et al., 2010). The sound signal travels from the cochlea to the auditory cortex (Söderström, 2024). The cochlea converts the vibrations to neural signals, preserving spectral details. These signals travel through the cochlea, and on to subcortical and cortical stages of the auditory system, where frequency, timing, and intensity-based information is further organized and processed (Zatorre, 2024). Speech is a complex signal, and in everyday listening, it rarely occurs in isolation without the presence of accompanying background noise. Therefore, before speech can be recognized or mapped onto meaning, the auditory system needs to continuously separate overlapping acoustic signals to identify the relevant acoustic stream. Auditory scene analysis (ASA) (Bregman, 1990) describes how the auditory system can segregate auditory objects in complex acoustic environments by grouping sounds that share temporal and spectral characteristics. Beyond perceptual grouping, the auditory cortex supports complex representations of sound. Insights from primate neurophysiology and human neuroimaging highlight hierarchical cortical pathways that represent auditory objects, spatial cues, and speech-relevant acoustic patterns (Scott, 2005). These maps and streams are thought to provide the neural foundation for human speech perception (Rauschecker & Scott, 2009).

2.2. Speech perception

Speech perception lies at the intersection of multiple scientific disciplines that together examine the pathway from ear to cognition. It connects the peripheral auditory system, which encodes acoustic information, with the neural and cognitive mechanisms that transform these continuously changing acoustic signals into meaningful linguistic pieces. This complex yet rapid process has been studied over several decades from multiple different perspectives. To this date, several theoretical frameworks have been developed to explain how the auditory system and cognitive mechanisms work together to accomplish the transformation from sound to meaning.

Hypotheses about the human brain's ability to perceive speech despite the complexity of the signal have been shaped by debates over if the speech is processed through mechanisms unique to language or through general auditory and cognitive principles. Early frameworks proposed different mechanisms to explain how listeners achieve fast and accurate speech processing. One influential perspective in speech perception was the Motor Theory of speech perception (Liberman et al., 1967; Liberman & Mattingly, 1985). It proposed that listeners perceive speech with the help of internal simulation of articulatory gestures and defined speech processing as "special" compared to other cognitive processes. Although a later review by Galantucci et al. (2006) re-evaluated the Motor Theory within a broader ecological framework, they weakened its central claim by arguing that while perception and production could be linked by shared systems, speech perception may not be a special process.

Episodic-based views of lexical access have proposed that the perceptual details of spoken words may be retained in memory and influence subsequent speech perception (Goldinger, 1998). Similarly, exemplar-based approaches propose that, for example, a phoneme is represented as distributions of stored tokens that preserve fine-grained acoustic/phonetic variation across experiences (Pierrehumbert, 2001).

Beyond questions of representational views, a large body of work has focused on the dynamics of lexical activation. They have examined how the spoken words are activated and compete with lexical candidates over time. Neighborhood activation models discuss activation as the speech signal unfolds with recognition speed and accuracy influenced by lexical frequency and lexical competition (Luce & Pisoni, 1998; Goldinger et al., 1989).

Classic models of spoken-word recognition, such as the Cohort and TRACE models, explained interactive processing in terms of partial acoustic inputs activating word candidates, and higher-level lexical context feeding back to assist perception (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986). Shortlist (connectionist) model (Norris, 1994; Norris et al., 1997) and Shortlist B (Bayesian) (Norris & McQueen, 2008) treat spoken word recognition as a process of competition among lexical candidates that is updated dynamically as the speech signal unfolds. Together these models highlight the incremental processing in speech perception as early acoustic information shaping expectations about upcoming information.

2.3. Anticipatory neural processing in auditory and language perception

One of the most prominent frameworks to explain the perception mechanism of the human brain is predictive coding (Clark, 2013; Friston et al., 2012; Friston, 2005; Rao & Ballard, 1999). Instead of passively waiting, the brain is considered to continuously generate expectations about upcoming information, actively comparing it to the sensory input, which reflects a neural response called prediction error, and leads to an update of the inner predictive model (Garrido et al., 2009; Friston et al., 2021). Thus, within the predictive coding framework, perception arises from continuous interplay between top-down and bottom-up signals (Friston & Kiebel, 2009).

At the sensory level, the auditory system shows sensitivity to patterns, and these can be tracked using electroencephalography (EEG). The N100, a negative-going evoked or event-related potential (ERP) following the onset of auditory stimulus peaking approximately around ~100 ms, has been reported as reflecting early auditory processing as it is sensitive to acoustic properties of the stimulus (Näätänen & Picton, 1987). Beyond responding to irregularities, N100 has also been linked to predictive processing and its amplitude is modulated by the predictability and informational value of upcoming sounds (Schröger et al., 2015) as well as by attention (Astheimer & Sanders, 2009, 2011).

The mismatch negativity (MMN) is elicited as a response to irregular, unexpected auditory stimuli (Näätänen, 2000). It has also been proposed as a neural index of prediction error, reflecting model updating (Friston, 2005; Wacongne et al., 2012). Importantly, such mechanisms are not limited to tone frequency changes; MMN was also reported for infrequent syllables and speech sounds (Näätänen et al., 2007).

Beyond scalp-recorded ERPs, Parras et al. (2017) investigated prediction error from single-neuron activity at various stages of the auditory pathway in rodents and observed mismatch responses consistent with prediction error already at subcortical levels and increasing toward the auditory cortex. These findings suggest the auditory system is sensitively tracking patterns and flagging changes.

Considering that language comprehension involves rapidly integrating complex and hierarchically structured information, it is a complex yet remarkably fast process extending basic sensory prediction. However, the human brain is still able to decode linguistic input within fractions of a second. Therefore, to understand the process and its aiding structures, it has been studied over decades (Kutas & Federmeier, 2011; Van Petten & Luka, 2012). Increasing recent evidence suggests that the brain is operating hierarchically during spoken language comprehension, with cortical regions representing information at distinct levels and time scales (Schmitt et al., 2021; de Heer et al., 2017; Heilbron et al., 2022).

Studies at the sentence level of speech perception have shown how a word is processed when it is supported by the sentence context (Brothers & Kuperberg, 2021; Staub, 2015). When a sentence starts unfolding, the brain activates possible continuations anticipatorily (Levy, 2008).

Event-related potentials have a long history in the study of language processing to map the temporal progression from early auditory encoding to semantic access (Federmeier et al., 2016). One of the widely researched ERP components is the N400, a negativity peaking approximately at 400 ms following stimulus onset. It was first reported by Kutas and Hillyard (1980) and discussed as the brain's reaction to semantic incongruency. It has been associated with semantic processing and with the effort required to integrate linguistic information into context (Lau et al., 2008). While N400 has been discussed widely regarding its function to index lexical access, several studies (MacGregor et al., 2012; Gosselke Berthelsen et al., 2020) have found that the brain is sensitive to lexicality already at around ~50 ms after disambiguation between words and pseudowords.

Similar to N400 studies, using linguistic violations, syntactic encoding processes have also been researched in language studies. The P600 was originally reported by Osterhout and Holcomb (1992) as a positive ERP shift evoked by syntactic anomalies, peaking around 600 ms after the anomaly. These findings were later extended suggesting the P600 is not limited to syntactic anomalies but is also observed when listeners attempt to repair thematic-role violations (Kuperberg et al., 2003). Late positivities are often argued to form part of the P300 component family (Polich, 2007). This positive-going evoked potential was first reported by Sutton et al. (1965) elicited by uncertainty and named P300. Later, it was considered as related to context updating and working memory (Donchin, 1981). Strauss et al. (2015) found the P300 was present during wakefulness but disappeared during sleep as a prediction-error response, while early mismatch responses (i.e., N1, N2) remained detectable across sleep stages. These results together suggest that the P300 reflects a response to uncertain events that require additional evaluative processing, and context updating (Donchin & Coles, 1988). In predictive coding, it can be thought to index model updating (Roll et al., 2023).

Earlier mismatch-based processes can be observed in the phonological mismatch/mapping negativity (PMN), which is a negative-going ERP component elicited by unexpected phonemes, reported by Connolly and Phillips (1994) to reflect pre-lexical processing (Connolly et al., 2001; Newman et al., 2003; Archibald & Joanisse, 2011; Connolly et al., 1992). The PMN is considered not to be tied to semantics as it has also been elicited in non-word stimuli (Newman & Connolly, 2009).

The pre-activation negativity (PrAN) is an ERP component indexing the predictive strength of anticipatory cues. It is reported as a negativity following the cue onset between 136-200 ms, indexing how the listener can anticipate the upcoming information (Söderström et al., 2016; Roll, 2022; Roll et al., 2023; Roll et al., 2015). PrAN has been proposed to reflect feed-forward prediction processing, unlike components signaling prediction error (i.e. MMN and N100) and has been dissociated from the P200 (Roll et al., 2013; Roll et al., 2023). Studies using Swedish word accents have shown that accent 1 evokes a larger PrAN compared to accent 2. This increased negativity is reported as the predictive potential of the cues. Thus, PrAN provides a window into how prosodic cues (word accents) can support prediction at early stages of processing.

Taken together, these ERP components illustrate how anticipatory processes in language operate at multiple hierarchical levels. From the earliest auditory responses to later components associated with phonological mapping, semantic access, pre-activation, and syntactic integration, ERP components collectively characterize the complex hierarchical nature of speech processing.

Considering the earliest stages of this hierarchy for speech perception rely on information encoded in the acoustic signal, the next section focuses on the acoustic cues including segmental and suprasegmental cues that provide evidence for phonological and lexical expectations.

2.4. Perceptual and neural processing of acoustic cues

Currently, there are several theoretical frameworks and empirical approaches investigating the acoustic characteristics of the speech signal to understand how listeners extract linguistic information from a dynamic, highly variable sound signal (Diehl et al., 2004; Holt & Peelle, 2022).

Given that the speech signal unfolds over time and the brain is considered to operate predictively, acoustic properties also contribute to this process. A large body of research showed that speech perception benefits from both segmental and suprasegmental cues (Roll et al., 2017; Söderström & Cutler, 2023; Roll, 2022; Gosselke Berthelsen et al., 2018; Roll et al., 2013; Söderström et al., 2012; Roll et al., 2011; Roll & Horne, 2011; Roll et al., 2010; Hjortdal et al., 2022; Roll et al., 2023; Kwon & Roll, 2024). Neural responses during spoken word recognition have been associated with graded lexical activation as the word unfolds (Söderström & Cutler, 2023).

Listeners can often identify spoken words before hearing the entire acoustic signal (Grosjean, 1980). The speech signal comprises multiple overlapping speech sounds. Coarticulation results in an acoustic signature where a segment is shaped by its surrounding phonetic context (Kent & Minifie, 1977; Fowler, 2005) and lexical processing is influenced by coarticulatory cues (Dahan et al., 2001; Marslen-Wilson

& Warren, 1994). Spectrotemporal dynamics in the signal due to coarticulation introduce variability, but this variability is structured rather than random. It has been shown to convey useful information that listeners can use as a cue to predict the upcoming sounds using graded acoustic patterns (Beddor et al., 2013; Bell-Berti & Harris, 1979). What listeners use behaviorally in these paradigms is consistent with how the auditory cortex represents speech through detailed spectrotemporal structure. Neural evidence from both invasive and non-invasive studies has suggested that the auditory cortex represents speech with detailed clarity and tracks the acoustic signal preserving the fine-grained acoustic patterns, allowing phonetic differences to be distinguished. Intracranial recordings in Heschl's gyrus (HG) for instance, suggest that, at the earliest auditory stages, the brain encodes detailed acoustic information (Khalighinejad et al., 2021). Using ECoG, electrodes have been shown to respond selectively to acoustic patterns encoded in line with articulatory phonetic distinctions emerging from population-level tuning (Mesgarani et al., 2014). These findings suggest that the primary and non-primary auditory cortex represents detailed acoustic features of speech in real time. Using non-invasive fast optical imaging, Toscano et al. (2018) reported that the posterior superior temporal gyrus (pSTG) and planum temporale remained sensitive to continuous acoustic cues during approximately the first 200 ms of processing. Complementing these findings, a magnetoencephalography (MEG) study suggested that the auditory cortex shows sensitivity to phonological ambiguity as early as 50 ms, while preserving subphonemic information over long timescales and re-evoking this information when later lexical input becomes available (Gwilliams et al., 2018).

Across several methods and studies, one consistent observation is that the auditory system preserves and uses fine-grained acoustic detail. Behavioral and electrophysiological studies have shown that coarticulation and other subphonemic cues guide listeners' expectations. Activity in auditory cortex has been reported as encoding such distinctions through detailed spectrotemporal patterns. Non-invasive evidence also echoes this sensitivity to continuous acoustic structure. Importantly, these details are not only decoded retrospectively, but they also support anticipatory perception. In line with the predictive use of acoustic cues, auditory regions such as primary auditory cortex provides the sensory evidence that feeds into higher levels of the hierarchy (Garrido et al., 2007). Higher-level predictions generate expectations about upcoming speech sounds, and activity in the STG can reflect the difference between expected and heard input when these prior expectations are violated (Gagnepain et al., 2012). Thus, segmental acoustic cues not only provide the blocks of phonetic structure but also form the basis for rapid predictions about how the speech signal will unfold.

Beyond segmental detail, speech also carries suprasegmental cues. Prosodic patterns are a linguistically meaningful dimension of speech and can be used predictively as a cue to generate expectations about upcoming information during speech perception (Roll, 2022). Swedish has a feature known as *word accent*, where pitch patterns (Accent 1 and Accent 2) are associated with the stressed syllable (Elert, 1964). In Central Swedish, Accent 1 is realized as a low tone (L*) and Accent 2, as a high tone (H*) (Bruce, 1987). Word accents are linked to upcoming morphological structure (Bruce, 1977; Riad, 2012). For instance, a word stem carrying accent 1 is a cue for an upcoming definite singular *-en* suffix, while accent 2 cues the plural suffix *-ar*. Thus, word accents provide an early prosodic cue that listeners can use to predict how the word is unfolding (Roll et al., 2010; Roll et al., 2017; Roll et al., 2013; Roll et al., 2015; Söderström et al., 2016; Gosselke Berthelsen et al., 2018; Roll & Horne, 2011; Roll et al., 2009; Söderström et al., 2012; Söderström, Horne, & Roll, 2017; Söderström et al., 2023).

When the suffix was cued by a word accent that did not match the upcoming signal, listeners showed increased response times (Söderström et al., 2012; Roll et al., 2017). Studies using Swedish word accents have shown that violations of generated expectations (i.e. hearing another suffix than what the word stem prosody suggested) elicited P600 (Novén, 2021; Roll et al., 2013; Roll et al., 2015) and N400 effects (Söderström, Horne, Mannfolk, et al., 2017; Novén, 2021; Roll, 2015). The responses mirrored prediction error signatures. The results of these studies demonstrate that prosodic cues such as word accents function as predictive cues, shaping early pre-activation and driving prediction error responses when the speech signal unfolds differently than expected based on stem tone and other expectations.

In summary, segmental and suprasegmental (including coarticulation and word accents) evidence converges in the view that listeners make use of the acoustic patterns of speech to anticipate what is coming next. Speech perception benefits from anticipatory cues and the speech signal contains coarticulatory information, fine-grained acoustic details and prosodic patterns jointly supporting real-time prediction. The auditory system preserves spectrotemporal detail at multiple levels, as evidenced by several studies reporting the brain encoding of acoustic patterns, sensitivity to coarticulatory cues, and rapid cortical responses to prosodic cues, such as the pre-activation negativity as well as mismatch effects. These studies show that the brain integrates acoustic cues across various timescales to generate and update predictions during the unfolding speech signal.

2.5. Acoustics, environment and listener-based constraints in perception

Spoken language unfolds rapidly and continuously, and it contains several spectral and prosodic patterns. This complexity requires the listener to integrate information across multiple spectral and temporal levels. In everyday communication, this dynamic signal varies across contexts. Speech sounds also show considerable acoustic variability. For instance, while some fricatives have relatively flat spectral energy (such as /f/), others, like /s/, present sharper spectral peaks (Jongman et al., 2000) and traces from surrounding segments (Perkell & Matthies, 1992). This uneven pattern suggests that the perceptual system must flexibly interpret the signal as it unfolds over time.

Speech perception typically occurs under acoustically challenging environments and consists of overlapping sounds and background noise (Cherry, 1953). The acoustic signal is affected by a set of environmental and listener-based constraints that limit how clearly speech can be encoded and how efficiently the signal can be parsed. These constraints reduce the quality of incoming information, increase processing demands, and influence the perceptual strategies listeners use to maintain comprehension (Van Hedger & Johnsrude, 2022).

The acoustic environment places possible constraints on perception by affecting how much bottom-up detail is available to listeners. Noise, distortion and poor signal-to-noise ratios affect the accessibility of fine-grained spectrotemporal and prosodic cues. However, not all cues are affected by noise the same way (see Mattys et al. (2012) for a review). For instance, prosodic cues have been shown to remain relatively robust under noise, allowing listeners to track durational patterns to locate word boundaries when acoustic detail is masked (Smith et al., 1989).

Work on acoustically distorted speech shows that when the signal becomes less clear and intelligibility declined, cortical activity varies with these conditions (Davis & Johnsrude, 2003). Similarly, when the acoustic clarity is reduced listeners increasingly rely on higher-order cortical regions. For instance, spectral degradation has been found to strengthen functional connectivity across cortical regions with semantic predictability facilitating comprehension under intermediate levels of noise (Obleser et al., 2007). In multi-talker listening contexts, hearing-aid noise reduction has been shown to modulate early and late cortical representations of competing talkers in noise, suggesting that changes in acoustic enhancement may influence hierarchical speech representations (Alickovic et al., 2021). More generally, hearing aid noise reduction can enhance neural representations of speech while reducing representations of background noise (Alickovic et al., 2020).

Taken together, under challenging listening conditions the brain does not give up, but instead continuously tries to decode and integrate the accessible information from the signal. However, the usefulness of cues and coping mechanisms depend not only on their spectral- or context-related informativeness but also on whether the listener is able to perceive and register them. Individuals differ in how they cope with speech perception under adverse conditions, and these differences influence the outcome of perceptual process. Under adverse listening conditions, when the input is challenging or ambiguous, listeners may rely more on their sensory and cognitive resources to follow the conversation.

Several listener-based factors have been linked to individual differences and have been shown to influence speech perception, including working memory capacity

(Rönnberg et al., 2013), cognitive resources (Herrmann & Johnsrude, 2020), and attention (Pichora-Fuller et al., 2016). In adverse listening conditions, performance is further shaped by the interaction between perceptual abilities, cognitive abilities and linguistic processes (Mattys et al., 2025). Hearing sensitivity is one of these fundamental listener-based constraints on perception. Standard hearing assessments typically measure thresholds up to 8 kHz and classify listeners as "normal hearing" if the thresholds fall within clinically normal limits (≤ 20 dB HL). However, recent studies suggested that sensitivity in the extended high-frequency (EHF) range also influences speech perception (Hunter et al., 2020). Elevated EHF thresholds have been associated with poorer speech in noise performance even in situations where standard audiograms are normal (Motlagh Zadeh et al., 2019; Yeend et al., 2019). Moreover, EHF sensitivity has been linked to benefits in spatial and cocktail party listening, talker orientation detection and in naturalistic listening conditions (Badri et al., 2011; Trine & Monson, 2020; Monson et al., 2019).

Taken together, these acoustic, environmental and listener-based constraints show that speech perception is shaped by the quality of the signal as well as the listener's ability to access and interpret it. The perceptual system must work with whatever information is available at a given moment, and this availability may vary across listening conditions and across individuals. Understanding the nature of these constraints, therefore, requires considering both the presence of acoustic detail and whether listeners are able to detect and make use of it.

3. Methods

This thesis employed a combination of behavioral paradigms, hearing assessments, and EEG recordings to investigate how listeners perceive and use acoustic information predictively for speech perception across different listening conditions. Across the studies, all participants were native Central Swedish speakers and had normal hearing based on pure-tone audiometry. Detailed sample characteristics of each study are reported in the original articles. Here, the methodological approaches are summarized.

3.1. Stimuli

In Study 1 and Study 2, the same set of stimuli was used. Twenty Swedish words with /s/ and twenty with /f/ onsets were selected to be used in a gating paradigm. The stimuli were recorded using a U47 FET microphone in a soundproof room. Using a RME Fireface UCX II sound interface and Universal Audio 6176 preamplifier, the recordings were made at 44.1 kHz sample rate. The speech materials were designed to capture the temporal dynamics of spoken word recognition and the cues accessible for prediction. To achieve this, an adapted gating paradigm was used. This paradigm presents listeners with a speech signal with incrementally longer portions of the words as “gates” (Grosjean, 1980). First, a Praat (Boersma & Weenink, 2021) script was used to determine the word onset based on a predefined intensity threshold. After this procedure, each word was divided into four gates using another Praat script. The first gate starting from the word onset was 15 ms, the second gate was 35 ms, third gate was 75 ms, and, finally, the fourth gate was 135 ms. The shortest gate was set to 15 ms starting from word onset to target early word-onset information.

Figure 1 illustrates the gates showing the sound waveform and spectrogram with gating boundaries marked in red. Acoustic analyses were also performed to test the differences between /s/ and /f/.

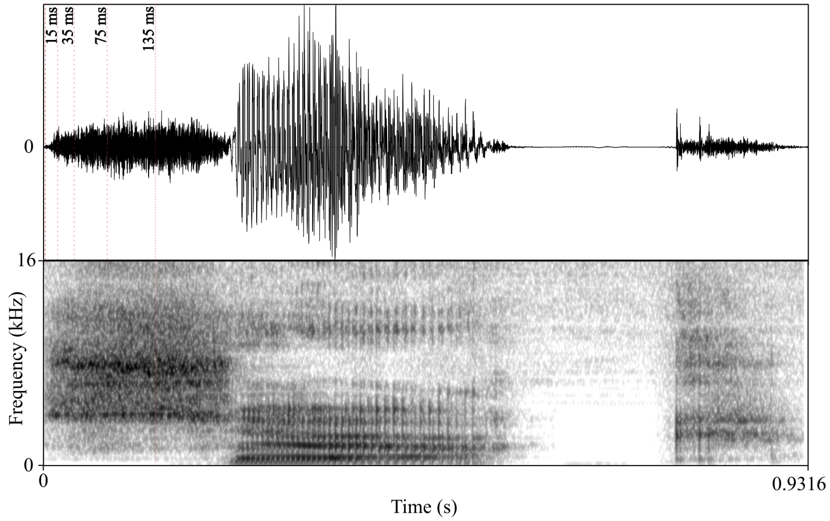


Figure 1. Example stimulus (word *sök)* from the gating paradigm. Sound waveform (top) and spectrogram (bottom), vertical lines marking the gate boundaries (e.g. 15 ms, 35 ms).

Center of gravity measures were extracted to analyze variance between and within fricatives (Wikse Barrow et al., 2022; Lulaci et al., 2022). The center of gravity of gates was computed using the built-in "centre of gravity" function in Praat. The effect of the following vowel on the onset fricative based on its articulatory features and the difference between /f/ and /s/ onsets were analyzed, confirming that /s/ and /f/ provided different useful acoustic cues at word onset. The band energy calculation function in Praat was used to calculate the high-frequency energy in onset fricatives. Detailed descriptions of the acoustic analyses are provided in the papers included in this thesis (Lulaci et al., 2025; Lulaci et al., 2024)

In Study 3, the auditory stimuli were identical to those used in Roll et al. (2015), consisting of Swedish sentences recorded by a native Central Swedish speaker (for details of the acoustic analyses see Roll et al. (2015)). All sentences ('NP got target word PP') followed the same structure, and all target words had one of the two Swedish word accents (accent 1 and accent 2). The following experimental conditions were created through cross-splicing: matched accent and suffix, and mismatched accent and suffix. To test the resilience of prosodic cues under challenging conditions, speech-shaped noise was created using an adapted Praat script from Winn (2024). Signal-to-noise-ratios (SNRs) were set to SNR 0 dB and

SNR -5 dB using Praat with root mean square (RMS) scaling. Figure 2 shows an example from the stimulus list.

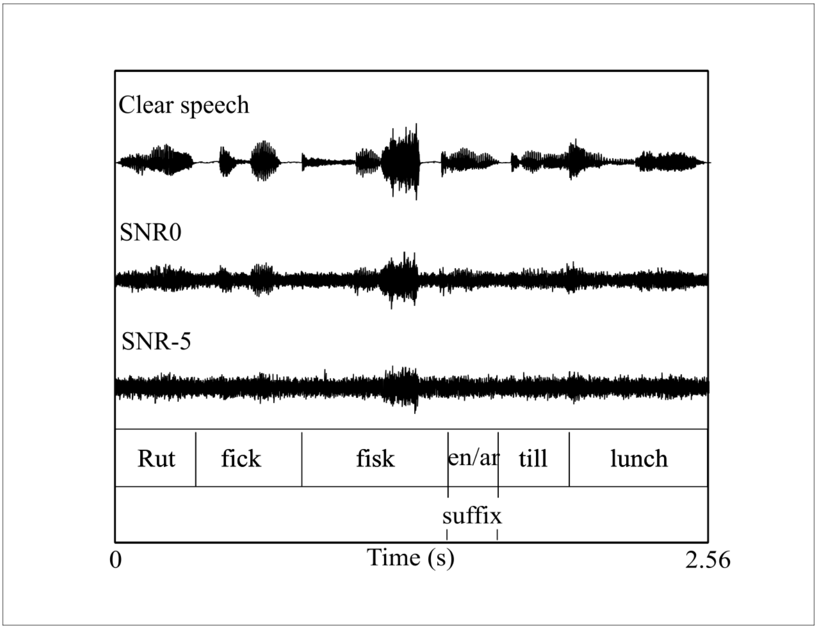


Figure 2. Sound waveform examples of the sentence stimuli at the three noise levels used in Study 3. The clear condition (top) shows the unmasked speech signal, while the 0 SNR dB and –5 SNR dB conditions illustrate the same sentence mixed with speech-shaped noise. These plots are provided for visualization.

Study 4 employed a cross-splicing paradigm with fricative-onset words. The words were selected from the list used for Study 1 and Study 2. Word onset fricatives were manually excised using Praat and spliced onto the continuation of another word. Boundaries were adjusted to the nearest zero crossing to avoid artifacts. This design allowed us to ensure that the stimuli sounded natural while introducing a mismatch between the word onset fricative and the following vowel + coda.

3.2. Hearing assessment

Hearing thresholds were tested for all participants as an inclusion criterion. Air-conduction pure tone hearing thresholds were assessed in 5 dB-sized steps using a Callisto audiometer, Interacoustics (IEC60645-1 2017/ANSI S3.6 2018). Radioear DD450 headphones were used to present pure tones at 0.25-8 kHz (ISO 389-8 2004, ANSI S3.6 2018) and EHF hearing thresholds at 9-16 kHz (ISO 389-5 2004, ANSI S3.6 2018). Pulsed tones were presented using the modified Hughson-Westlake method (Carhart & Jerger, 1959; Hughson & Westlake, 1944). All participants had normal hearing (250–8000 Hz ≤20 dB HL). In Study 2, which addresses extended high-frequency hearing thresholds, thresholds above 8 kHz were also tested for use in individual difference analyses.

3.3. Behavioral methods

Behavioral data were collected in all four studies. In studies 1 and 2, participants completed the gating paradigm described above. A two-alternative forced-choice (2AFC) paradigm was used; participants were asked which word they heard after listening to each gate, starting from the word onset. In studies 1 and 2, accuracy data were collected, and temporal dynamics were obtained from the gate information. Trial order was randomized and stimuli were presented in blocks. In studies 1 and 2, multiple unique recordings per word were used to reduce stimulus-specific repetition effects. In studies 3 and 4, accuracy, selection preference and response times were collected.

3.4. Electroencephalography

EEG is a non-invasive technique that measures the brain's electrical activity from the scalp using electrodes. The activity recorded comes from post-synaptic potentials of large populations of neurons, largely cortical pyramidal cells, and spans over tens to hundreds of milliseconds rather than action potentials of neurons (Luck, 2014). Neural activity was recorded in studies 3 and 4 to analyze the brain responses to matched and mismatched stimuli. EEG was recorded using 64 channel Easycap, and a SynAmps 2 amplifier through the NeuroScan Curry 7. Scalp electrodes' impedances were kept below 5kΩ and mastoids were kept below 1kΩ. EEG data were pre-processed using EEGLAB (Delorme & Makeig, 2004) in MATLAB. The data were re-referenced to the mastoid average.

After bandpass filtering at 0.05–30 Hz, the continuous EEG data were epoched starting from -200 to 800 ms relative to stimulus onset.

3.4.1. Event related potentials

Event-related potentials (ERPs) are voltage fluctuations in the EEG recorded at the scalp that are time-locked to specific events. ERPs provide a great opportunity to understand human sensory and cognitive processes with high temporal precision. The millisecond-level temporal resolution of ERPs makes them well-suited to research questions concerning rapid cues. ERP components are usually defined by their polarity and latency in the waveform, even though their latencies may vary across different experiments (Luck, 2014). An example is the “P600,” a denomination used for voltage deflection with a peak around 600 ms following structurally unexpected stimuli (Osterhout & Holcomb, 1992). ERP analyses focused on components associated with speech and language processing as outlined in Section 2.3. Time windows of interest were defined based on established ERP literature for each ERP component (e.g. N400, P300, PMN). Analyses were not restricted to predefined electrode sites; instead, effects were examined across the scalp with interpretation guided by typical spatial distributions reported for each component (see Figure 3 for channel locations).

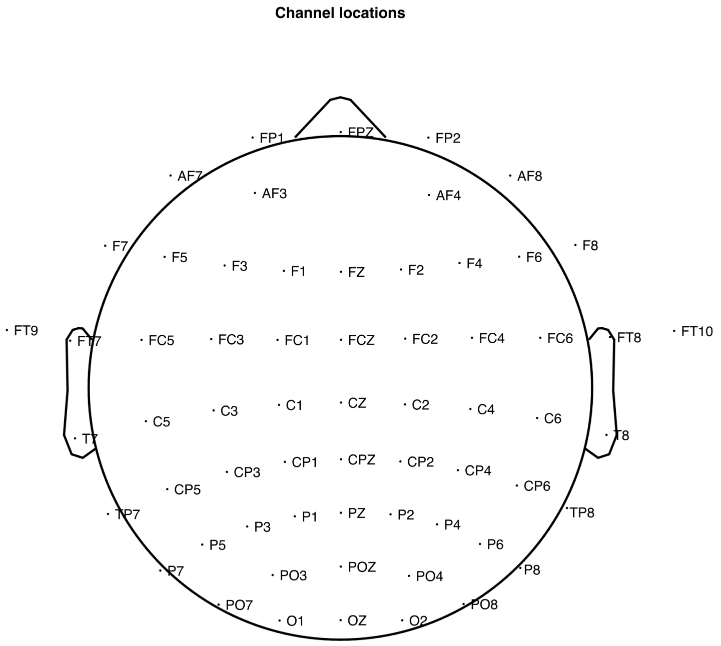


Figure 3. Electrode layout of the 64-channel EEG cap used in the experiments

3.5. Statistical analysis

3.5.1. Behavioral analysis

Across four studies, behavioral data were analyzed using linear mixed-effects models (LMMs) and generalized linear mixed models (GLMMs), allowing the estimation of experimental effects while accounting for variability, including random effects and fixed effects (Bates et al., 2015). Analyses were conducted in R

Studio (R, 2024). Experimental factors were entered as fixed effects and random intercepts for both participants and items were included in the model. Sound files were uniquely coded and included as random intercepts in the statistical models to account for individual and item-level variability. Behavioral outcomes were described in terms of accuracy, response times, or selection preferences across the experiments. Over the four studies, to account for multiple comparisons, Bonferroni, family-wise error (FWE), or false discovery rate (FDR) correction was applied to p-values based on the particular study design.

3.5.2. Neural analysis

The neural analyses focused on temporal dynamics, making the event-related potential (ERP) technique ideal, given its millisecond-scale time resolution. Neural data were analyzed in MATLAB (MathWorks, 2024) and R (R, 2024). Electroencephalography (EEG) signals were time-locked to stimulus onset and averaged across trials, providing event-related potentials (ERPs). Mean ERP amplitudes were averaged from predefined time windows based on previous literature (e.g., PMN: 150–300 ms). Nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) including 10,000 permutations were used for ERP analysis using Fieldtrip (Oostenveld et al., 2011). Clusters ($p<0.05$) were FWE-corrected.

4. Investigations

4.1. Study 1 – Temporal dynamics of coarticulatory cues to prediction

Study 1 (Lulaci et al., 2024) was the first step of this thesis. In this study, fine-grained acoustic cues were used to test how early acoustic information can be used for speech perception. Using an adapted behavioral gating paradigm (Grosjean, 1980), this study aimed to investigate the temporal dynamics of predictive cues of within-word coarticulation. Speech unfolds rapidly; listeners do not wait until a word is finished; instead, the comprehension process starts as soon as the word onset is heard (Warren & Marslen-Wilson, 1987; Norris et al., 2016). Because speech is coarticulated, each speech sound is affected by the surrounded sounds. For instance, a vowel influences the preceding word onset consonant, and some consonants carry these cues more noticeably than others. Interestingly, /s/ has been reported as being influenced by following rounded vowels several times in prior research (Lubker & Gay, 1982; Perkell & Matthies, 1992). This study tested one of the research questions of this thesis: how fast listeners can predict the upcoming speech sounds and, finally, the word, when hearing only part of the word onset. Using Swedish words with /f/ and /s/ onsets, we tested how quickly and accurately the listeners can predict the words based on fine-grained acoustic cues.

This study has two aims: to understand, first, how quickly listeners can identify the word, and second, how much of the acoustic cues the word onsets carried from the upcoming vowel to make this differentiation as seamless as possible. The design of the study focused on word onset fricatives, which allowed us to test our hypothesis that word onset fricatives carry enough cues to give leverage to the listeners to identify the words while they are listening, already at the early partial information of the word. Using an adapted gating paradigm, we created four gates. Starting from the word onset, we divided the word onset fricatives into four gates: 15 ms, 35 ms, 75 ms, and 135 ms. This way we aimed to test the word recognition threshold of listeners both for understanding the acoustic cues that exist in the acoustic signal and to assess how much of these cues can be used by listeners predictively to identify the word and how these cues differentiate from each other based on articulatory differences.

Listeners significantly predicted the upcoming speech sounds and the identity of the word by listening to the onset fricative for only a fraction of a second. The results showed ultra-rapid effects of predictive coarticulatory cues, with a mere fifteen milliseconds of an onset fricative yielding reliable behavioral judgments about the upcoming vowel with /s/ onset words based on coarticulatory lip rounding. For /f/ onset words, other vowel features such as backness and height exhibited a slightly slower (75 ms) – but still rapid – time course of identification.

This study set the stage for the rest of this thesis, showing that recognition can occur from such minimal information, and that the precision of prediction depends on auditory sensitivity to fine-grained acoustic detail through segmental and suprasegmental cues in the speech signal.

4.2. Study 2 – Extended high-frequency hearing sensitivity facilitates predictive speech perception

Study 2 (Lulaci et al., 2025) investigated the relationship between extended high-frequency hearing sensitivity and listeners’ ability to predict words using fine-grained auditory cues. People vary in their spoken-word recognition performance. One factor contributing to this difference could be hearing sensitivity in the extended high-frequency (EHF) range. By linking EHF hearing to time-resolved spoken-word recognition in a gating paradigm, this study aimed to investigate how efficiently the individual hearing profile shapes the perception and use of fine-grained acoustic cues as the speech unfolds.

Standard pure tone audiometry assessment measures the frequencies up to 8 kHz. However, there is a growing body of evidence that frequencies beyond 8 kHz support speech perception in several situations. Recent reviews have shown relations between EHF sensitivity to speech perception in noise, sound localization, aging, and early signs of cochlear dysfunction (Hunter et al., 2020). While prior studies have highlighted the importance of EHF thresholds and high frequency information in the speech signal, its contribution to predictive speech processing remained unexplored. This study aimed to explore the connection between EHF hearing sensitivity and predictive mechanism for speech perception to allow listeners to identify words from fine-grained, minimal acoustic cues. The study addresses the gap in the literature by testing the relation between EHF thresholds beyond conventional clinical range and spoken word recognition performance based on predictive coarticulatory cues.

The design of Study 2, built directly on the first study, which showed that words can be identified from extremely short gates (15 ms) starting from fricative onsets. Here, the same gating paradigm with /f/ and /s/ onset words was combined with pure tone audiometric assessment, extending beyond the conventional hearing test range, including EHF thresholds. Participants heard fricative-onset words presented for a fraction of a second and attempted to identify them after listening to each gate at 15, 35, 75, and 135 ms. This design allowed us to test the link between EHF hearing sensitivity and a possible advantage in accurately identifying the words using acoustic cues predictively. Twenty native speakers of Central Swedish participated in the study. Participants' hearing were within normal hearing range (interaural difference ≤ 5 dB HL, 0.25-8 kHz < 20 dB HL) and were young adults (mean age = 24.6 years, SD = 3.7, range = 20–33 years). In the analysis, EHF hearing thresholds were tested while controlling for pure tone average (PTA) of standard hearing frequencies. This allowed us to assess unique variance explained by EHF hearing sensitivity and test the possible collinearity with PTA.

The results demonstrated that better EHF hearing sensitivity provided an advantage in using acoustic cues to accurately identify target words. However, this effect was only observed for /s/ onset words. It persisted when controlling for PTA, suggesting that EHF hearing sensitivity explained variance beyond the conventional audiometric range. In contrast, for /f/ onset words, no reliable relation between EHF sensitivity and accuracy was found. However, /f/ onset word accuracy correlated with PTA in the 135-ms gate. As the results of Study 1 suggested, /s/ onset words may carry more distinct cues. Study 2 highlighted the high frequency spectral detail of /s/ which could be used by listeners with better EHF hearing sensitivity as predictive cues. For /f/, where the acoustic profile is diffuse without prominent peaks and carries weaker high-frequency detail compared to /s/, EHF hearing sensitivity provided little benefit.

These results suggested that the EHF hearing thresholds affect predictive spoken word recognition performance, particularly when the speech signal contains sufficient fine-grained acoustic cues to support the prediction.

4.3. Study 3 – Neural correlates of prosodic cues in predictive speech perception in noise

Submitted manuscript

Study 3 (Lulaci et al., submitted-a), investigated the role of prosodic cues in speech perception in noise. Everyday communication rarely occurs in silence. Competing talkers, background noise, and unpredictable interruptions make speech perception challenging. Under these listening situations, actively predicting what is coming next rather than listening passively could be one way to explain the human brain's successful perception.

Swedish word accents Accent 1 and Accent 2 differ in their prosodic contour. This suprasegmental distinction on the word stem provides cues that can be used predictively (Roll et al., 2010; Roll et al., 2013; Roll, 2015; Roll et al., 2015; Söderström et al., 2016; Roll, 2022). The aim of this study was to test prosodic cues' support to predictive speech perception under adverse listening conditions. Swedish word accents were used as controlled cues of predictive prosodic information in noise. In the experiment, participants listened to sentences containing target words with either plural or singular endings, with the critical suffix being matched or mismatched with the word accent (the prosodic contour of the word stem). Speech-shaped noise was used to mask the speech signal to reduce intelligibility while keeping the overall spectral profile similar, mainly causing energetic masking (Culling & Stone, 2017). During the experiment, three listening conditions were interleaved: quiet, 0 dB SNR, and -5 dB SNR. Participants listened to the sentences with target words and were asked to decide whether the target word was plural or singular. EEG was recorded to measure the ERP responses.

The results showed that predictive processing was modulated by both accent type and listening conditions. Accent 1 mismatched words elicited an N400 increase consistently across all the listening conditions, even under high noise masking (SNR-5), indicating that listeners used prosodic cues predictively throughout the experiment for accent 1 words, while a P600 was only observed in the quiet listening condition. In contrast, accent 2 elicited an increased N400 only in the quiet listening condition, and a P600 at SNR0. No effects were observed under the highest noise level in this experimental setting (SNR-5). Although the ERPs reflect cortical processing, prosody relies on strict early encoding of temporal modulations (Moore, 2008), and the subcortical inferior colliculus (IC) is responsible for tracking temporal structure and following its amplitude, playing an essential role in speech and tone processing, particularly under adverse conditions (Söderström, 2024). Together with the auditory cortex, it predicts individual differences in pitch discrimination (Bianchi et al., 2017). From this perspective, an interpretation of these results could also be that accent 1's prosodic contour remained usable as a cue even in severe noise while the accent 2 pitch contour was suppressed by energetic masking. Noise can already affect the encoding of prosodic cues at subcortical levels prior to cortical processing. A difference between accent 1 and accent 2 was also observed in behavioral findings. Accent 1 mismatch trials showed lower accuracy than match trials while accent 2 did not show any differences based on accuracy. Response times were also measured and analyzed, and longer response times were consistently observed for mismatch trials in each noise level for accent 1, while accent 2 showed minimal difference between match and mismatch response times. Accent 1 was observed to be predictively informative.

The pre-activation negativity (PrAN) is an ERP component elicited by word onsets beginning around 136 ms from fundamental frequency onset (Söderström et al., 2016; Roll et al., 2023; Roll et al., 2015; Söderström & Cutler, 2023). While previous studies have reported PrAN in quiet listening settings, study 3 tested PrAN in challenging listening conditions with interleaved design. The results shown that a PrAN elicited by the accent 1 vs. accent 2 contrast appeared under high noise (SNR-5). This pattern was interpreted as the PrAN indexing a functionally adaptive mechanism. Specifically, the interleaved design meant that even quiet trials were not truly "quiet." Listeners remained in an attentive state of alertness throughout the experiment, which may have shaped the emergence of PrAN for accent 2, evening out the difference with respect to accent 2, except under the hardest listening condition.

Overall, Study 3 suggested that Swedish word accents continued to function as predictive cues even when the speech signal was masked by noise under challenging listening conditions. These results extend earlier work on word accents by showing that the predictive function of word accents is not limited to quiet listening but rather stays resilient in more challenging settings. Additionally, the ERP results suggested that predictive processing is flexible rather than static: N400, P600 and PrAN component patterns shifted as a function of both noise level and accent type, suggesting the brain changes strategies to maintain comprehension in adverse listening conditions.

4.4. Study 4 – Neural dynamics of rapid acoustic cues in spoken-word prediction

Submitted manuscript

Study 4 (Lulaci et al., submitted-b) investigated how the auditory system uses early coarticulatory cues in the speech signal to generate expectations during the recognition of words spoken in isolation. While a larger body of work has focused on speech perception in sentence contexts marked by consistent higher-level lexical expectations (Altmann & Kamide, 1999; Brothers & Kuperberg, 2021; DeLong et al., 2005; Federmeier, 2007), this study investigated speech perception at the within-word level with a focus on the first few hundred milliseconds of acoustic cues. The aim of this study was to test how listeners use fine-grained coarticulatory cues within word-onset phonemes to predict the upcoming speech sounds to words, and how this process unfolds in real time.

The experiment used naturally produced Swedish real words (CVC) organized into onset-matched lexical competitor sets. The onset fricative was manually identified in Praat (Boersma & Weenink, 2021). To create mismatch tokens, the fricative onset from one word was cross spliced onto the entire continuation (vowel + coda) of a competitor word within the same onset set. Match tokens were presented unaltered. Splice boundaries were moved to the nearest zero crossing to prevent artifacts, and to preserve the natural temporal patterns. Participants were presented with real words (CVC) beginning with /s/ and /f/ onsets. In each set, all words began with the same fricative (/s/ or /f/) but differed in the offset of the word (vowel + coda), supporting different lexical options. This structure allowed us to investigate how listeners use early acoustic cues in the onset to generate lexical predictions and how they respond when later input either confirms or contradicts that expectation. The pairs presented in each trial were always the competitor words from the same onset list. Importantly, in each mismatch trial, the two visually presented response options were the same real words used in the splice. Within each onset-matched set, every word appeared in both roles: providing the onset for multiple competitor continuations and providing the continuation for multiple competitor onsets. No word was ever paired with itself. This resulted in bidirectional cross-splicing. Each unique splice configuration was repeated equally. This design made it possible to analyze the neural mechanism behind the results of the first study (Lulaci et al., 2024). Neural activity was tracked and recorded using EEG starting from word onset during online listening.

Event-related potentials locked to word onset and vowel onset were analyzed. In the onset analysis, both onsets followed by unrounded and rounded vowels were analyzed. Word-onset locked ERP results of /s/-onsets followed by rounded vowels and /s/-onset words followed by unrounded vowels showed a divergence starting from 45 ms up to 70 ms following the word onset for /s/-onset words. However, this effect could not be observed in /f/-onset words.

The vowel onset was the first disambiguation point for listeners to perceive if the onset was followed by its continuation or spliced to a competitor word offset. The neural responses to /f/- and /s/-onset words differed both temporally and topographically.

For /s/-onset words from the vowel onset, listeners showed mismatch-related responses that differed from /f/-onset word trials. Mismatches elicited a phonological mapping negativity (PMN) (Connolly & Phillips, 1994) for s/-onset words following vowel onset (150–300 ms). This negativity reflects the violation of the listener’s expectation about the upcoming phoneme based on the information already extracted and processed at that point in time. Crucially, the PMN emerged even though the onset itself was identical across match and mismatch conditions, suggesting that listeners had formed a reliable prediction about the upcoming input based on onset acoustics alone. However, this neural response was only observed for /s/ while /f/-onset words did not elicit any PMN. Following PMN, /s/-onset words elicited an N400 effect (300–400 ms). Following this N400, a late positivity was observed from 500 to 650 ms.

In contrast, /f/-onset words displayed a different profile than /s/-onset words. Mismatch trials elicited a positivity starting from 300 ms (300–400 ms) and extending into the later time window (500–650 ms). These differences aligned with the behavioral data; mismatch trials slowed down responses only for /s/-onset words.

In addition to investigations of neural responses to match and mismatch outcomes this study also tested neural sensitivity prior to the disambiguation point. This allowed us to assess rapid coarticulatory cue processing before later segmental information confirmed or violated lexical expectations. The findings suggested that word onset coarticulatory cues modulate neural processing during spoken word recognition. For /s/ onset words, neural sensitivity to vowel rounding was observed as an increased negativity within ~45 ms of word onset and interpreted as indicating rapid use or detection of subphonemic information.

Taken together, both behavioral and neural patterns indicate that prediction during speech perception is not uniformly applied across phonemes. Instead, the system weighted the cues based on their acoustic informativeness. Better-diagnosed segments like /s/-onsets generated early and constrained expectations that spread across phonological and lexical levels as shown by modulations of PMN and N400 components. Less acoustically informative phonemes like /f/ delayed prediction and shifted processing in time. This study provides evidence that fine-grained rapid coarticulatory information can drive anticipatory neural processing during spoken word recognition, even when words are presented in isolation.

Overall, Study 4 suggests that listeners rapidly use fine-grained acoustic details and that predictive processing depends on the informativeness of the unfolding onset signal. The findings contribute to our understanding of how the auditory system integrates anticipatory onset cues with later segmental input during real-time speech perception. The results show that the auditory system combines coarticulatory onset information with later segmental input online, in which early auditory cues guide the unfolding activation of phonological and lexical representations. Finally, the human brain uses rapid sensory information for spoken word recognition, and different violations reflect distinct neural mechanisms temporally and topographically.

5. General Discussion

Speech perception is a dynamic process that continuously anticipates what is coming next (Heilbron et al., 2022; Heilbron & Chait, 2018; Friston et al., 2021). Across the four studies of this thesis, converging behavioral and neural evidence made apparent an adaptive and resilient feature of speech perception using predictive processing: participants did not wait for complete words or even phonemes before beginning to interpret the signal. These findings suggested an adaptive and strategically driven predictive mechanism that actively assesses accessible cues. Rapidly unfolding coarticulatory cues, prosodic cues, and hearing sensitivity all contributed to speech perception with varying degrees. Across behavioral and electrophysiological evidence, predictive processing showed a robust and resilient pattern, aiding the comprehension process, and reorganizing itself when the cues were masked, misleading or weak.

The first study used a gating paradigm to test how early acoustic cues can be used predictively to anticipate upcoming sounds in spoken word recognition. Results showed that listeners could predict upcoming sounds and finally identify words using only partial short gates of onset phonemes as early as 15 ms into word onsets. The second study tested whether this ability was affected by hearing sensitivity, focusing on extended high-frequency hearing thresholds. Individuals with better extended high-frequency hearing thresholds made more efficient use of predictive cues. This suggests that extended high-frequency hearing influences how efficiently listeners can perceive and use fine-grained acoustic information predictively. The third study investigated whether prosodic cues, specifically Swedish word accents, remain reliable cues when the speech signal was masked by speech-shaped noise. In mismatch trials, N400 and P600 effects persisted under adverse listening conditions although their topography shifted and the amplitude weakened. These results indicate that noise affected the predictive use of prosodic cues but did not eliminate them. Furthermore, a pre-activation negativity (PrAN) for accent 1 as compared to accent 2 was observed solely under the highest noise-masked listening condition (SNR–5 dB). This suggest that pre-activation was used as a functionally adaptive mechanism supporting comprehension differentially under increased uncertainty. The fourth study examined how coarticulatory cues influence the neural dynamics of prediction in real time. Cross-spliced stimuli were used to create match and mismatch conditions between the word onset fricatives and the rest of the word (vowel + coda). The results highlight that when fine-grained cues signaling the upcoming sound are available at word onset, the brain registers them rapidly, as early as ~45 ms. Coarticulatory cues were found to be used predictively. That is, mismatch between the onset and the rest of the word elicited different neural responses across fricatives: /s/ onset words showed PMN and N400 effects, as well as late positivity, whereas /f/ onset words only evoked a P300 response. This pattern suggested that listeners were sensitive to fine-grained coarticulatory cues and that the processing was affected by the spectral properties and accessible cues within the unfolding speech signal.

In line with Grosjean (1980), recognition of the unfolding spoken words does not wait for the word to end. Moreover, several studies to date have shown that coarticulatory cues can be used to support speech perception (Beddor et al., 2013; Salverda et al., 2014; Gow & McMurray, 2007). Building on these findings, we sought to determine how early such cues may begin to guide spoken word recognition.

Across the first two studies, coarticulatory cues were shown to facilitate rapid offline word recognition, suggesting that the acoustic cues available for predictive use in speech may begin earlier than commonly assumed. While previous studies has demonstrated incremental interpretation of speech as the acoustic-phonetic information unfolds (Salverda et al., 2014; Swingley et al., 1999), the present findings suggest that use of acoustic cues can start at an even earlier point in the unfolding signal. The salience of coarticulatory cues improved recognition for very short onset gates, indicating that listeners do not have to wait for the word to fully unfold to identify it. Instead, they appear to make use of the earliest fine-grained spectral cues available. Importantly, the results also showed differences in cue strength in fricatives. For instance, recognition was faster for /s/ than /f/, suggesting that certain spectral cues may be more perceptually salient and more useful for predictive processing. Although recent work has highlighted dynamic and hierarchical coding of speech in the auditory cortex, the temporal dynamics remain debated. Importantly no clear consensus has emerged regarding pre-lexical processing (Obleser & Eisner, 2009). It remains unclear when the brain first begins to form predictive phonological information from the acoustic signal. However, findings of Study 1 helps narrow this time window by showing that listeners can perceive and use acoustic cues informatively starting as early as 15 ms from word onset.

Study 2 extended this observation by demonstrating that the efficiency of early predictive processing depends, in part, on hearing sensitivity, particularly extended-high frequency hearing thresholds. Individuals with better EHF sensitivity showed enhanced recognition of the early /s/ onset cues when they were accessible. Fricatives differ based on their spectral properties (Shadle et al., 2023), and /s/ has stronger spectral peaks in high-frequency energy regions (Strevens, 1960). Therefore processing of /s/ may have been strongly affected by EHF hearing, while /f/, which has been reported as a spectrally flat and more turbulent sound (Hughes

& Halle, 1956), remained less affected. This adds to increasing evidence that frequencies above 8 kHz can support speech perception (Hunter et al., 2020) and suggests that predictive speech perception relies on both cognitive expectations and the sensory ability to process detailed acoustic information at word onset. As mentioned in Hunter et al. (2020), from a biological perspective, the human sensory system is tailored to features useful for survival, and EHF hearing provides a sound localization advantage. With previous findings showing that EHF hearing also supports speech perception in noise in addition to sound localization (Hunter et al., 2020; Masterton et al., 1969), the present results suggest another facilitative effect of EHF sensitivity: that of predictive speech processing.

Study 1 and 2 established that listeners can use fine-grained coarticulatory cues predictively as early as at word onset. This was observed even during spoken word recognition in isolation, and this ability depended partly on hearing sensitivity. Study 4 tested if this predictive processing could be observed online at the neurophysiological level. Words from the stimulus list of studies 1 and 2 were used but the onset sounds were cross-spliced to intentionally create a prediction conflict. Using ERPs, we tested how the brain responds when predictive coarticulatory cues were violated by the unfolding speech signal, focusing on anticipatory neural activity starting from the earliest stages of processing.

These results further complemented studies 1 and 2 showing a similar pattern; cue differences shape the timing and type of the neural response. In the mismatch conditions, /s/-onset words showed clear prediction error effects in the shape of increased response times and an early (150–300 ms) PMN. Following the PMN, mismatched /s/-onset words elicited N400 effect (300–400 ms), which was interpreted as violated onset-based predictions influencing later stages of lexical semantic congruency and followed by a late positivity. However, /f/-onset words only elicited only a late positivity, interpreted as a P300 indexing auditory reassessing and model updating after hearing an acoustically unexpected token. One way to explain this is that, compared to /s/, /f/ onset words' onset cue based anticipation was weaker or less decisive, shifting processing toward later stages.

While PMN, P300 and N400 effects highlight differences in later expectation violation and evaluative processes, the Study 4 findings suggested neural sensitivity at an even earlier time scale. Within ~45 ms of /s/ word onset, a neural difference between rounded and unrounded vowel contexts emerged. This rapid sensitivity may reflect the early registration and integration of coarticulatory information.

Study 3 tested a different type of cue than studies 1, 2, and 4. Here, prosodic cues using Swedish word accents were tested. Accent 1 consistently triggered prediction-related activity across listening conditions including SNR –5 dB, indicating that prediction was still active even when the signal was degraded. However, accent 2 did not show this robust effect under the highest noise condition. A PrAN for accent 1 compared to accent 2 was observed only under the highest level of noise, SNR –

5 dB, corresponding to the most challenging listening condition. One possible interpretation of this pattern is PrAN indexing a functionally adaptive process that supports pre-activation to elevate the predictive perception under challenging listening conditions. On the other hand, at the highest noise level accent 2 prosodic cues may be less accessible, limiting its effectiveness in noise. These findings extend the growing body of work on Swedish word accents (Roll, 2022; Roll et al., 2023) by showing that word accents can support predictive speech processing even under adverse listening conditions. Although sentence-level prediction can benefit from and be guided by contextual cues (DeLong et al., 2005), the current studies intentionally minimized these influences. In studies 1, 2, and 4, the stimuli consisted of words in isolation, and Study 3 used a neutral carrier sentence rather than a semantically constraining sentence. This design allowed us to limit sentence-level contextual prediction and isolate the acoustic cues involved in prediction.

Across the four studies, the converging result was that speech perception is supported by predictive mechanisms that are graded and adaptive rather than fixed. In all studies, prediction behaved as a part of a dynamic system that reorganizes itself according to the strength and accessibility of available cues. These results position prediction as a graded rather than binary success-or-failure mechanism that operates under uncertainty and selects whichever cue offers the strongest support to achieve successful speech perception. This view aligns with the predictive coding framework (Friston, 2005): listeners continuously engage with stimuli to minimize surprise in the long run. This showed itself clearly in Study 3, where different ERPs elicited by mismatch conditions were modulated by both noise level and accents.

In study 3, with increasing noise, the N400 persisted although the scalp distribution shifted toward a more frontal and widespread negativity for accent 1. In contrast, the P600 disappeared as listening become effortful and prediction became less stable. Thus, the shift in N400 topography with increased background noise may potentially reflect the effortful listening. The presence of N400 effects despite the noise indicates that listeners continued to track the accent 1 prosodic cues and use them predictively to detect mismatches between expected and heard suffixes. This could be related to accent 1's robust predictive nature. It has been shown that when the speech signal was degraded, the brain flexibly shifts its neural processing strategies, engaging more effortful processing as predictability decreases (Obleser & Kotz, 2010). Listening to speech under masked conditions engages additional neural networks involved in auditory attention and control (Evans et al., 2016). Accent 1 was still providing enough cues to give support during the most challenging listening conditions while accent 2 was not useable due the noise masking. A processing difference between accent 1 and accent 2 has been reported several times since Roll et al. (2010). However, this difference of their cue performance is not necessarily due to the acoustic difference between the word accents (Roll, 2015; Roll et al., 2013). Rather, accent 1 has been reported as a more reliable cue based on lexical statistics across several word accent studies

(Söderström et al., 2016; Hjortdal et al., 2024). In phonological terms, accent 1 might be the default accent, whereas accent 2 is marked (Riad, 2014) and needs to be more clearly perceived to be processed.

The finding that PrAN increased for accent 1 as compared to accent 2 only in SNR–5 dB suggests that accent 1 served as a stronger cue for pre-activation when sensory input became least reliable. The influence of prior knowledge becomes stronger when the sensory information is reduced (Sohoglu et al., 2012). In the present thesis, the general principle was mirrored by the increased pre-activation observed for accent 1 in the highest noise condition. These findings provide insights into how and to what extend listeners use prosodic cues predictively to support speech perception under increased ambiguity. A possibility is that, in the presence of moderate noise and clear speech, accent 2 was used for pre-activation to a similar extent as accent 1 when perceivable. However, under the highest noise masking, only accent 1 remained robust enough to support predictive processing while accent 2 did not, increasing the difference between the word accents.

Another early ERP component possibly indexing pre-activation prior to the phonological disambiguation point in Study 4 suggests that the auditory processing begins evaluating coarticulatory information early in the unfolding signal, within 45–70 ms. Early ERP responses in comparable time windows have sometimes been associated with lexical access in word–pseudoword and learning paradigms (Gosselke Berthelsen et al., 2020). In the present thesis, coarticulatory cues were already present in the signal and accessible to listeners, as demonstrated in studies 1 and 2. While one possible interpretation is that the observed divergence between the two ERP conditions may reflect lexical pre-activation, the pattern can also be accounted for by rapid encoding of fine-grained coarticulatory detail. Across both behavioral and neural measures, /s/ onset functioned as a stronger cue than /f/ onset. In study 1, listeners were faster and more accurate when identifying /s/ onset words, suggesting that rapid acoustic information was sufficient to begin narrowing the set of lexical candidates. This pattern was replicated in Study 4, where only /s/ onset words showed response-time differences between match and mismatch conditions and ERP responses differed between /s/ and /f/ onsets. These findings suggest that the predictive use of early onset cues depends on the strength and clarity of the acoustic information available in the signal. This pattern does not require lexical access to begin within the early ERP window (~45 ms), but it remains compatible with the hypothesis that early sensory encoding supports subsequent lexical processing.

These findings highlight how finely tuned the human auditory system is to the acoustic details of speech. Across the studies in this thesis, listeners used rapid and fine-grained cues (including high-frequency energy, coarticulatory information, as well as prosodic cues in noise), to predict within-word continuations. All these cues influenced both behavioral performance and neural responses.

Acoustic-phonetic features are encoded across distributed neural populations in the auditory cortex, supporting robustness of speech perception under variability (Obleser et al., 2010). Phonetic features have been found to be represented in the STG via distributed neural populations. These populations have also been shown to respond to speech sound cues and phonetic features such as articulatory gestures (Leonard et al., 2024). Speech perception is processed in a detailed, layered system where information is already structured before lexical access. The auditory cortex tracks speech simultaneously at multiple temporal scales with neural activity showing both syllabic and phonemic dynamics in the acoustic signal (Giroud et al., 2024).

Across all four studies in this thesis, a consistent pattern emerged. Listeners relied on whatever acoustic information was available. Whether the relevant cues derived from fine-grained coarticulatory information, high-frequency spectral detail, prosodic cues, or accessible cues in a masked speech signal, listeners were able to use these cues predictively. Taken together, the findings suggest that speech perception does not unfold as a strictly linear process. The findings point to interactive and parallel mechanisms operating at multiple levels of analysis, aligning with models of speech perception with co-existing layers and hierarchical representations (Davis & Johnsrude, 2007). Overall, these studies suggest the predictive use of acoustic cues to facilitate spoken language recognition and reflect a dynamic and adaptive system where fine-grained acoustic detail and contextual information are continuously processed and integrated.

Across these findings, a common theme may be inferred: speech perception is resilient and supported by continuous acoustic cues, that remain informative even when the signal is brief, masked, or contains only subtle acoustic detail.

By combining behavioral, audiological, and EEG measures, this thesis shows how rapid acoustic detail can influence prediction to facilitate speech perception and support comprehension under adverse listening conditions. In this way, these findings contribute to a broader view of speech perception, suggesting a dynamic and adaptive process where listeners listen actively rather than waiting until the signal has fully unfolded.

6. Conclusions

This thesis examined how listeners extract and use fine-grained acoustic information to navigate spoken language and understood speech in real time. By focusing on the predictive use of acoustic cues, including coarticulation and prosody, it investigated both behavioral and neural dynamics of speech perception. Across four studies combining behavioral, EEG, and audiological assessment, the findings consistently showed that speech perception is a dynamic process where acoustic and phonetic information is actively used as the signal unfolds. Listeners were able to use available acoustic information to generate predictions rapidly within the first tens to hundreds of milliseconds of the signal.

In a gating paradigm, partial acoustic information available at word onset carried sufficient cues to guide listeners' interpretations, suggesting that predictive constraints can be established very early in processing. Listeners relied on graded coarticulatory information rather than waiting for fully specified phonemic input, with predictive use emerging as early as 15 ms for /s/ onset words and later for /f/ onset words. These results show that spoken-word recognition is initiated through continuous tracking of fine-grained acoustic structure from the earliest moments of the signal.

Individual differences in extended high-frequency hearing further modulated the efficiency with which early acoustic cues were used. Listeners with better extended high-frequency sensitivity were able to make more effective use of the earliest available information, particularly for stimuli containing relatively higher-frequency energy. Although all participants had clinically normal hearing, these findings suggest that hearing acuity can influence how rapidly fine-grained acoustic cues are used during predictive speech processing.

Under adverse listening conditions, predictive processing did not disappear but adapted according to cue accessibility. Word-level prosodic cues continued to support prediction in noise, with Swedish word accents showing differential robustness as masking increased. Even at the most challenging noise levels, predictive neural responses were preserved, indicating that listeners adjust their perceptual strategies rather than abandoning prediction when acoustic information is limited.

Importantly, electrophysiological evidence demonstrated sensitivity to coarticulatory onset information within early neural time windows at word onset, prior to the availability of disambiguating segmental input, providing direct neural support for rapid predictive processing during spoken-word recognition. These findings point to a close relationship between the structure of the auditory signal and neural responses during speech perception and support a view of speech processing as a continuous and adaptive process.

Across all four studies listeners used continuous and fine-grained acoustic cues from the beginning of the speech signal. Predictive constraints emerged rapidly from the speech signal itself, adapted to changes in listening conditions, and were observed even when words were presented in isolation, without sentence-level context. These findings support the role of early acoustic encoding in shaping perceptual interpretation and suggest a close relationship between the auditory signal and neural responses during speech perception. Under challenging listening conditions or with partial cues, prediction does not fail, but adjusts according to what the signal allows, reflecting a flexible adaptive processing pattern.

By combining behavioral, auditory and neural evidence, this thesis shows that speech is processed continuously, closely tracking the spectrotemporal dynamics of the speech signal as it unfolds.

7. Outstanding issues and future directions

There are some aspects raised in this thesis that would benefit from further investigation in future studies. One contribution of the thesis was showing that rapid fine-grained acoustic cues can be used predictively to anticipate the upcoming sounds. However, the stimulus lists consisted of isolated words and were restricted to two fricative onsets (/f/ and /s/), in order to allow precise control of acoustic detail. A natural extension of this work would be to expand the onset diversity and to embed words in continuous or conversational speech. Future studies could also examine whether graded differences in vowel roundedness further modulate the availability and use of coarticulatory cues beyond the categorical distinctions. This would help future studies to test whether similar predictive dynamics emerge across a wider range of cues.

Another relevant issue concerns individual variation in extended high-frequency hearing sensitivity. In study 2, individuals with better hearing thresholds showed more efficient use of fine-grained acoustic cues relative to people with elevated hearing thresholds. Despite containing a relatively high number of trials (3,040), the sample size was modest (20 participants). Future research could be further developed by testing the relation between predictive use of cues and hearing acuity in larger and more diverse populations, including clinical groups with hearing loss. This would help to determine how robust the effect is and how it operates beyond normal-hearing listeners.

Attention may also have influenced performance due to the 2AFC task, since listeners were required to choose between two alternatives. While the rapid temporal dynamics and cue specific effects throughout the studies suggest that prediction is not likely to be reduced to attentional engagement, its potential contribution could be investigated further. The noise paradigm in Study 3 used speech-shaped noise to mask the target sentences. Future work could extend this approach by introducing increased ecological validity, for instance using competing talkers or babble noise. Combining these paradigms with physiological measures of listening effort (i.e. pupillometry) or subcortical responses such as auditory brainstem response (ABR) or frequency following response (FFR) could help to clarify how prediction becomes more effortful when bottom-up evidence is limited and ambiguous, or how listeners adjust their strategies when the signal is masked and degraded.

This thesis contributed to a more integrated understanding of how auditory processing supports spoken-language comprehension, with implications for individual differences in hearing and for communication in challenging listening environments. A more applied direction concerns the relevance of the present findings for speech-hearing technologies. The link between extended high-frequency hearing sensitivity and predictive use of fine-grained coarticulatory cues raises the possibility that access to fine-grained acoustic information may play a functional role in speech perception. Future studies could test whether this information can be supported or enhanced in hearing aid signal processing. The findings of Study 3 also highlight a potential value of word accents (prosodic cues) in adverse listening conditions. Future studies could investigate whether word accents remain accessible and predictively used by listeners with hearing loss including when using hearing-aids.

Future research could investigate how a wider set of speech sounds can contribute to rapid predictive speech processing. In particular, further researching PrAN and the predictive use of word accents in adverse listening conditions may help to clarify how prosody supports prediction during everyday listening, where the signal is often uncertain and masked.

References

Alickovic, E., Lunner, T., Wendt, D., Fiedler, L., Hietkamp, R., Ng, E. H. N., & Graversen, C. (2020). Neural Representation Enhanced for Speech and Reduced for Background Noise With a Hearing Aid Noise Reduction Scheme During a Selective Attention Task [Original Research]. *Frontiers in Neuroscience*, *Volume 14 - 2020*. https://doi.org/10.3389/fnins.2020.00846

Alickovic, E., Ng, E. H. N., Fiedler, L., Santurette, S., Innes-Brown, H., & Graversen, C. (2021). Effects of Hearing Aid Noise Reduction on Early and Late Cortical Representations of Competing Talkers in Noise [Original Research]. *Frontiers in Neuroscience*, *Volume 15 - 2021*. https://doi.org/10.3389/fnins.2021.636060

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264. https://doi.org/10.1016/S0010-0277(99)00059-1

Archibald, L. M. D., & Joanisse, M. F. (2011). Electrophysiological responses to coarticulatory and word level miscues. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1275-1291. https://doi.org/10.1037/a0023506

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology*, *80*(1), 23-34. https://doi.org/10.1016/j.biopsycho.2008.01.015

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, *49*(12), 3512-3516. https://doi.org/10.1016/j.neuropsychologia.2011.08.014

Badri, R., Siegel, J. H., & Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *The Journal of the Acoustical Society of America*, *129*(2), 852-863. https://doi.org/10.1121/1.3523476

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1 - 48. https://doi.org/10.18637/jss.v067.i01

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350-2366. https://doi.org/10.1121/1.4794366

Bell-Berti, F., & Harris, K. S. (1979). Anticipatory coarticulation: some implications from a study of lip rounding. *The Journal of the Acoustical Society of America*, *65*(5), 1268-1270. https://doi.org/10.1121/1.382794

Bianchi, F., Hjortkjær, J., Santurette, S., Zatorre, R. J., Siebner, H. R., & Dau, T. (2017). Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. *NeuroImage*, *163*, 398-412. https://doi.org/10.1016/j.neuroimage.2017.07.057

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer (version 6.1. 54)*. In *http://www.praat.org/*

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language*, *116*, 104174. https://doi.org/10.1016/j.jml.2020.104174

Bruce, G. (1977). *Swedish word accents in sentence perspective* [Doctoral Thesis, Lund University]. Lund.

Bruce, G. (1987). Nordic prosody IV: papers from a symposium In K. Gregersen & H. Basbøll (Eds.), *How floating is focal accent?* (Vol. 7, pp. 41-49). Odense University Press.

Carhart, R., & Jerger, J. F. (1959). Preferred Method For Clinical Determination Of Pure-Tone Thresholds. *Journal of Speech and Hearing Disorders*, *24*(4), 330-345. https://doi.org/10.1044/jshd.2404.330

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the acoustical society of America*, *25*, 975-979. https://doi.org/10.1121/1.1907229

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204. https://doi.org/10.1017/s0140525x12000477

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J Cogn Neurosci*, *6*(3), 256-266. https://doi.org/10.1162/jocn.1994.6.3.256

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*(1), 1-18. https://doi.org/10.1016/0093-934X(92)90018-A

Connolly, J. F., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Neuroreport*, *12*(2), 237-243. https://doi.org/10.1097/00001756-200102120-00012

Culling, J. F., & Stone, M. A. (2017). Energetic Masking and Masking Release. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The Auditory System at the Cocktail Party* (pp. 41-73). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_3

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534. https://doi.org/10.1080/01690960143000074

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, *23*(8), 3423. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1), 132-147. https://doi.org/10.1016/j.heares.2007.01.014

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. *The Journal of Neuroscience*, *37*(27), 6539-6557. https://doi.org/10.1523/jneurosci.3267-16.2017

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121. https://doi.org/10.1038/nn1504

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149-179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493-513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357-374. https://doi.org/10.1017/S0140525X00058027

Elert, C.-C. (1964). *Phonologic Studies of Quantity in Swedish. Based on Material from Stockholm Speakers*. Almqvist & Wiksell.

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *J Cogn Neurosci*, *28*(3), 483-500. https://doi.org/10.1162/jocn_a_00913

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Kutas, M., & Dickson, D. S. (2016). Chapter 45 - A Common Neural Progression to Meaning in About a Third of a Second. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 557-567). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00045-6

Fowler, C. A. (2005). Parsing coarticulated speech in perception: effects of coarticulation resistance. *Journal of Phonetics*, *33*(2), 199-213. https://doi.org/10.1016/j.wocn.2004.10.003

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, *360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci*, *364*(1521), 1211-1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, *3*, 130. https://doi.org/10.3389/fpsyg.2012.00130

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2021). Active listening. *Hearing Research*, *399*, 107998. https://doi.org/10.1016/j.heares.2020.107998

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Curr Biol*, *22*(7), 615-621. https://doi.org/10.1016/j.cub.2012.02.015

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon Bull Rev*, *13*(3), 361-377. https://doi.org/10.3758/bf03193857

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences*, *104*(52), 20961-20966. https://doi.org/10.1073/pnas.0706274105

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol*, *120*(3), 453-463. https://doi.org/10.1016/j.clinph.2008.11.029

Giroud, J., Trébuchon, A., Mercier, M., Davis, M. H., & Morillon, B. (2024). The human auditory cortex concurrently tracks syllabic and phonemic timescales via acoustic spectral flux. *Science Advances*, *10*(51), eado8915. https://doi.org/10.1126/sciadv.ado8915

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol Rev*, *105*(2), 251-279. https://doi.org/10.1037/0033-295x.105.2.251

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501-518. https://doi.org/10.1016/0749-596X(89)90009-0

Gosselke Berthelsen, S., Horne, M., Brännström, K. J., Shtyrov, Y., & Roll, M. (2018). Neural processing of morphosyntactic tonal cues in second-language learners. *Journal of Neurolinguistics*, *45*, 60-78. https://doi.org/10.1016/j.jneuroling.2017.09.001

Gosselke Berthelsen, S., Horne, M., Shtyrov, Y., & Roll, M. (2020). Different neural mechanisms for rapid acquisition of words with grammatical tone in learners from tonal and non-tonal backgrounds: ERP evidence. *Brain Research*, *1729*, 146614. https://doi.org/10.1016/j.brainres.2019.146614

Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. *Papers in laboratory phonology*, *9*(173-200).

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267-283. https://doi.org/10.3758/BF03204386

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In Spoken Word Recognition, the Future Predicts the Past. *The Journal of Neuroscience*, *38*(35), 7585-7599. https://doi.org/10.1523/jneurosci.0065-18.2018

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54-73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016. https://doi.org/10.1016/j.heares.2020.108016

Hjortdal, A., Frid, J., Novén, M., & Roll, M. (2024). Swift Prosodic Modulation of Lexical Access: Brain Potentials From Three North Germanic Language Varieties. *Journal of Speech, Language, and Hearing Research*, *67*(2), 400-414. https://doi.org/10.1044/2023_JSLHR-23-00193

Hjortdal, A., Frid, J., & Roll, M. (2022). Phonetic and phonological cues to prediction: Neurophysiology of Danish stød. *Journal of Phonetics*, *94*, 101178. https://doi.org/10.1016/j.wocn.2022.101178

Holt, L. L., & Peelle, J. E. (2022). The Auditory Cognitive Neuroscience of Speech Perception in Context. In L. L. Holt, J. E. Peelle, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Speech Perception* (pp. 1-12). Springer International Publishing. https://doi.org/10.1007/978-3-030-81542-4_1

Hughes, G. W., & Halle, M. (1956). Spectral Properties of Fricative Consonants. *The Journal of the Acoustical Society of America*, *28*(2), 303-310. https://doi.org/10.1121/1.1908271

Hughson, W., & Westlake, H. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Trans Am Acad Ophthalmol Otolaryngol*, *48*(Suppl), 1-15.

Hunter, L. L., Monson, B. B., Moore, D. R., Dhar, S., Wright, B. A., Munro, K. J., Zadeh, L. M., Blankenship, C. M., Stiepan, S. M., & Siegel, J. H. (2020). Extended high frequency hearing and speech perception implications in adults and children. *Hearing Research*, *397*, 107922. https://doi.org/10.1016/j.heares.2020.107922

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252-1263. https://doi.org/10.1121/1.1288413

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, *5*(2), 115-133. https://doi.org/10.1016/S0095-4470(19)31123-4

Khalighinejad, B., Patel, P., Herrero, J. L., Bickel, S., Mehta, A. D., & Mesgarani, N. (2021). Functional characterization of human Heschl's gyrus in response to natural speech. *Neuroimage*, *235*, 118003. https://doi.org/10.1016/j.neuroimage.2021.118003

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117-129. https://doi.org/10.1016/S0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203-205. https://doi.org/10.1126/science.7350657

Kwon, J., & Roll, M. (2024). Neural semantic effects of tone accents. *Neuroreport*, *35*(13), 868-872. https://doi.org/10.1097/wnr.0000000000002077

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933. https://doi.org/10.1038/nrn2532

Leonard, M. K., Gwilliams, L., Sellers, K. K., Chung, J. E., Xu, D., Mischler, G., Mesgarani, N., Welkenhuysen, M., Dutta, B., & Chang, E. F. (2024). Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature*, *626*(7999), 593-602. https://doi.org/10.1038/s41586-023-06839-2

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177. https://doi.org/10.1016/j.cognition.2007.05.006

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev*, *74*(6), 431-461. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36. https://doi.org/10.1016/0010-0277(85)90021-6

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *The Journal of the Acoustical Society of America*, *71*(2), 437-448. https://doi.org/10.1121/1.387447

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. The MIT Press.

Lulaci, T., Söderström, P., & Roll, M. (2025). Extended High-Frequency Hearing Sensitivity Facilitates Predictive Speech Perception. *Hearing Research*, 109453. https://doi.org/10.1016/j.heares.2025.109453

Lulaci, T., Söderström, P., & Roll, M. (submitted-a). Neural correlates of prosodic cues in predictive speech perception in noise.

Lulaci, T., Söderström, P., & Roll, M. (submitted-b). Neural dynamics of rapid acoustic cues in spoken-word prediction.

Lulaci, T., Söderström, P., Tronnier, M., & Roll, M. (2024). Temporal dynamics of coarticulatory cues to prediction [Original Research]. *Frontiers in Psychology*, *Volume 15 - 2024*. https://doi.org/10.3389/fpsyg.2024.1446240

Lulaci, T., Tronnier, M., Söderström, P., & Roll, M. (2022). The time course of onset CV coarticulation. Proceedings of Fonetik 2022: Fonetik 2022-the XXXIIIrd Swedish Phonetics Conference, Royal Institute of Technology, Stockholm, Sweden.

MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, *3*(1), 711. https://doi.org/10.1038/ncomms1715

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychol Rev*, *101*(4), 653-675. https://doi.org/10.1037/0033-295x.101.4.653

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29-63. https://doi.org/10.1016/0010-0285(78)90018-X

Masterton, B., Heffner, H., & Ravizza, R. (1969). The Evolution of Human Hearing. *The Journal of the Acoustical Society of America*, *45*(4), 966-985. https://doi.org/10.1121/1.1911574

MathWorks. (2024). *MATLAB (R2024b)*. In The MathWorks, Inc.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953-978. https://doi.org/10.1080/01690965.2012.705006

Mattys, S. L., O'Leary, R. M., McGarrigle, R. A., & Wingfield, A. (2025). Reconceptualizing cognitive listening. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2025.09.014

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006-1010. https://doi.org/10.1126/science.1245994

Monson, B. B., Rock, J., Schulz, A., Hoffman, E., & Buss, E. (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. *Hearing Research*, *381*, 107773. https://doi.org/10.1016/j.heares.2019.107773

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J Assoc Res Otolaryngol*, *9*(4), 399-406. https://doi.org/10.1007/s10162-008-0143-x

Motlagh Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, W., Hunter, L. L., & Moore, D. R. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proc. Natl. Acad. Sci. U.S.A.*, *116*(47), 23753-23759. https://doi.org/10.1073/pnas.1903315116

Näätänen, R. (2000). Mismatch negativity (MMN): perspectives for application. *International Journal of Psychophysiology*, *37*(1), 3-10. https://doi.org/10.1016/S0167-8760(00)00091-X

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544-2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114-121. https://doi.org/10.1016/j.biopsycho.2008.04.008

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640-647. https://doi.org/10.1111/1469-8986.00065

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev*, *115*(2), 357-395. https://doi.org/10.1037/0033-295x.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4-18. https://doi.org/10.1080/23273798.2015.1081703

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cogn Psychol*, *34*(3), 191-243. https://doi.org/10.1006/cogp.1997.0671

Novén, M. (2021). *Brain anatomical correlates of perceptual phonological proficiency and language learning aptitude* [Doctoral Thesis, Lund University]. Lund.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, *13*(1), 14-19. https://doi.org/10.1016/j.tics.2008.09.005

Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex*, *20*(3), 633-640. https://doi.org/10.1093/cercor/bhp128

Obleser, J., Leaver, A., VanMeter, J., & Rauschecker, J. P. (2010). Segregation of Vowels and Consonants in Human Auditory Cortex: Evidence for Distributed Hierarchical Organization [Original Research]. *Frontiers in Psychology*, *Volume 1 - 2010*. https://doi.org/10.3389/fpsyg.2010.00232

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *The Journal of Neuroscience*, *27*(9), 2283-2289. https://doi.org/10.1523/jneurosci.4663-06.2007

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806. https://doi.org/10.1016/0749-596X(92)90039-Z

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Perkell, J. S., & Matthies, M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel/u/: within- and cross-subject variability. *The Journal of the Acoustical Society of America*, *91*(5), 2911-2925. https://doi.org/10.1121/1.403778

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. https://doi.org/10.1097/aud.0000000000000312

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In L. B. Joan & J. H. Paul (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137-158). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

R. (2024). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,*. In https://www.R-project.org/

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87. https://doi.org/10.1038/4580

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*, *12*(6), 718-724. https://doi.org/10.1038/nn.2331

Riad, T. (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, *122*(13), 1352-1379. https://doi.org/10.1016/j.lingua.2012.07.001

Riad, T. (2014). *The Phonology of Swedish*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199543571.001.0001

Roll, M. (2015). A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics*, *38*(2), 149-162. https://doi.org/10.1017/S0332586515000189

Roll, M. (2022). The predictive function of Swedish word accents [Hypothesis and Theory]. *Frontiers in Psychology, Volume 13 - 2022*. https://doi.org/10.3389/fpsyg.2022.910787

Roll, M., & Horne, M. (2011). Interaction of right- and left-edge prosodic boundaries in syntactic parsing. *Brain Research*, *1402*, 93-100. https://doi.org/10.1016/j.brainres.2011.06.002

Roll, M., Horne, M., & Lindgren, M. (2009). Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – An ERP study. *Journal of Neurolinguistics*, *22*(1), 55-73. https://doi.org/10.1016/j.jneuroling.2008.06.001

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain Research*, *1330*, 114-123. https://doi.org/10.1016/j.brainres.2010.03.020

Roll, M., Horne, M., & Lindgren, M. (2011). Activating without Inhibiting: Left-edge Boundary Tones and Syntactic Processing. *Journal of Cognitive Neuroscience*, *23*(5), 1170-1179. https://doi.org/10.1162/jocn.2010.21430

Roll, M., Söderström, P., Frid, J., Mannfolk, P., & Horne, M. (2017). Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters*, *658*, 57-61. https://doi.org/10.1016/j.neulet.2017.08.030

Roll, M., Söderström, P., & Horne, M. (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, *1520*, 116-120. https://doi.org/10.1016/j.brainres.2013.05.013

Roll, M., Söderström, P., Horne, M., & Hjortdal, A. (2023). Pre-activation negativity (PrAN): A neural index of predictive strength of phonological cues. *Laboratory Phonology*, *14*. https://doi.org/10.16995/labphon.6438

Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., & Horne, M. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language*, *150*, 14-21. https://doi.org/10.1016/j.bandl.2015.07.009

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances [Review]. *Frontiers in Systems Neuroscience*, *Volume 7 - 2013*. https://doi.org/10.3389/fnsys.2013.00031

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145-163. https://doi.org/10.1016/j.jml.2013.11.002

Schmitt, L.-M., Erb, J., Tune, S., Rysop, A. U., Hartwigsen, G., & Obleser, J. (2021). Predicting speech from a cortical hierarchy of event-based time scales. *Science Advances*, *7*(49), eabi6070. https://doi.org/10.1126/sciadv.abi6070

Schnupp, J., Nelken, I., & King, A. J. (2010). *Auditory Neuroscience: Making Sense of Sound*. The MIT Press. https://doi.org/10.7551/mitpress/7942.001.0001

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641-664. https://doi.org/10.1111/ejn.12816

Scott, S. K. (2005). Auditory processing — speech, space and auditory objects. *Current Opinion in Neurobiology*, *15*(2), 197-201. https://doi.org/10.1016/j.conb.2005.03.009

Shadle, C. H., Chen, W.-R., Koenig, L. L., & Preston, J. L. (2023). Refining and extending measures for fricative spectra, with special attention to the high-frequency rangea). *The Journal of the Acoustical Society of America*, *154*(3), 1932-1944. https://doi.org/10.1121/10.0021075

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech, Language, and Hearing Research*, *32*(4), 912-920. https://doi.org/10.1044/jshr.3204.912

Söderström, P. (2024). *Phonetics in the Brain*. Cambridge University Press. https://doi.org/10.1017/9781009161114

Söderström, P., & Cutler, A. (2023). Early neuro-electric indication of lexical match in English spoken-word recognition. *PLOS ONE*, *18*(5), e0285286. https://doi.org/10.1371/journal.pone.0285286

Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-Activation Negativity (PrAN) in Brain Potentials to Unfolding Words [Original Research]. *Frontiers in Human Neuroscience*, *Volume 10 - 2016*. https://doi.org/10.3389/fnhum.2016.00512

Söderström, P., Horne, M., Mannfolk, P., van Westen, D., & Roll, M. (2017). Tone-grammar association within words: Concurrent ERP and fMRI show rapid neural pre-activation and involvement of left inferior frontal gyrus in pseudoword processing. *Brain and Language*, *174*, 119-126. https://doi.org/10.1016/j.bandl.2017.08.004

Söderström, P., Horne, M., & Roll, M. (2017). Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research*, *46*(2), 271-280. https://doi.org/10.1007/s10936-016-9434-2

Söderström, P., Lulaci, T., & Roll, M. (2023). The use of lexical tone in the segmentation of speech. Annual Conference of the Australian Linguistic Society,

Söderström, P., Roll, M., & Horne, M. (2012). Processing morphologically conditioned word accents. *The Mental Lexicon*, *7*(1), 77-89. https://doi.org/10.1075/ml.7.1.04soe

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *The Journal of Neuroscience*, *32*(25), 8443-8453. https://doi.org/10.1523/jneurosci.5069-11.2012

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311-327. https://doi.org/10.1111/lnc3.12151

Strauss, M., Sitt, J. D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., & Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. *Proc. Natl. Acad. Sci. U.S.A.*, *112*(11), E1353-E1362. https://doi.org/10.1073/pnas.1501026112

Strevens, P. (1960). Spectra of Fricative Noise in Human Speech. *Language and Speech*, *3*(1), 32-49. https://doi.org/10.1177/002383096000300105

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, *150*(3700), 1187-1188. https://doi.org/10.1126/science.150.3700.1187

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73-108. https://doi.org/10.1016/s0010-0277(99)00021-9

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language*, *184*, 32-42. https://doi.org/10.1016/j.bandl.2018.06.006

Trine, A., & Monson, B. (2020). Extended High Frequencies Provide Both Spectral and Temporal Information to Improve Speech-in-Speech Recognition. *Trends in Hearing*, *24*, 233121652098029. https://doi.org/10.1177/2331216520980299

Van Hedger, S. C., & Johnsrude, I. S. (2022). Speech Perception Under Adverse Listening Conditions. In L. L. Holt, J. E. Peelle, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Speech Perception* (pp. 141-171). Springer International Publishing. https://doi.org/10.1007/978-3-030-81542-4_6

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *The Journal of Neuroscience*, *32*(11), 3665-3678. https://doi.org/10.1523/jneurosci.5003-11.2012

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, *41*(3), 262-275. https://doi.org/10.3758/BF03208224

Wikse Barrow, C., Włodarczak, M., Thörn, L., & Heldner, M. (2022). Static and dynamic spectral characteristics of Swedish voiceless fricatives. *The Journal of the Acoustical Society of America*, *152*(5), 2588-2600. https://doi.org/10.1121/10.0014947

Winn, M. (2024). *Praat scripts [computer program]* http://www.mattwinn.com/praat/Make_SSN_from_LTAS_selected_sounds.txt

Yeend, I., Beach, E. F., & Sharma, M. (2019). Working Memory and Extended High-Frequency Hearing in Adults: Diagnostic Predictors of Speech-in-Noise Perception. *Ear and Hearing*, *40*(3), 458-467. https://doi.org/10.1097/aud.0000000000000640

Zatorre, R. (2024). *From Perception to Pleasure: The Neuroscience of Music and Why We Love It*. Oxford University Press. https://doi.org/10.1093/oso/9780197558287.001.0001

Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Joint Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-90055-70-0

9 789190 055700