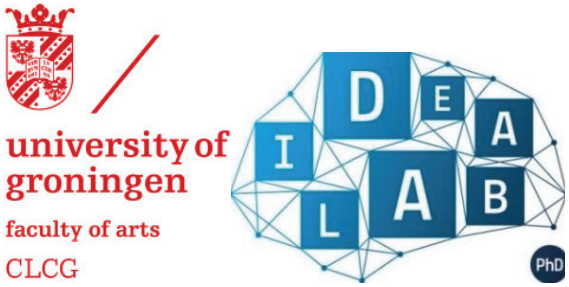
Recognition and cortical haemodynamics of vocal emotions—an fNIRS perspective

Ryssa Moffat



The research reported in this thesis has been carried out under the auspices of the Center for Language and Cognition Groningen (CLCG), the Behavioral and Cognitive Neuroscience (BCN) of the University of Groningen, and the International Doctorate for Experimental Approaches to Language And Brain (IDEALAB) of the Universities of Groningen (NL), Newcastle (UK), Potsdam (DE) and Macquarie University, Sydney (AU).

Publication of this thesis was financially supported by the Graduate School of Humanities (GSH).

Groningen Dissertations in Linguistics 214

Cover: iStock

Printed by Ipskamp Printing (NL) - www.ipskampprinting.nl

Recognition and cortical haemodynamics of vocal emotions— an fNIRS perspective

PhD thesis

to obtain the joint degree of PhD

at the University of Groningen, University of Potsdam, Macquarie University
and
Newcastle University

on the authority of the

Rector Magnificus of the University of Groningen Prof. C. Wijmenga,
President of the University of Potsdam, Prof. O. Günther,
Deputy Vice Chancellor of Macquarie University, Prof. S. Bruce Downton,
and Vice Chancellor of Newcastle University, Prof. Ch. Day

and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 2 June 2022 at 9:00 hours

by

Ryssa Evelyn Moffat

born on June 11th, 1993
in Otonabee, Canada

Supervisors
Prof. D. Başkent
Prof. D. McAlpine

Co-supervisors
Dr. L. van Yper
Dr. R. Luke

Assessment committee
Prof. H. Bortfeld
Prof. F. Dick
Prof. Y.R.M. Bastiaanse
Prof. D. Howard

Acknowledgements

I am very grateful to the following people for being involved in my PhD journey: First, I thank Prof David McAlpine, Dr Lindsey van Yper, Dr Robert Luke, and Prof Deniz Başkent for supervising me, sharing your wealth of expertise and guiding me from the conception of the project through to the submission of this thesis.

For conceiving of and coordinating IDEALAB, checking up on my progress and wellbeing, I thank Prof Lyndsey Nichols, Prof Roelien Bastiaanse, Prof David Howard and Prof Barbara Höhle. For navigating with me through the administrative mire of a multi-university degree, Lesley McKnight, Katie Webb, Alice Pomstra, Anja Papke, and Marijke Wubbolts, and to Joelle Jagersma for the various Dutch translations required along the way.

I thank Dr Anne van der Kant, Dr Romy Räling, and Prof Isabelle Wartenburger for a thorough introduction to fNIRS in the University of Potsdam's BabyLAB. Dr Jörg Buchholz for sharing his expertise in recording speech stimuli and the recording equipment. Elise Tobin for her portrayal of vocal emotions used as stimuli, and the MQ Phonetics & Phonology Lab for assistance designing stimuli.

For sharing programming and signal processing expertise, I thank Dr Jaime Undurraga, Dr Mathieu Recugnat, and Juan Mucarquer. Dr Andrew Etchell for assistance programming experiments. For technical support in the fNIRS lab and many borrowed technical bits and bobs, the Research and Innovation specialists, Craig Richardson, Marcus Ockenden, and Wendy Tham. Dr Peter Petocz, Louis Klein and Maria Korochkina for generously sharing your statistical prowess.

For the company and friendship, I thank the wonderfully approachable members of the Department of Cognitive Science at Macquarie University and all of the IDEALAB gang. Particular thanks to those who made my daily activities at Macquarie University a joy, through lively sound-boarding and highly effective co-procrastination, Dr Amanda Fullerton, Hannah White, Kurt Shulver, Louis Klein, Dr Kelly Miles, Sara Hjortborg, Maria Korochkina, JP Faundez, and Juan Mucarquer for being excellent sounding-boards, enabling appropriate amounts of procrastination.

Finally, enormous thanks to my family–Astrid, Dave, Keiton, and Skye, Grandma Doris– for picking up the phone at all hours. And to Dennis Rickers, Julia Detheridge, Lyle Halliday, Amy Whyte and the Turtle Hunters Ocean Swimming group, for their constant encouragement through this process.

Alongside the people named above, so many more generous souls in Potsdam, Groningen, Sydney and beyond enriched my PhD journey. Thank you all!

Table of contents

Table of figures

Table of tables

List of acronyms

BOLD blood oxygen level dependent
CI cochlear implant
CW continuous wave
DPF differential pathlength factor
EEG electroencephalography
F0 fundamental frequency
fMRI functional magnetic resonance imaging
fNIRS functional near infrared spectroscopy
GLM generalised linear model
GLMM generalised mixed effect model
HA hearing aid
HbO oxygenated haemoglobin
HbR deoxygenated haemoglobin
HbT total haemoglobin
HRM haemodynamic response magnitude
IFG inferior frontal gyrus
LME linear mixed effect model
MBLL modified Beer-Lambert law
MEG magnetoencephalography
MFG middle frontal gyrus
NH normal hearing
NIR near-infrared
PET positron emission tomography
PVF partial volume factor
ROI region of interest
SCI scalp coupling index
STC superior temporal cortex
STG superior temporal gyrus
TDDR temporal derivative distribution repair

Chapter 1| Introduction

1.1 What are emotions?

Over the past century, more than one hundred definitions have been proposed for the term 'emotion' (Plutchik, 2001; Scherer, 2005). Among these, there is a consensus that emotion is the subjective experience of the body's physiological response to a stimulus (e.g., Damasio, 1994; Frijda, 1986; Scherer, 2005). Scherer (2005) describes a model of emotion with five components: a cognitive component which evaluates the stimulus, a neurophysiological component that creates physical symptoms, a motivational component that prepares a reaction, a motor expression component that transmits a facial or vocal reaction, and a subjective feeling component which monitors the internal experience. When an internal or external stimulus initiates a synchronised change in all of these components, an emotion is experienced. Should a vocalisation result from the experience, it will characterise the experience and can be referred to as a vocal emotion. In general, vocal emotions contain sufficient emotional information for a listener to infer the experienced emotional state from the acoustic signal.

The five components in Scherer's (2005) model combined with the infinite number of unique internal and external stimuli explain the heterogeneity of the emotions that humans experience. This heterogeneity is at the centre of the debate on the classification of emotions, which yields two main accounts: a dimensional and a discrete account. Dimensional accounts organise emotions along two or more continua. Russell (1980) posited that all emotional states could be described using valence (positive to negative) and arousal (high to low). In this account, *happy* is high arousal and positive valence, while *angry* is high arousal and negative valence (Figure 1.1). Additional dimensions, including activation, intensity, potency, expectedness, effort, pleasantness, and submissiveness have been proposed (Fontaine et al., 2007; Goudbeek & Scherer, 2010; Russell & Mehrabian, 1977; Smith et al., 1985; Watson & Tellegen, 1999). As such, the dimensional account offers a tool for a rich characterisation of the spectrum of human emotions.

In comparison, the discrete account considers different emotional states (i.e., happy and sad) to be sufficiently distinct to be categorised as discrete emotions. In discrete accounts, a reduced number of emotions are proposed as 'basic universal emotions', which coincide with evolutionary functions that support survival (Ekman, 1992; Griffiths, 2004; Izard, 2007; Ortony & Turner, 1990; Prinz, 2004). A common criticism of the discrete account is that it does not consider the variability observed in the experience of emotion. Despite this shortcoming, the discrete account provides functional labels, which are necessary to obtain behavioural and objective neural measures of vocal emotion recognition accuracy.

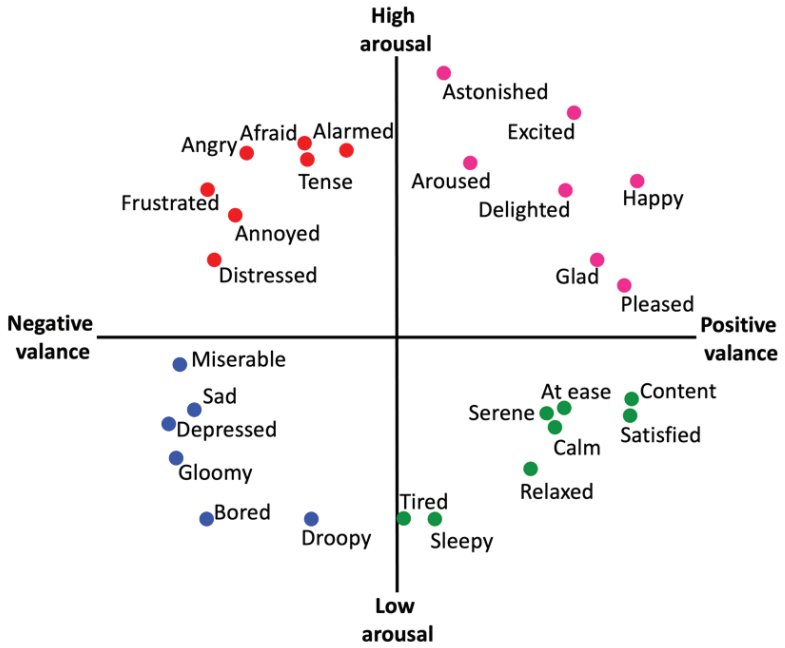


Figure 1.1 Circumplex model of affect. Emotions are categorised along continua of valence and arousal. Figure adapted from Russell (1980).

1.1.1 Recognition of emotions in speech

By evaluating the subjective feeling of an emotional state, the experiencer may recognise the felt emotion, whether explicitly or implicitly. Other individuals may perceive the emotion signalled by the experiencer's facial, vocal, and/or bodily motor expressions. To recognise the emotion, the perceiver must extract sufficient information from the perceived motor expression and match this information to existing representations stored in semantic and procedural memory (Dricu & Frühholz, 2020). If the extraction or matching of emotional information to an existing representation is unsuccessful, the perceiver will fail to recognise the emotional state. Deficits in emotion recognition can render social interactions more challenging (Hall et al., 2009; Schmid Mast & Hall, 2018), and have negative consequences for social well-being (Carton et al., 1999; O'Connell et al., 2021; Shimokawa et al., 2001; Simcock et al., 2020) and quality of life (Luo et al., 2018; Schorr et al., 2009). Moreover, deficits in emotion recognition can negatively impact emotional development in childhood (Netten et al., 2015; Rieffe & Terwogt, 2000; Wiefferink et al., 2012) as well as academic (Agnoli et al., 2012; Halberstadt & Hall, 1980) and career success (Kranefeld & Blickle, 2021; Momm et al., 2015; Scherer & Scherer, 2011).

In certain settings, access to the cues from the visual (face and body) channel conveying

emotional information is limited—for example, through a telephone conversation, or in a darkened room. In these situations, emotional information is perceived through the auditory (voice) channel. Recognition of vocal emotions, i.e., emotional states conveyed in the voice, develops throughout childhood, reaches adult-like maturity in early teenage-hood (Brosgole & Weisman, 1995; Morningstar et al., 2018; Nagels et al., 2020), and seems to decline with advancing age (Christensen et al., 2019; Lambrecht et al., 2012; Paulmann et al., 2008; Schmidt et al., 2016; Zinchenko et al., 2018). Healthy adults with normal hearing (NH) recognise vocal emotions with ease (Cannon & Chatterjee, 2019; Chatterjee et al., 2015; Demenescu et al., 2014; D. House, 1994; Jiang et al., 2015; Juslin & Laukka, 2001; Luo et al., 2007; Metcalfe, 2017; Paulmann et al., 2008; Tinnemore et al., 2018; Van Bezooijen et al., 1983).

Recognition of vocal emotions is known to present a challenge for some clinical populations. Difficulty recognising emotions in speech has been observed in populations diagnosed with Asperger's syndrome (Lindner & Rosén, 2006; Mazefsky & Oswald, 2007), autism spectrum disorders (Philip et al., 2010; Tobe et al., 2016), and alexithymia (Goerlich-Dobre et al., 2014), schizophrenia (Corcoran et al., 2015; Gold et al., 2012; Simpson et al., 2013), unipolar depression (Koch et al., 2018), social anxiety (McClure & Nowicki, 2001), Parkinson's disease (Breitenstein, Van Lancker, et al., 2001; S. L. Buxton et al., 2013; Péron et al., 2012), disordered eating (Kucharska-Pietura et al., 2003), and alcoholism (Kornreich et al., 2013). In these populations, assuming an listener has normal hearing, the acoustic signal transmitted to the listener would be intact, and the difficulties then likely stem from a non-auditory neurally-based perception or processing impairment (Frühholz & Staib, 2017).

Individuals with hearing loss, who use cochlear implants (CIs) or hearing aids (HAs) also seem to have difficulties in recognising vocal emotions (Everhardt et al., 2020; Goy et al., 2018; Jiam et al., 2017). The prevalence of hearing impairment increases with advancing age (Lin et al., 2011; Wilson & Tucci, 2021), which in turn is linked to reduced recognition of vocal emotions (Christensen et al., 2019; Lambrecht et al., 2012; Paulmann et al., 2008; Schmidt et al., 2016; Zinchenko et al., 2018). Recently, Christensen et al. (2019) demonstrated that age and hearing impairment influence recognition of vocal emotions separately, meaning that for adult hearing-impaired listeners, the deficit can be attributed to reduced acuity in decoding the acoustic signal, resulting from damage to the hair cells (Bernstein & Oxenham, 2006; Moore, 1995; Plack, 2018), compounded by signal degradations that may result from the CI or HA signal transmission. Because the deficit originates from a degradation of the acoustic signal or decoding thereof, preliminary investigation of vocal emotion recognition in degraded listening conditions can be used to gain a better understanding of how NH and hearing-impaired listeners utilise acoustic cues to recognise vocal emotions behaviourally and to in-

vestigate the neural representations of vocal emotions associated with successful and unsuccessful vocal emotion recognition. This thesis investigates these aspects of vocal emotion recognition in NH listeners to lay the groundwork for similar investigations in CI and HA listening.

1.1.2 Acoustic characteristics of vocal emotions

Listeners extract emotional meaning in speech from emotional prosody (Banse & Scherer, 1996; Frick, 1985; Juslin & Laukka, 2003). Features conveying cues to emotional prosody can be quantified as acoustic features, according to their physical properties, or perceptual features, as they are perceived by a listener.

The primary acoustic features conveying acoustic cues to vocal emotion recognition are the fundamental frequency (F0), intensity, and speech rate (Frick, 1985; Juslin & Laukka, 2001; Murray & Arnott, 1993; Scherer et al., 1991). F0 is a measure of the rate of vibration of the vocal folds and is the acoustic correlate of voice pitch. The trajectory of F0 across an utterance, also referred to as the pitch contour, is commonly reported to be the strongest indicator of the emotional category conveyed in speech (Banse & Scherer, 1996; Globerson et al., 2013; Hammerschmidt & Jürgens, 2007; Leitman et al., 2010; Metcalfe, 2017; Mozziconacci, 1998; Patel et al., 2011; Pell, 1998; Rodero, 2011; Scherer et al., 2003; Schmidt et al., 2016). Intensity can be quantified as the sound pressure level (SPL) of the speech signal, an acoustic correlate of loudness. Speech rate, the number of syllables pronounced per second, is an acoustic counterpart of tempo. Intensity and speech rate also vary sufficiently in mean value and variability between emotions to contribute to the acoustic profiles of vocal emotions (Banse & Scherer, 1996; Breitenstein, Lancker, et al., 2001; Pakosz, 1983; Scherer & Oshinsky, 1977). Voice quality, i.e., tense and lax sounding voice, is a perceptual feature with many acoustic correlates, (such as the distribution of power in higher and lower frequencies and the relative distance between formants, i.e., broad spectral maxima in the speech signal; Gobl & Ní Chasaide, 2003; Laukkanen et al., 1997; Waaramaa & Leisiö, 2013). Other acoustic features contributing to the perception of vocal emotion include the precision with which utterances are articulated, pause structure, onset of phonation in voiced speech sounds, and duration of stressed and unstressed syllables (Carlson et al., 1992; Laukka et al., 2005; Mozziconacci & Hermes, 2000; Murray & Arnott, 1993).

With respect to dimensions in vocal emotions, the consensus is that arousal ratings are most clearly reflected in the acoustic signal (Bachorowski, 1999; Goudbeek & Scherer, 2010; Patel et al., 2011; Rilliard et al., 2018; Sauter et al., 2010). The relationships described between dimensions and discretely labelled vocal emotions in the literature are inconsistent (Bachorowski, 1999; Goudbeek & Scherer, 2010; Laukka et al., 2005; Rilliard et al., 2018; Sauter et al., 2010), likely due to methodological differences

such as the dimensions and emotions under investigation, and whether the utterances conveying vocal emotions were spontaneous or acted. Nonetheless, the findings can be broadly summarised as follows: Vocal emotions rated as having higher arousal (e.g., anger) are associated with higher F0 mean and variability, higher intensity mean and variability, slower speech rate, and tense voice quality (Bachorowski, 1999; Goudbeek & Scherer, 2010; Laukka et al., 2005). Emotions with more positive valence (e.g., happiness), are associated with low mean F0, large F0 variability, low mean intensity and low intensity variability, fast speech rate, and lax voice quality (Bachorowski, 1999; Goudbeek & Scherer, 2010; Laukka et al., 2005). Sadness, rated as negative valence and lower arousal, is associated with a lower, and less variable, mean intensity (Bachorowski, 1999; Goudbeek & Scherer, 2010).

Listeners' abilities to successfully recognise vocal emotions in foreign languages suggests that acoustic characteristics of vocal emotions are, at least in part, universal, providing support for the discrete account (Bryant & Barrett, 2008; Jürgens et al., 2013; Koeda et al., 2013; Laukka et al., 2013; Sauter, 2013; Scherer et al., 2001; Thompson & Balkwill, 2006; Van Bezooijen et al., 1983; Waaramaa & Leisiö, 2013). However, evidence also exists suggesting that acoustic characteristics of vocal emotions are culture-dependent (Bryant & Barrett, 2008; Jürgens et al., 2013; Koeda et al., 2013; Sauter, 2013; Scherer et al., 2001; Van Bezooijen et al., 1983). Even at the level of a single speaker, the acoustic characteristics of basic emotions vary substantially (e.g., when happiness is sub-divided into finer categories of contentment and joy, Scherer 1986). Laukka et al. (2012) provide an explanation reconciling the considerable variability in the acoustic characteristics of vocal emotions and the discrete account: Expressions of vocal emotions are 'ideal-based goal-derived', meaning that while the prosodic form of an expressed emotion can vary greatly, listeners are able to associate the expression with a prototype. This also explains why acted vocal emotions tend to be more acoustically prototypical than spontaneous vocal emotions (e.g., Laukka et al., 2012; Wilting et al., 2006). In this thesis, emotions will be referred to with discrete labels, the reasoning being purely pragmatic: Although the dimensional account reflects the complexity and diversity of emotional experiences, discrete labels facilitate the acquisition of the behavioural accuracy scores with which vocal emotion recognition can be assessed.

1.1.3 Acoustically degraded speech and recognition of vocal emotions

In individuals with sensorineural hearing loss, at moderate or higher levels of hearing loss, the damage to inner and outer hair cells can reduce temporal and spectral acuity (Bernstein & Oxenham, 2006; Moore, 1995; Plack, 2018). To mitigate sensory loss, individuals may use therapeutic devices such as hearing aids (HAs), which amplify sounds, and in severe cases of hearing loss, may also compress the intensity range to

fit the limited dynamic acoustic range, or cochlear implants (CIs), which convert an acoustic signal to a channelised electrical signal with a heavily compressed intensity range and limited spectral resolution, to match the electric stimulation limitations of the nerve (Başkent et al., 2016; Plack, 2018). As a result, cues related to F0 are weakly transmitted in the speech signal delivered by modern CIs (Başkent et al., 2016; Chatterjee & Peng, 2008; Everhardt et al., 2020; Plack, 2018). While HAs can transmit the low frequencies from which F0 can be extracted, HA-related factors including listeners' individual amplification settings, microphone placement, and the resulting lack of pinna filtering, may alter the overall shape of the spectrum and the temporal properties of the signal. Further, the dynamic range of intensity may also be reduced due to the need to match the reduced dynamic range resulting from hearing loss (Lesica, 2018). These collectively could alter the prosodic cues related to vocal emotions (Goy et al., 2018). CI and HA users have difficulty identifying emotions in speech (e.g., for HA: Goy et al., 2018; Most & Aviner, 2009; Waaramaa et al., 2018, and for CI: Chatterjee et al., 2015; Everhardt et al., 2020; Gilbers et al., 2015; Jiam et al., 2017; Luo et al., 2007; Most & Aviner, 2009; Nakata et al., 2012; Pak & Katz, 2019; Panzeri et al., 2021; Pereira, 2000; Peters, 2006; Waaramaa et al., 2018), likely resulting from the combination of an altered auditory perceptual system (due to hearing loss) and the altered signal delivered by the hearing devices (matched to the listener's hearing loss profile).

The categorical invariance in listeners' perception of acoustically variable speech is a pervasive phenomenon in speech perception (e.g., Holt & Lotto, 2008). Unsurprisingly, this also applies to vocal emotions (Laukka et al., 2012). Toscano & McMurray (2010) propose that listeners cope with the acoustic variability by weighting acoustic features according to the usefulness of the cues they provide, i.e., listeners adapt to altered speech signals by affording less weight to uninformative acoustic cues and more weight to informative acoustic cues. According to the weighting-by-reliability hypothesis (Toscano & McMurray, 2010), listeners may rely more on other acoustic cues to compensate for any reduced reliability, or usefulness, of F0 as a cue for processing vocal emotion, although the evidence is inconclusive. To examine this, studies have presented CI listeners and NH listeners with full-spectrum and noise-vocoded vocal emotions respectively, and manipulated intensity and speech-rate cues (Chatterjee et al., 2015; Everhardt et al., 2020; Gilbers et al., 2015; Luo et al., 2007; Metcalfe, 2017). By presenting vocal emotions with normalised and intact overall intensity, Luo et al. (2007) found that CI and NH listeners can make use of overall intensity cues to recognise vocal emotions. Hegarty & Faulkner (2013) report that CI listeners are not able to make use of intensity information, making use of duration information instead to recognise emotions. Gilbers et al. (2015) normalised each overall intensity and speech rate cues separately and found no evidence for compensatory use of cues from acoustic features in NH or CI listeners. Additional evidence that listeners can reweight acoustic features

based on their usefulness comes from studies examining question/statement intonation and sentence stress—which are primarily conveyed by the pitch contour (P. Warren, 1999). NH listeners who heard noise-vocoded CI simulations, and CI listeners who listened to full-spectrum speech, made use of intensity (Marx et al., 2015; Meister et al., 2011; Peng et al., 2012, 2009) and duration (Peng et al., 2012) information to recognise intonation patterns and sentence stress. When vocal emotions and aspects of linguistic prosody are considered together, the evidence suggests that listeners faced with uninformative F0 cues rely more heavily on intensity cues than timing cues, such as speech rate and duration.

1.1.4 Neuroimaging measures of vocal emotion recognition

Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG) and magnetoencephalography (MEG), enable investigation of neural mechanisms contributing to the processing of vocal emotions. A model proposed by Frühholz et al., (2016) synthesises findings from neuroimaging techniques—mainly fMRI and PET—to describe the neural network involved in processing vocal emotion (Figure 1.2): The processing of emotion conveyed in speech is initiated by the transmission of the acoustic signal from the peripheral auditory system to the amygdala and the auditory cortex, where the acoustic signal is decoded and an auditory percept is formed. The auditory cortex decodes changes in the signal over time (Ethofer et al., 2012; Frühholz & Grandjean, 2013a), while the amygdala contributes a valence appraisal (Frühholz et al., 2014). Moreover, a functional differential of the decoding of slower and faster signal components is thought to exist, in which faster components, such as phonetic characteristics of speech sounds, are decoded by the left auditory cortex and slower components, such as F0 contours, are decoded by the right (Flinker et al., 2019; Meyer et al., 2004; Poeppel, 2003).

The percept formed by the auditory cortex and amygdala is then subjected to higher-level processes in the frontal brain regions including the medial frontal cortex, the orbitofrontal cortex, and the inferior frontal gyrus. The medial frontal cortex is suggested to evaluate and appraise the percept (Ethofer, Kreifelts, et al., 2009), the orbitofrontal cortex may enrich that percept with learned reward/punishment information (Kringelbach & Rolls, 2004), and the inferior frontal gyrus prepares a response while categorising the emotion (Frühholz & Grandjean, 2013b). These higher-level processes are suggested to be mediated by the insula (Kotz et al., 2013). The hippocampus is believed to facilitate access to episodic memory (Frühholz et al., 2014) and the basal ganglia to decode rhythm information (Grahn & Brett, 2007). Finally, the cerebellum may initiate automatic motor responses to vocal emotions and encode basic rhythmic information (Schwartze & Kotz, 2013).

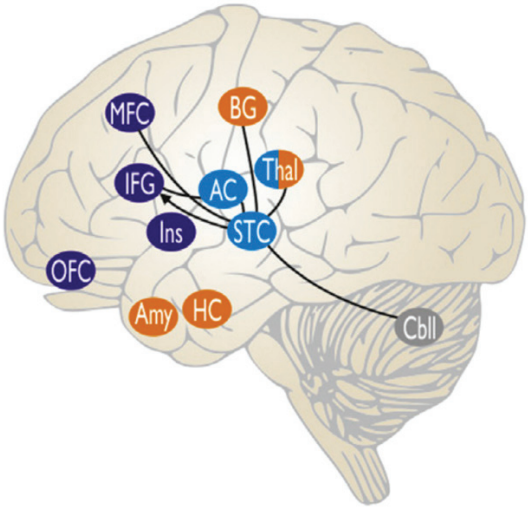


Figure 1.2. Schema of the network for processing emotional prosody. Light blue components belong to the core auditory processing network; AC=auditory cortex, STC=superior temporal cortex, Thal=thalamus. Dark blue components belong to fronto-insular regions; MFC=medial frontal cortex, IFG=inferior frontal gyrus, OFG=orbitofrontal cortex, Ins=insula. The orange components belong to the supporting subcortical network; Thal=thalamus, BG=basal ganglia, Amy=amygdala, HC=hippocampus. Cbll=cerebellum is grey, denoting a supporting role in processing vocal emotions. Figure adapted from (Frühholz, Trost, et al., 2016).

Frühholz et al.'s (2016) model describes the network of brain regions composing the ventral auditory perceptual stream (i.e., the 'what' stream), but not those included in the dorsal stream (i.e., 'where/how' stream; Arnott et al., 2004; Belin & Zatorre, 2000; Rauschecker, 2012). The ventral stream, from the primary auditory cortex to the broader superior temporal cortex and the frontal regions, is responsible for the extraction of acoustic features (Arnott et al., 2004), the generation of a percept, and its appraisal drawing upon semantic memory representations (Dricu & Frühholz, 2020). Positive correlations between activity in the bilateral superior temporal gyri and mean F0, mean intensity, and duration offer support for the function of this pathway (Wiethoff et al., 2008). The dorsal pathway, linking the primary auditory cortex with prefrontal and parietal regions, draws upon procedural memory traces to evaluate (Dricu & Frühholz, 2020) or mirror (J. E. Warren et al., 2006) the perceived audio motor sequence as extracted from spectral changes in the speech signal (Belin & Zatorre, 2000). The functionality of the dorsal stream applies to audition in general rather than specifically to vocal emotions, providing a potential explanation for why it is not described in depth in Frühholz et al.'s (2016) model.

Frühholz et al.'s (2016) model provides a framework of the neural networks enabling successful processing of vocal emotions in normal-hearing adults with typical cognition.

This model serves as a robust point of reference for investigations into the neural mechanisms which may explain the deficit in recognising vocal emotions in the clinical populations described above. See Frühholz & Staib (2017) for a review pertaining to autism, schizophrenia, and mood disorders. Further, and consistent with recent fMRI studies (e.g., Giordano et al., 2021), the model reconciles the discrete and dimensional accounts of emotion classification: Subcortical structures, such as the amygdala may be more sensitive to dimensions like valence, while frontal and temporal brain regions are involved in categorisation of perceived emotions into functional categories (Frühholz, Trost, et al., 2016).

Personality and cognitive characteristics may also explain variability among NH listeners in the patterns of neural activity evoked by vocal emotions. Concerning gender, fMRI studies report that men show increased activation to vocal emotions compared to women in the inferior frontal gyri (Beaucousin et al., 2011), or right middle frontal gyrus (D. Wildgruber et al., 2002), plausibly indicating increased attention and semantic processing (Beaucousin et al., 2011). Using fMRI, increased neuroticism has been associated with increased activity in the somatosensory cortex during vocal emotion processing, suggesting increased reliance on the dorsal pathway (Brück et al., 2011), and increased activation of the insula during the processing of vocal emotions has been associated with stronger empathetic ability (Sachs et al., 2018).

Little is known about the neural activity associated with processing vocal emotions with uninformative, i.e., acoustic cues. While neural processing of acoustically degraded vocal emotions is poorly documented, various EEG studies have investigated vocal emotion processing in CI listeners, a population in which the F0 cues important for emotion recognition are highly compromised (Agrawal et al., 2012, 2013; Deroche et al., 2019). Deroche et al. (2019), for instance, found enhanced late evoked potentials (600–850ms post-stimulus onset) in NH, but not CI, listeners for emotional relative to unemotional vocalisations, indicating reduced neural differentiation between emotional and unemotional speech in CI listeners. Cartocci et al. (2021) also reported that reduced emotion recognition accuracy in paediatric CI listeners is associated with increased right hemisphere activity in the gamma band—proposed to reflect increased effort to extract emotional meaning. These EEG studies overcome the electrical artefacts in the EEG signal typically caused by electrical interference from the CI device. Further, they exploit EEG's strong temporal resolution to distinguish between neural decoding of the acoustic signal and appraisal of vocal emotions, revealing that increased bottom-up processing is required to decode the degraded speech signal and that CI listeners exhibit reduced appraisal of the conveyed emotion. However, neural activity evoked by acoustically degraded emotions remains relatively unstudied, as does the relationship between individual listeners' abilities to recognise vocal emotions in degraded speech and their cortical activity while doing so.

1.2 Functional near-infrared spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) uses low levels of near-infrared (NIR) light to provide an indirect measure of cortical neural activity by quantifying changes in cerebral blood oxygenation (Yücel et al., 2017). fNIRS is a non-invasive, portable technique with better temporal resolution than fMRI and PET, although relatively poor compared to EEG and MEG (Table 1.1; Pinti et al., 2018). fNIRS is well suited to auditory research as it is very quiet, thus not interfering with the auditory stimulation (S. C. Harrison & Hartley, 2019). Being light-based, fNIRS is suitable for use with ferrous metals (i.e., HAs, CIs, pacemakers or dental appliances). Further, the recorded signal is not susceptible to electrical noise (e.g., from HAs, CIs or pacemakers; Bortfeld, 2019; Saliba et al., 2016), like EEG and MEG (e.g., Undurraga et al., 2021) and is robust to movement compared to other neuroimaging techniques (Table 1.1). Its main drawbacks, however, are limited penetration depth (~15 mm into the head, 5–8 mm into the cortex; Huppert, 2016; Strangman et al., 2013) and lower spatial resolution compared to fMRI, PET and MEG. Moreover, fNIRS cannot be used to generate structural images (Pinti et al., 2020) and is more reliable at the group than the individual-participant level (Wiggins et al., 2016).

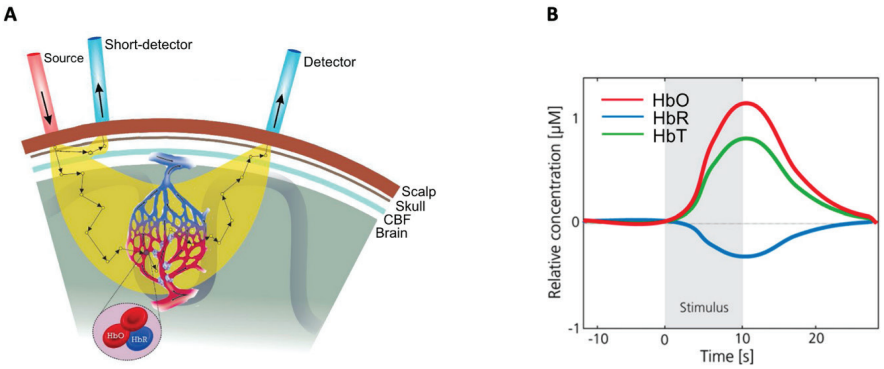


Figure 1.3. Schema of fNIRS measurement and example haemodynamic response. A) Schema of fNIRS recording technique, in which NIR light is continuously emitted from the sources and passes through the scalp into the cortex. The photons are diffusely reflected (scattered omnidirectionally), and a certain proportion of the photons are absorbed as they pass through chromophores in haemoglobin. The unabsorbed photons, which scatter along a banana-shaped arc from the source to a detector are detected as they re-emerge. Short-detectors can be used to measure and regress extracerebral concentrations of haemoglobin out of the cortical signal. Figure adapted from Mohammadi-Nejad et al. (2018) under a CC BY-NC-ND 4.0 license. B) Example of haemodynamic response elicited by 10-s stimulus, i.e., time-course of oxygenated (HbO), deoxygenated (HbR), and total haemoglobin (HbT). Of note, the R in HbR refers to the 'reduced' form or state of the haemoglobin (see Muirhead & Perutz, 1963), rather than changes in concentration. Grey area indicates stimulus duration relative to time-course. Figure adapted from Scholkmann et al. (2014).

1.2.1 Haemodynamic responses

In brain-imaging techniques such as fMRI and fNIRS, task-evoked cortical activity is characterised by a 'haemodynamic response' (Figure 1.3B). To record haemodynamic responses, fMRI and fNIRS exploit a physiological phenomenon called neurovascular coupling, which is based on the premise that active neurons metabolise oxygen at an increased rate: To meet the metabolic demands of active neurons, local arterial blood flow is increased through vasodilation (Wolf et al., 2002). The increased local blood flow causes an increase in oxygenated haemoglobin (HbO) and results in increased local blood volume, which in turn reduces the concentration of deoxygenated haemoglobin (HbR; Ferrari & Quaresima, 2012; Huppert et al., 2006). These opposite—negatively correlated—dynamics of HbO and HbR (collectively called chromophores) are characteristic of neural activity (Wolf et al., 2002). While, the blood oxygen level dependent (BOLD) response measured with fMRI is an indirect measure of changes in HbR concentration (Huppert et al., 2006; Logothetis & Wandell, 2004), fNIRS is used to measure haemodynamic responses in both HbO and HbR. The BOLD response and the haemodynamic response measured in HbR with fNIRS are strongly correlated (Huppert et al., 2006).

Ultimately, fNIRS uses NIR light to quantify the presence of HbO and HbR in the cortex, thereby recording haemodynamic responses and providing an indirect measure of cortical activity (Figure 1.3A, 1.3B). Changes in local cortical blood volume can also be quantified using fNIRS, by summing the measured HbO and HbR to calculate total haemoglobin (HbT; Figure 1.3B; Wolf et al., 2002). Importantly, a haemodynamic response can be observed in either chromophore, yet a negative correlation between HbO and HbR provides the strongest evidence that the measured response is of neural origin (Wolf et al., 2002). Accordingly, a consensus has emerged that measures of both chromophores should be reported in fNIRS studies, for the sake of transparency, generalisability, and replicability (Yücel et al., 2021).

Measured haemodynamic responses elicited by motor, visual, and auditory tasks are largest in amplitude over the associated primary motor or sensory cortical areas, decreasing in amplitude further from these areas (e.g., Mushtaq et al., 2019; Plichta et al., 2006, 2011; Shader et al., 2021; Watanabe et al., 2012; Wijeakumar et al., 2012). In general, the concentration of HbO begins to increase and the concentration of HbR begins to decrease 1–2 s after stimulus onset, reaching maximal deflection (peak for HbO, minimum for HbR) ~6–7 s post-task-onset, plateauing for the duration of stimulation and returning to baseline after stimulus-offset (Hong & Nguyen, 2014; Hong & Santosa, 2016; Huppert et al., 2006). A transient decrease in HbO or increase in HbR relative to the baseline may be observed before or after the main stimulus-evoked deflection (e.g., Khan et al., 2020). The initial dip, immediately following stimulus-onset is believed

Table 1.1. Comparison of neuroimaging techniques

	fNIRS	fMRI	PET	EEG	MEG
Measures	HbO, HbR	BOLD (HbR)	Glucose metabolism	Electric potentials	Magnetic fields[a]
Spatial resolution	1-3 cm	0.3 mm	4 mm	5–9 cm	2–3 mm[a]
Penetration depth	5–8 mm into cortex[c]	Whole head	Whole head	Cortex	Whole head
Temporal resolution	Up to 10 Hz	1–3 Hz	<0.1 Hz	>1000 Hz	290–1000 Hz[b]
Robustness to motion	Good	Poor	Poor	Poor	Poor
Loudness	Very quiet	Very loud	Very quiet	Very quiet	Very quiet
Cost	Low	High	High	Low	High
Limiting factors	May cause headaches when many sensors are used	Not suitable for children, participants with ferrous metal implants, claustrophobia	Invasive; exposes participants to low levels of radiation	Electrical devices cause artefacts in signal; may cause headaches	Ferrous metal implants and electrical devices cause artefacts in signal

Notes. fNIRS=functional near-infrared spectroscopy, fMRI=functional magnetic resonance imaging, PET=positron emission tomography, EEG=electroencephalography, MEG=magnetoencephalography. Adapted from Pinti et al. (2018) under a CC BY 4.0 license with information supplemented from [a]Singh (2014), [b]Velmurugan et al. (2014), and [c]Huppert (2016).

to reflect the rate of neuronal oxygen metabolism briefly outpacing the rate of local blood flow (R. Buxton, 2010), whereas the post-stimulus-dip may reflect a slow return of local blood volume or both local blood flow and oxygen metabolism to baseline (E. Y. Liu et al., 2019). With respect to the negatively correlated time courses of the two chromophores, the reduction in HbR may lag a few seconds behind the increase in HbO. Although this lag is not always observed, it may reflect faster changes in blood flow (as indexed by HbO) than blood volume (as indexed by HbR; Huppert et al., 2006; H. L. Liu et al., 2000). Variation in these aspects of response morphology, i.e., time to peak, the symmetry of time courses of HbO and HbR, delay between stimulus-offset and return to baseline may vary as a function of the task domain and the measured brain areas (e.g., Shader et al., 2021; Toronov et al., 2007; Weder et al., 2018). Moreover, the amplitude of measured haemodynamic responses may be larger for cortical regions closer to the scalp, i.e., on gyri rather than sulci, or below thinner regions of the skull (Brigadoi & Cooper, 2015; Haeussinger et al., 2011).

Motor tasks, such as finger-tapping, are well-suited to validating an fNIRS lab setup. They elicit increased concentrations of HbO in primary motor cortex, which remain elevated for the duration of the task (e.g., Rahimpour et al., 2020; Sato et al., 2007). The concomitant reduction in HbR lags 1–2 s behind HbO when compared to a baseline of rest, but not when compared to a baseline involving finger-tapping on the other hand (Boden et al., 2007; Jasdzewski et al., 2003). Additionally, HbR may reach the minimum as late as 15 s post-task-onset, which in turn delays return to baseline (Leff et al., 2011). The haemodynamic responses evoked by finger-tapping can be observed in individual participants (e.g., Huppert et al., 2006), likely due to the location of the primary motor cortex associated with hand and finger movements on the pre-central gyrus (Yousry et al., 1997) which results in a relatively short distance between the location of optodes on the scalp and the brain.

Auditory tasks, such as the listening tasks investigated in this thesis, activate the primary auditory cortex, located on Heschl's gyrus, which is part of the superior temporal gyrus (STG; Dick et al., 2012). The primary auditory cortex is situated ~25 mm below the scalp (Ohnishi et al., 1997), while fNIRS measures reliably from ~15 mm below the scalp (Strangman et al., 2013), meaning that very few of the emitted photons reflected back to the detector will pass through Heschl's gyrus. However, auditory haemodynamic responses can be reliably recorded with fNIRS from the superior temporal gyrus at the group level (STG; Harrison & Hartley, 2019; van de Rijt et al., 2018). In adults, the HbO response generally plateaus for the length of the stimulus, returning to baseline up to 10 s post-offset, (e.g., Lawrence et al., 2018; Plichta et al., 2011). Weder et al. (2018) recently observed that haemodynamic responses evoked by 18-s-long auditory stimuli remained elevated for the duration of the stimulus in the anterior (rostral) part of

the STG, whereas haemodynamic responses in the posterior (caudal) parts of the STG peaked shortly after the stimulus onset and offset. The symmetry of the HbO and HbR time courses evoked by auditory stimuli remains a topic of discussion with some studies reporting a lag in HbR relative to HbO (Quaresima et al., 2012), and others reporting no lag (Mushtaq et al., 2019; D. Zhang et al., 2018).

1.2.2 Continuous-wave fNIRS

During fNIRS recordings, light-emitting sources and light-receiving detectors—collectively called optodes—are positioned snuggly on the participant's scalp with the hair moved aside. With a continuous wave (CW) spectrometer, NIR light is continuously emitted from the sources and passes through the scalp into the cortex. The photons are diffusely reflected (scattered omnidirectionally), and a proportion of the photons are absorbed as they pass through chromophores in haemoglobin. The remaining photons, which scatter along a 'banana-shaped' arc from a source to a detector, are detected as they re-emerge (Figure 1.3A). The attenuation of the light, i.e., the difference between the intensity of the emitted and detected light, is recorded as raw intensity, yielding raw time-series data (Scholkmann, Kleiser, et al., 2014).

To obtain measurements of HbO and HbR, most CW systems, such as the NIRScoutX described in Chapter 2.2, use two NIR wavelengths between 650–950 nm. Although a third chromophore, cytochrome oxidase, can also be measured using wavelengths within this range (Figure 1.4; Zhu et al., 2012), this thesis will investigate changes in HbO and HbR. At 800 nm, HbO and HbR have the same absorption coefficient, i.e., isosbestic point (Figure 1.4). To optimise sensitivity to HbO and HbR, two wavelengths between 695–770 and 830–850 nm are usually used. These wavelengths, within the NIR spectrum, have a low absorption coefficient for the components of the scalp, skull and brain, including haemoglobin, water, lipids, melanin, and collagen (Scholkmann, Kleiser, et al., 2014). This means that the proportion of NIR light absorbed by these tissues is a) sufficient to be measured, and b) small enough to enable the NIR light to penetrate ~15 mm below the scalp (Strangman et al., 2013), equating to the 5–8 most superficial mm of the cortex after accounting for the thickness of the scalp (Huppert, 2016). The depth of penetration of a source-detector pair is approximately equal to half the source-detector separation on the scalp. In adults, a source-detector separation of 25–30 mm yields the best compromise between signal-to-noise-ratio and depth sensitivity, with a spatial resolution of 1–3 cm (Pinti et al., 2020; Quaresima & Ferrari, 2019).

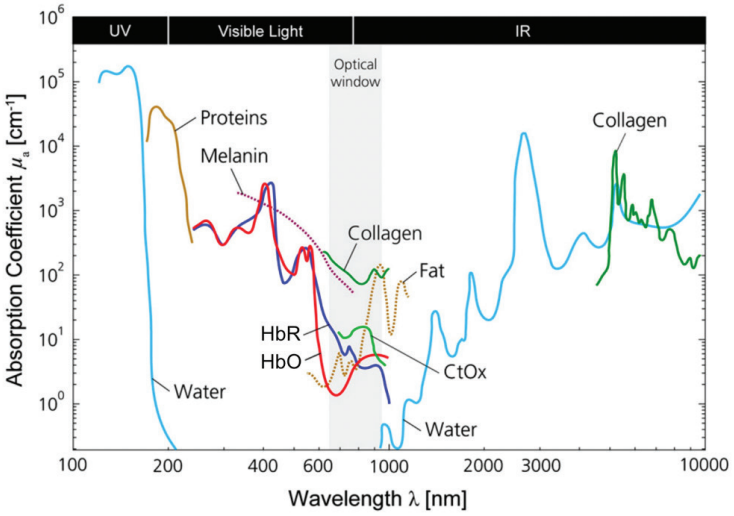


Figure 1.4. Absorption spectra for chromophores found in human tissue. UV=ultraviolet, IR=infrared, HbO= oxygenated haemoglobin, HbR=deoxygenated haemoglobin, CtOx=cytochrome oxidase. Adapted from Scholkmann et al. (2014).

To obtain relative concentrations of HbO and HbR, raw intensity for each wavelength is converted to optical density, then to relative concentrations using the Modified Beer Lambert Law (MBLL; Delpy et al., 1988; Kocsis et al., 2006). The MBLL accounts for a partial pathlength factor, which is the product of the differential pathlength factor (DPF) and the partial volume factor (PVF). The DPF is used to calculate the effective distance travelled by photons as they scatter through biological tissue, which can vary with age, brain region, and wavelength (Duncan et al., 1995; Scholkmann & Wolf, 2013). The PVF is a factor to account for the reduced proportion of emitted photons that pass through cerebral tissue, as opposed to extracerebral tissue (Strangman et al., 2003). CW spectrometers quantify chromophores in relative concentrations based on the assumption of a constant DPF across brain regions, as well as participants (of variable sex and age). As this assumption is known to be incorrect, best practice is to compare evoked haemodynamic activity within a brain region, rather than between regions, and to match participants for sex and age (Scholkmann, Kleiser, et al., 2014).

1.2.3 fNIRS signal components

A common aim in fNIRS studies is to quantify the neural activity evoked by a specific task. However, fNIRS is very sensitive to physiological noise and motion artefacts (Huppert, 2016); measured changes in HbO and HbR resulting from any origin other than neurovascular coupling (Kirilina et al., 2012; R. Saager & Berger, 2008). The raw signal contains extracerebral and cerebral components, both of which can be evoked by

the task, or occur independently. Furthermore, the cerebral signal component contains both neural and systemic signals, while the extracerebral signal component consists mainly of systemic signals (Figure 1.5; Tachtsidis & Scholkmann, 2016). The latter are generated by heart rate (~1 Hz), respiration rate (~0.3 Hz), and blood pressure regulation (i.e., Mayer waves, ~0.1 Hz; Julien, 2006).

Extracerebral sources of physiological noise can be accounted for very effectively with specific channels that only record extracerebral changes in HbO and HbR. These are commonly called short channels and are source-detector pairs separated by ~8 mm (Figure 1.3). With short channels of the optimal (8 mm) length as a regressor, ~95% of the extracerebral signal component in the long-channel signal can be accounted for (Brigadoi & Cooper, 2015). Extracerebral physiological signal components have been demonstrated to be heterogeneous across different scalp regions, but symmetrical across hemispheres (Y. Zhang et al., 2015).

Accounting for physiological noise, beyond short-channel regression, depends on the approach to the statistical analysis (Tak & Ye, 2014). Analyses based on epoched and averaged time-courses of the haemodynamic response apply a variety of signal pre-processing steps including principal component analyses (e.g., Defenderfer et al., 2017; Yücel et al., 2014), or common average reference calculated across channels (e.g., Bauernfeind et al., 2014), in conjunction with band-pass filters (e.g., 0.05–0.7 Hz to remove ~1 Hz cardiac signals and slow drifts). Further signal improvement can be achieved by enhancing the theoretically and empirically supported negative correlation between HbO and HbR (Cui et al., 2010).

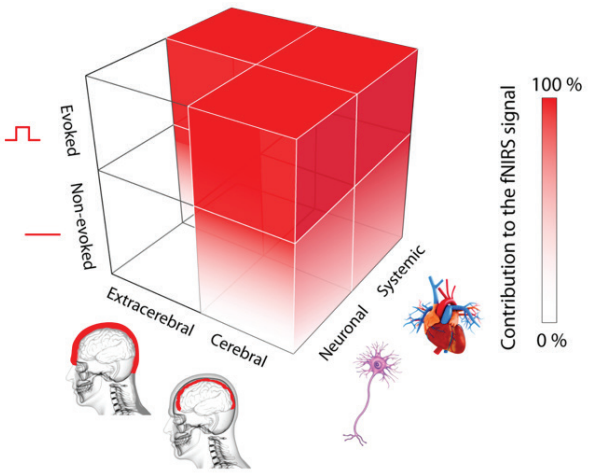


Figure 1.5. Composition of fNIRS signal. Illustration of the origins of fNIRS signal components and their relative contributions. Figure adapted from Tachtsidis & Scholkmann (2016) under a CC BY 3.0 license.

Analyses based on general linear models (GLMs; Huppert, 2016; Luke et al., 2021; Ye et al., 2009) predict the amplitude of each participant's measured haemodynamic response to each stimulus condition, which is modelled using a canonical haemodynamic response function (e.g., Friston et al., 1998; Glover, 1999) convolved with the trial occurrence and length. GLMs assume non-correlated noise, i.e., spherical noise. This assumption is violated by the correlated, non-spherical nature of the physiological—cardiac, respiratory, blood pressure—noise in the fNIRS signal (Huppert, 2016). To account for the correlated noise, an autoregressive noise model can be calculated in the GLM, per channel or for the whole of the data, effectively down-weighting the physiological noise (Barker et al., 2013; Huppert, 2016).

Interestingly, some systemic signals are considered 'noise' in the context of extracting haemodynamic responses from the fNIRS signal, but can still provide valuable information about cerebral processing. For example, heart rate can provide a reliable measure of signal quality; the scalp-coupling index (SCI; Pollonini et al., 2014) calculates the correlation coefficient between the heart rate in the HbO and HbR signals for each channel. A higher index indicates better coupling of the optodes with the scalp. This metric can be used as a criterion for channel rejection.

Finally, fNIRS is also sensitive to motion artefacts, although less so than other neuroimaging techniques. These result from physical movement by the participant, which can alter the contact quality between the optode and the scalp. Changes in pressure of contact or angle of optode relative to the scalp can introduce large deflections in the signal (Orihuela-Espina et al., 2010). In the epoch-averaging approach, avenues for removing motion artefacts include excluding epochs with peak-to-peak amplitudes exceeding a certain cut-off, or employing signal-repairing algorithms (e.g., Cui et al., 2010; Fishburn et al., 2019; Scholkmann et al., 2010). In the GLM approach, where the autoregressive noise model is applied, the addition of the iteratively re-weighted least-squares procedure can repair motion artefacts, reducing the risk of Type I errors (Barker et al., 2013; Huppert, 2016).

1.2.4 Illuminating vocal emotions with fNIRS

Speech-evoked haemodynamic responses can be measured with fNIRS over the bilateral STG (Bortfeld, 2019; Quaresima et al., 2012; van de Rijt et al., 2018). Similarly, fNIRS studies report that vocal emotions evoke activation of bilateral STG, with larger haemodynamic responses in the right hemisphere (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021), consistent with findings obtained with fMRI (e.g., Witteman et al., 2012). With fNIRS, significant cortical activity is reported in left IFG for happy speech (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021), right MFG for sad (Anuardi & Yamazaki, 2019), as well as bilateral IFG (Gruber et al.,

2020), left MFG (Zhen et al., 2021) and bilateral MFG (Sonkaya & Bayazıt, 2018) for angry. Gruber et al.'s (2020) finding of bilateral IFG activity evoked by angry speech is based on a region of interest analysis (ROI; responses averaged across channels, although only in left and right IFG). The remaining fNIRS studies on vocal emotions analyse haemodynamic responses in individual channels (Anuardi & Yamazaki, 2019; Sonkaya & Bayazıt, 2018; Steber et al., 2020; D. Zhang et al., 2018; Zhen et al., 2021); an approach that has been suggested to be less reliable than ROI analyses (Wiggins et al., 2016). fNIRS studies of visual emotion perception (for review see Westgarth et al., 2021), many of which employ channel-wise analyses, are also inconsistent in their findings, reporting activity in a variety of cortical regions.

Interest in the utility of fNIRS for investigating the cortical representations of vocal emotions is relatively recent and the existing studies involving adults report variable patterns of cortical activity (described above, see also Anuardi & Yamazaki, 2019; Gruber et al., 2020; Sonkaya & Bayazıt, 2018; Steber et al., 2020; D. Zhang et al., 2018; Zhen et al., 2021). Before fNIRS can be used to investigate recognition of vocal emotions in populations such as CI users, the technique's sensitivity to the cortical representations of vocal emotions must be determined.

1.3 Aims of this thesis

Considerable scientific effort has been paid to understanding the acoustic profiles of vocal emotions (e.g., Banse & Scherer, 1996; Frick, 1985; Juslin & Laukka, 2003), and listeners' behavioural capacity to recognise emotions in natural and degraded listening situations (e.g., Christensen et al., 2019; Everhardt et al., 2020; Scherer et al., 1991). The outcomes of these studies suggest that substantial variability exists between listeners' individual abilities to recognise different emotions in speech. This variability has yet to be assessed in neuroimaging studies, particularly those that might seek to explain the neural underpinnings of successful and failed recognition of vocal emotions. More specifically, the relationship between the accuracy with which listeners recognise emotions in speech and cortical activity evoked by vocal emotions has yet to be assessed. This relationship is of particular interest in cases where reduced fidelity of acoustic cues (F0, intensity, and speech rate) might lead to reduced accuracy of vocal emotion recognition. The aims of this thesis are:

To develop a set of vocal emotion stimuli with which to obtain behavioural and neuro-physiological measures of vocal emotion processing in natural and degraded speech.

i. To determine, when variation within acoustic features (i.e., F0, intensity, speech rate) is attenuated, rendering acoustic cues to vocal emotions less informative, whether normal-hearing listeners can extract sufficient information from intact cues to recognise vocal emotions.

ii. To determine if fNIRS is sensitive to cortical signatures evoked by discrete vocal emotions conveyed in natural speech.

iii. To determine if cortical activation evoked by vocal emotions, as measured with fNIRS, reflects the accuracy with which normal-hearing listeners recognise emotions in acoustically degraded speech.

1.4 Thesis structure

Chapter 2. General methods
This chapter details the selection of stimulus items and the signal processing procedures applied to degrade the speech signal, followed by an in-depth description of the fNIRS equipment and laboratory, as well as the experimental protocol and data processing pipeline.

Chapter 3. Withholding emotion: Attenuating variation in voice pitch, intensity, and speech rate yields differentially reduced accuracy of vocal emotion recognition
The first experimental chapter investigates NH listeners' abilities to recognise emotions conveyed in speech with uninformative acoustic cues to vocal emotions and to make use of more informative secondary cues. The findings indicate listeners recognise emotions conveyed in natural speech with ease. The accuracy with which listeners recognise emotions is greatly reduced for emotions conveyed in speech with uninformative F0 cues and mildly reduced for emotions conveyed in speech with uninformative intensity and speech-rate cues. The data demonstrate that listeners were not able to use intensity or speech-rate cues to compensate for uninformative F0 cues to vocal emotions.

Chapter 4. Illuminating emotion: Assessing functional near-infrared spectroscopy's sensitivity to discrete vocal emotions
The second experimental chapter assesses the sensitivity of fNIRS to cortical haemodynamic activity evoked by discrete vocal emotions (i.e., *angry*, *happy*, *sad*) conveyed in natural speech. The results indicate that fNIRS is sensitive to haemodynamic activity evoked by speech in the bilateral auditory brain regions, with evidence of a right-lateralisation evoked by emotional speech. Cortical activations did not differ significantly between any discrete vocal emotion and unemotional speech, indicating that fNIRS is

not sensitive to cortical representations of discrete vocal emotions, but rather to more general aspects of auditory processing.

Chapter 5. Deciphering emotion: Cortical responses to vocal emotions with attenuated voice pitch variations as measured by functional near-infrared spectroscopy
The third experimental chapter investigates cortical haemodynamic activity evoked by vocal emotions conveyed in natural speech and speech with uninformative F0 cues. Particular attention was paid to the relationship between listeners' abilities to recognise vocal emotions conveyed in speech with uninformative F0 cues and their cortical activity for vocal emotions. The findings indicate that vocal emotions conveyed in natural speech and speech with uninformative F0 cues evoke haemodynamic activity in the bilateral auditory brain regions, although the measured activity differs significantly from the activity evoked by the control silent condition only in the right hemisphere. A significant association was observed between listeners' abilities to recognise vocal emotions with uninformative F0 cues and their haemodynamic activity in the right auditory brain regions, whereby reduced accuracy is associated with increased haemodynamic activity.

Chapter 6. General discussion
This chapter summarises and synthesises the relevance of the findings from Chapters 3–5, as they pertain to the research aims examined in this thesis.

Chapter 7. Conclusions

Chapter 2| General methods

2.1 Stimulus creation

Speech stimuli conveying vocal emotions were created to investigate vocal emotion recognition and processing in normal-hearing (NH) listeners. To examine NH listeners' usage of the cues conveyed by acoustic features in their recognition of vocal emotions, cues in the main acoustic features conveying emotional information, i.e., fundamental frequency (F0), intensity, and speech rate, were systematically rendered less informative, one or two at a time. Subsequently, stimuli were generated for a) behavioural N-alternative forced-choice tasks and b) block-design functional near-infrared spectroscopy (fNIRS) experiments.

2.1.1 Generation of pseudo-sentences

The speech stimuli consisted of six-syllable sentences with real function-words and rhyming pseudo-content-words, such as "*the ziffox is dorval*" and "*the miffox is borval*" (see Table 2.1). The combination of real function-words and pseudo-content-words served to respect English phonotactics and syntax while eliminating confounds due to emotional associations related to semantic content, e.g., "snake" being associated with negative emotions (Demenescu et al., 2014; Pell et al., 2009; Scherer et al., 1991; D. Zhang et al., 2018). The rhyming pseudo-words allow for speech-sound-specific phonetic information such as vowel length and consonant voicing (i.e., glottal fold vibration) to be held constant between sentences. This is important as vowel length and F0 can be influenced by the surrounding consonants (A. S. House & House, 1961; Whalen et al., 1993); rhyming pseudo-words minimise these differences while ensuring that F0 measurements, taken from voiced segments, are extracted from comparable phonetic environments. Finally, varying voiced consonants only in the pseudo-word-initial position, minimised the influence of the phonetic characteristics of individual speech sounds on comparisons of mean F0, intensity, and speech rate between stimuli and emotions.

The pseudo-words *miffox* and *borval* were designed to meet these criteria. The final consonant cluster /ks/ in *miffox* does not cause elision of the voiced schwa in *is*, /əz/. The /l/ coda of *borval* is voiced, facilitating the identification of the speech offset for each sentence, which is needed to calculate the speech rate. For the pseudo-words *miffox* and *borval*, six additional rhyming pseudo-words were created by replacing the initial voiced consonant with another voiced consonant, yielding a total of seven sentences (Table 2.1).

2.1.2 Recording of sentences

To create stimuli suitable for Australian listeners, one 27-year-old female native speaker of standard Australian English, from Sydney, Australia, was recorded pronouncing the sentences in each 'angry', 'happy', 'sad', and 'neutral/unemotional' prosodies. *Angry*, *happy* and *sad* emotions were selected on the basis that they are a subset of Ekman's basic emotions (1992), each of which occupies a different quadrant in Russell's (1980) dimensional account, as shown in Chapter 1, Figure 1.1: *angry* is high arousal and negative valence, *happy* is high arousal and positive valence, *sad* is low arousal and negative valence. Although *unemotional*/*neutral* is not considered a basic emotion, it was added to facilitate comparisons between emotional and unemotional speech, and may be best described as mid-level arousal and mid-level valence.

During a single recording session in an anechoic chamber, recordings were made using a Røde NT1 microphone with a pop cover (RØDE Microphones, Silverwater, Australia) connected to an RME Fireface UC sound card (RME, Haimhausen, Germany). The sentences were digitally recorded using Adobe Audition (version 13.0.8.43; Adobe Systems, Mountain View, USA) on a personal computer at a sampling rate of 48 kHz (32 bits mono). For each emotion, each of the seven sentences in Table 2.1 was recorded at least four times. The rendition of each item, for each emotion, with the fewest recording artefacts (i.e., mouth noises created by spittle or lips), mispronunciations and other background noises, was selected and saved as a discrete WAV file using Adobe Audition. Each item was manually trimmed at zero crossings from the onset of the first consonant (initial increase in energy and initial glottal pulse) to the offset of the final consonant (final decrease in energy and final glottal pulse) using Praat (version 6.1.16; Boersma & Weenink, 2018).

Table 2.1. Recorded stimulus sentences

Item #	Sentence
1	The ziffox is dorval
2	The biffox is jorval
3	The diffox is morval
4	The niffox is zorval
5	The giffox is lorval
6	The miffox is borval
7	The riffox is gorval

2.1.3 Further adjustments

Further adjustments were made to the selection of recordings with the input of lab members and colleagues with experience in the speech and language field in informal pilot listening sessions. First, to ascertain whether the stimuli categories were behaviourally discriminable, 10 experienced listeners were asked to indicate the emotion conveyed by each stimulus in a 4-alternative forced-choice task (7 sentences for each of 4 emotions, N trials=28). During a first round, the response alternatives were labelled as *angry*, *happy*, *sad* or *neutral*. *Happy* and *sad* were commonly misidentified as *neutral* stimuli (consistent with confusions reported by Paulmann, Pell, & Kotz, 2008; Scherer, Banse, & Wallbott, 2001). In a second round, 10 new experienced listeners completed the same task, except this time *neutral* was re-labelled *unemotional*. With these alternatives, there were almost no confusions between happy and unemotional (rate of *happy* stimulus being misidentified as *unemotional*=0.01), but the confusions between *sad* and *unemotional* remained (rates for *unemotional* stimulus misidentified as *sad*=0.27, *sad* stimulus mislabelled *unemotional*=0.21). Accuracy was near ceiling in all four emotions (*angry* *M*=1, *SD*=0; *happy* *M*=0.94, *SD*=0.1; *sad* *M*=0.80, *SD*=0.18; *unemotional* *M*=0.71, *SD*=0.17) indicating that the stimuli are behaviourally discriminable.

Next, to reduce the number of retained recordings, the five sentences with the best expression of each emotion were identified. Eleven new experienced listeners were asked to judge the relative strength with which the relevant emotion was conveyed in each item. All seven sentences for each emotion were presented on a computer screen at once, and listeners were instructed to play each sentence and order them from 'most to least strongly' conveying the relevant emotion. Listeners could play the sentences as often as they wished. The two sentences for each emotion most frequently ordered as 'least strongly' were discarded, yielding sets of five sentences per emotion (Table 2.2).

Table 2.2. Retained sentences for each emotion

Emotion	Retained sentences	Discarded sentences
angry	1, 2, 4, 5, 6	3, 7
happy	1, 3, 4, 5, 7	2, 6
sad	2, 3, 4, 6, 7	1, 5
unemotional	1, 2, 3, 6, 7	4, 5

2.1.4 Acoustic measures and acoustic analysis

The mean F0 (Hz), mean intensity (dB, relative to the maximum measured root mean square, RMS), and mean speech rate (syllables per second) were measured for each sentence using a Praat script (Figure 2.2; Table 2.3).

To confirm that the emotions are acoustically distinct, i.e., that the emotions differ significantly from one another in at least one acoustic feature, a statistical analysis was performed using R (version 3.6.3; R Core Team, 2020) and the RStudio IDE (version 1.3.959; RStudio Team, 2019). An alpha level of 0.05 was used for all statistical tests. To assess the presence of differences between emotions, a Kruskal-Wallis test was performed for each F0, intensity, and speech rate, revealing significant differences between emotions, i.e., F0 ($\chi^2(3, N=20)=16.07$, $p=0.001$), intensity ($\chi^2(3, N=20)=17.86$, $p<0.001$), and speech rate ($\chi^2(3, N=20)=11.75$, $p=0.008$).

Subsequently, to test for differences between emotions within each acoustic feature (Figure 2.2), post-hoc pairwise Wilcoxon rank sum tests were computed with false discovery rate correction for multiple comparisons (FDR; Benjamini & Hochberg, 1995). There is evidence to support significant differences in F0 between the following pairs: mean F0 is higher for each *angry*, *happy* and *sad* relative to *unemotional* (all $W=25$, $p=0.008$), as well as *angry* relative to *sad* ($W=24$, p=0.008), and *happy* relative to *sad* ($W=23$, $p=0.008$). Mean F0 does not differ significantly between *angry* and *happy* ($W=13$, $p=1$). Mean intensity differs significantly between all pairs (all $W=25$, $p=0.008$). Mean speech rate is significantly higher for *happy* than each *angry*, *unemotional*, and *sad* (all $W=25$, $p=0.008$). Mean speech rate does not differ significantly between *angry* and *unemotional* ($W=6$, $p=0.222$), *angry* and *sad* ($W=8$, $p=0.421$), or *sad* and *unemotional* ($W=13$, $p=1$).

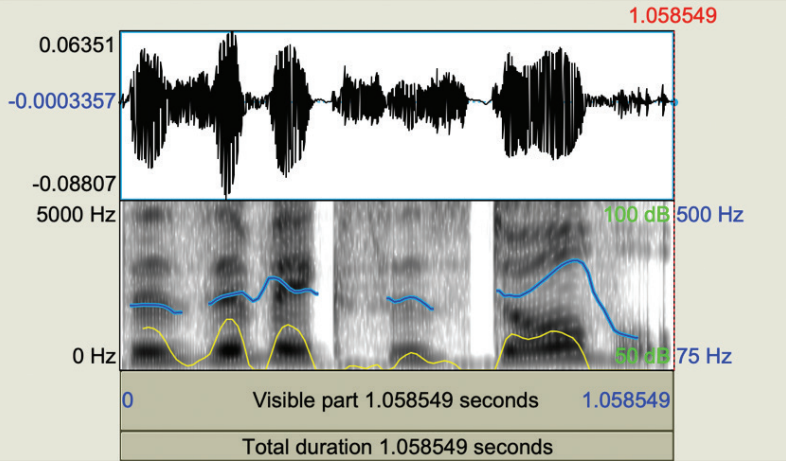


Figure 2.1. Praat analysis window for a *happy* recording (sentence 1). Above, the waveform of the recording with the RMS on the y-axis and time (s) on the x-axis. Below, the spectrogram with frequency (Hz) on the y-axis, and each the F0 (blue line) and the intensity (yellow line) contours overlaid on the spectrogram.

Table 2.3. Mean and standard deviation of each acoustic feature for recordings of each emotion

Emotion	**F0** (Hz)		**Intensity** (dB)		**Speech rate** (syl/sec)	
	M	***SD***	***M***	***SD***	***M***	***SD***
angry	241.47	6.14	-0.99	1.01	4.50	0.07
happy	243.31	10.29	-10.97	0.85	5.57	0.18
sad	215.97	11.44	-17.76	1.44	4.70	0.31
unemotional	167.04	8.37	-15.04	0.58	4.68	0.26
Overall mean	216.95	32.70	-11.19	6.59	4.86	0.47

Note. dB is relative to the maximum measured RMS

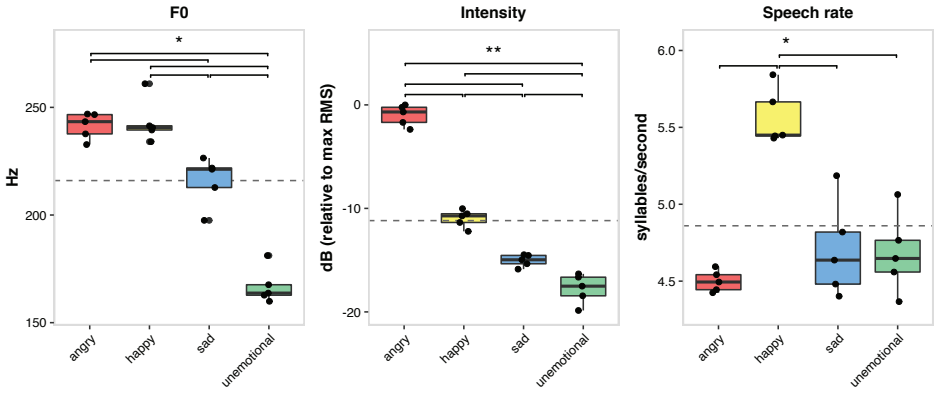


Figure 2.2. Analyses of acoustic feature within each emotion. The dashed line indicates the overall mean across emotions. Significance for Wilcoxon rank sum test, *p<0.05; **p<0.01.

Despite the small number of stimuli and variability within each emotion, each emotion patterns in an acoustically discriminable fashion. The mean values and the relative positions of emotions within each acoustic feature match those previously reported (for F0 and rate, Paulmann & Uskul, 2014; for F0, Gilbers et al., 2015; Pell, 1998, for intensity, Luo 2007, for rate, Most & Aviner, 2009). The direction of differences between emotions is also similar to those previously described for each acoustic feature (Belin et al., 2008; Murray & Arnott, 1993; Paulmann & Kotz, 2008; Pollermann & Archinard, 2002; Yildirim et al., 2004).

2.1.5 Creation of conditions

Next, through the creation of five speech conditions, F0, intensity, and speech-rate cues were systematically rendered uninformative by attenuating variations in the given feature. The five conditions were named according to the attenuated feature(s): a) *natural* (all variations in features intact), b) *intensity+rate*, c) *F0*, d) *intensity+F0*, e) *rate+F0*. In the natural condition, the signal was not altered, thus leaving variations in investigated acoustic features intact. To create the *intensity+rate*, *F0*, *intensity+F0*, and *rate+F0* conditions, variations in the investigated features were attenuated sequentially. Variations in F0 were attenuated first, followed by variations in intensity, then speech rate (e.g., for the *rate+F0* condition, variations in F0 were attenuated before variations in speech rate). Next, the speech stimuli were prepared for each experiment type (i.e., individual sentences for behavioural experiments, 5-sentence blocks of each emotion-condition pair for fNIRS experiments), and uniformly normalised to presentation sound level (Figure 2.3).

The following section will describe the signal processing procedures applied to the speech signal to attenuate variations in each F0, intensity and speech rate, as well as the procedures used to prepare speech for presentation in behavioural and fNIRS experiments (Figure 2.3). After attenuating variations in a cue, the acoustic properties of the stimuli were measured to ensure that the selected signal processing procedure did attenuate variations to the target value (Figure 2.4). The process of attenuating variations in any cue can alter other aspects of the speech signal. This is unavoidable, and where possible, measures were taken to minimise these alterations. The procedures, their success, and any observed influence on other untargeted acoustic features or the signal in general will be described next.

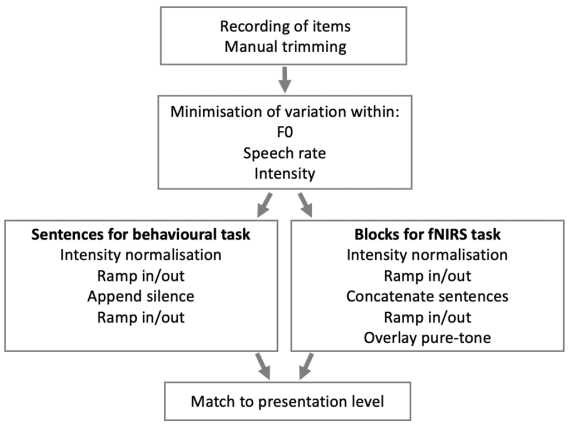


Figure 2.3. Stimulus creation pipeline. Flowchart illustrating the order of steps taken to create single-sentence stimuli for behavioural experiments and 5-sentence blocks fNIRS experiments.

Attenuation of F0 variation

Variations in F0 were attenuated using the PyWORLD vocoder python package (version 0.2.11; Morise et al., 2016). First, the F0 contour is extracted using Distributed Inline-filter Operation algorithm (DIO; Drugman et al., 2019). The F0 floor and ceiling were set to 100 and 800 Hz, respectively. The estimated F0 contour was then refined using StoneMask (Al-Radhi et al., 2019)— an algorithm that reduces the effects of noise on the estimated F0 values. Second, the spectral envelope was extracted using CheapTrick (Morise, 2015) with its default parameters (a spectral recovery parameter of -0.15 and automatically computed FFT size that allows accurate representation of F0 floor). Third, the aperiodicity was estimated using Definitive Decomposition Derived Dirt-Cheap (d4c; Morise, 2016). The threshold for the aperiodicity-based voiced/unvoiced decision was set to the default 0.85. Finally, the original F0 contour was replaced with a constant F0 of 217 Hz—the mean F0 across all stimuli—and the speech signal was reconstructed based on the target F0, spectrogram, and aperiodicity.

As displayed under the experimental condition *F0* in Figure 2.4, manipulating F0 may have a small influence on the RMS of *angry* stimuli; the RMS mean, median and variability are slightly increased for *angry* in the *F0* condition. Attenuating variations in F0 does not appear to have influenced speech rate or introduced any obvious abnormalities to the general signal.

Attenuation of intensity variation

Variations in intensity were attenuated using a Praat 'intensity-neutralizer' script shared by the UCLA (University of California, Los Angeles) Phonetics Lab (Vicenik, n.d.-a). To attenuate variations in intensity to the target value across a WAV file (here: -11.19 dB relative to max RMS—the mean across all natural stimuli), the function measured the intensity of the signal at 10-ms-intervals. At each interval, the difference between the measured and target intensity was obtained. Next, at each interval, the intensity was scaled to the target intensity of -11.19 dB (relative to max RMS).

Attenuating variations in intensity caused a slight reduction of the mean F0 of *angry* stimuli, as illustrated under the intermediate condition *Intensity* in Figure 2.4. It did not influence speech rate. One drawback of the applied procedure is that it flattened the intensity of all speech segments and periods of silence, with the latter resulting in periods of white noise (illustrated in Figures 2.5 and 2.6, *intensity+rate* and *intensity+F0*). In 25% of the *intensity+rate* stimuli and 35% of the *intensity+F0* stimuli, attenuating variations in intensity introduced a 20 to 30-ms clicks. Attempts to resolve this issue failed to systematically reduce the unwanted noise without influencing the overlap between words. Rather than further distorting the speech signal, clicks were

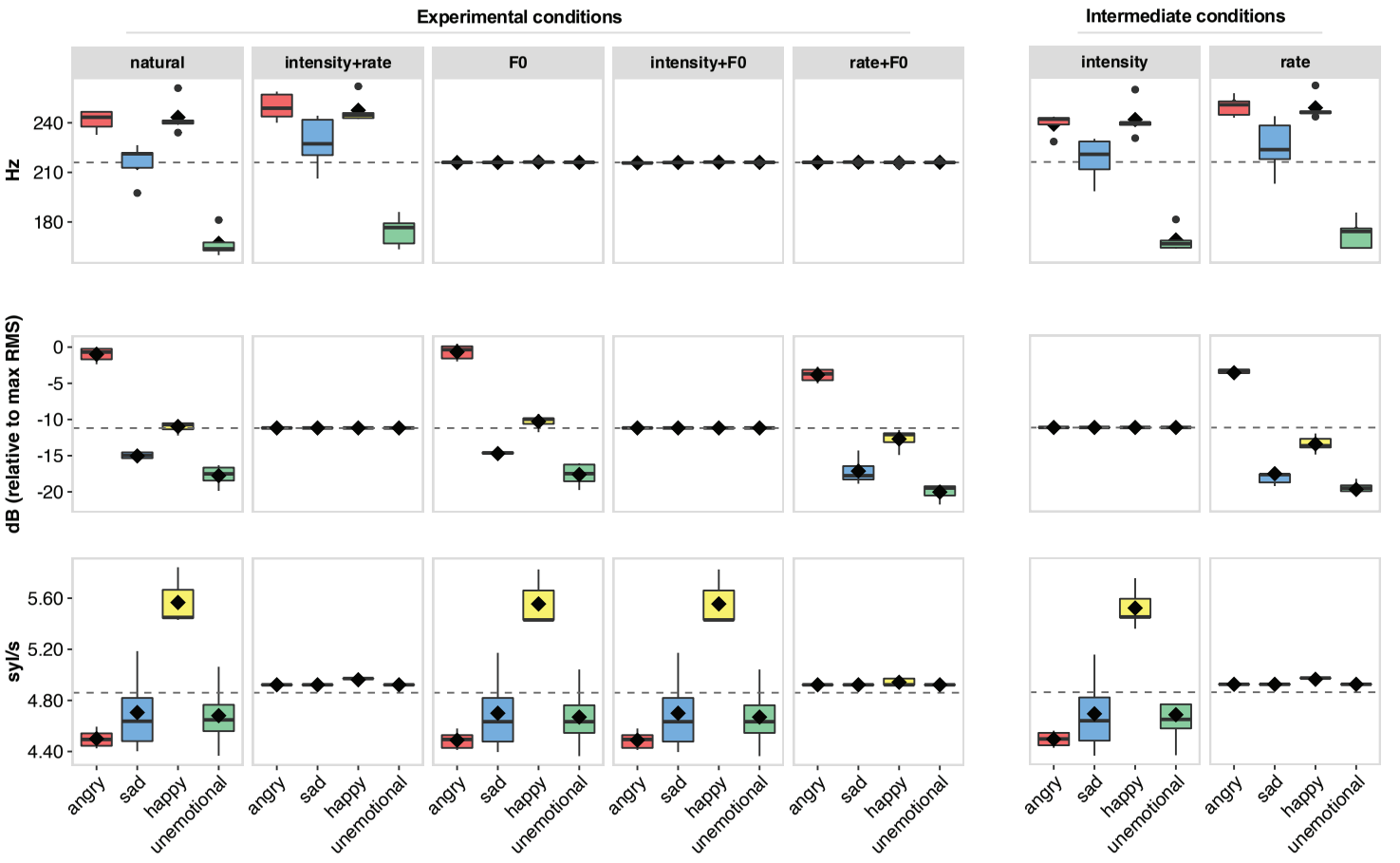
accounted for as a difference between stimulus sentences by including 'sentence' as a random effect when using generalised mixed-effects models (GLMMs) to predict the accuracy of vocal emotion recognition. Note that the conditions containing stimuli with clicks were not included in fNIRS experiments.

Attenuation of speech rate variation

Variations in speech rate were attenuated at the sentence level using the audioTSM python package (Muges, 2017). The phase-vocoder function was used to adjust the speed of the speech by a scaling factor (equal to the original duration divided by the target duration). This function implements the Waveform Similarity-based Overlap-Add procedure (WSOLA; Verhelst & Roelands, 1993). WSOLA uses the overlap-add (OLA) procedure in which the signal is windowed, and the windows are concatenated with a fixed amount of overlap, which lengthens or shortens the signal. The 'waveform similarity' is an additional procedure that allows a variable, instead of fixed, amount of overlap, depending on the similarity of the concatenated windows, to preserve periodic signal components, i.e., reduce aperiodic artefacts. The target duration was set to 1.24 s, which is equal to 4.82 syl/s. The duration of the resulting WAV files was M=1.22 s (SD=0 s), which is equal to 4.92 syl/s. This duration is slightly shorter than the target duration. This deviation from the target, as well as the minimal variability between files (~0.004 s), can be attributed to the precision of the WSOLA procedure.

Attenuating variations in speech rate resulted in a reduction of the RMS for all emotions, as evidenced under the intermediate condition *rate* in Figure 2.4. There was also a small increase in mean F0 for all emotions except *happy*, mirroring the relative change in speech rate (i.e., increased for all emotions except *happy*).

Figure 2.4. *(Next page)* Measured acoustic properties for each emotion within each experimental condition, and two intermediate conditions. The first row shows mean F0 in Hz, the second shows mean intensity in dB relative to the max RMS, the third shows mean speech rate in syl/s. The columns indicate how the signal has been changed. The columns for experimental conditions illustrate the acoustic properties of the experimental stimuli and the supplementary columns for intermediate conditions illustrate the acoustic properties of conditions that were not included in the experiments (i.e., *intensity* and *rate* alone). The intermediate conditions provide insight into the alterations of the speech signal incurred by attenuation of variations within those features. Where variations in a feature were not attenuated (i.e., F0 in the *intensity+rate* condition), the distribution of the emotions resembles the *natural* condition. Where variations in a feature have been attenuated (i.e., F0 in *F0* condition), all emotions have the same mean value as the target value (dashed line; mean across all emotions).


Experimental conditions
Intermediate conditions
natural
intensity+rate
F0
intensity+F0
rate+F0
intensity
rate
Hz
240
210
180
dB (relative to max RMS)
0
-5
-10
-15
-20
syl/s
5.60
5.20
4.80
4.40
angry
sad
happy
unemotional

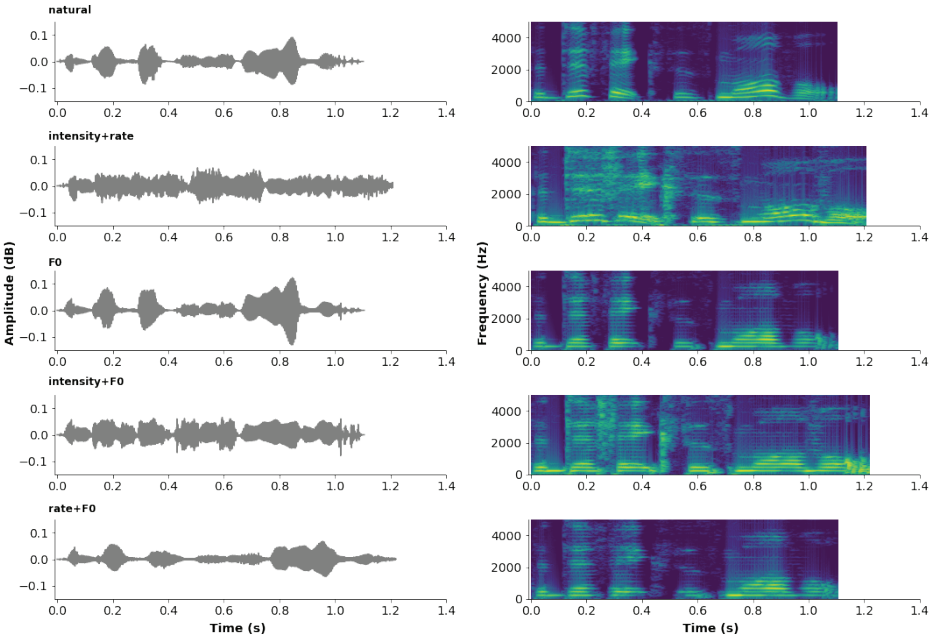


Figure 2.5. *Happy,* waveforms and narrowband spectrograms per condition.

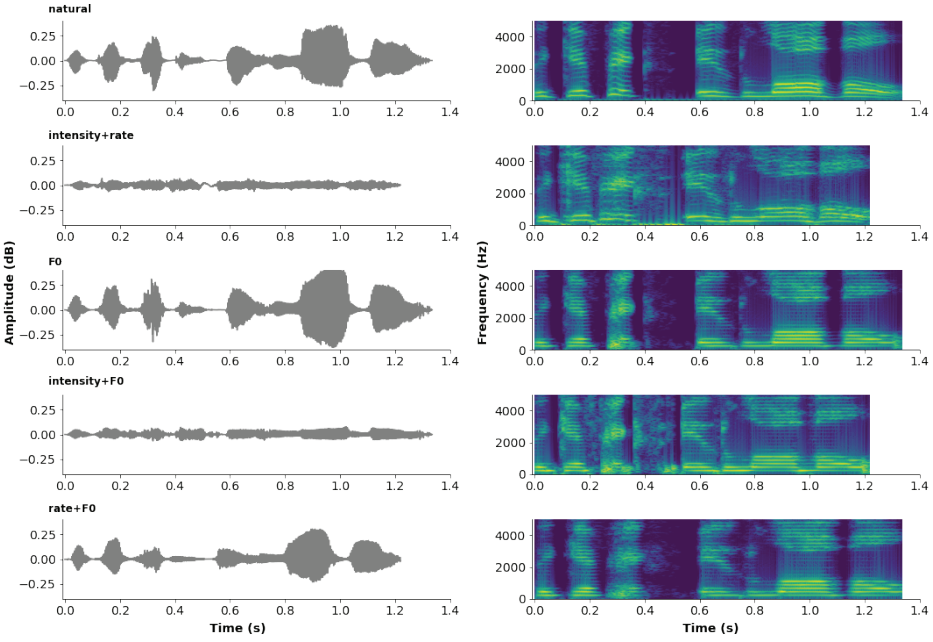


Figure 2.6. *Angry*, waveforms and narrowband spectrograms per condition.

2.1.6 Preparing stimuli for behavioural and fNIRS experiments

Intensity normalisation

With the exception of the sentence-length stimuli presented in the first behavioural experiment on vocal emotion recognition (Chapter 3) where intensity was investigated as a cue for emotion recognition, the overall intensity of all stimuli was normalised (Chapters 4 and 5). To ensure that the overall intensity of all stimuli was equal, a UCLA Phonetics Lab Praat 'intensity-scaler' script (Vicenik, n.d.-b) was adapted for this purpose. Using this script, the signal of each WAV file was multiplied by a scaling factor to obtain an overall intensity of -11.19 dB (relative to max RMS). Subsequently, a 10-ms ramp was applied to beginning and end of each sentence stimuli using the Pysox python package (version 1.4.0; Bittner et al., 2016). This served to smooth onset and offsets of speech, alleviating clicks.

Blocks for fNIRS experiments

Using the intensity-normalised and ramped stimuli, a block was created for each emotion in each of the *natural* and *F0* conditions. Blocks were generated by concatenating the five sentences for each emotion with the Pysox package. The sentences within a block were separated by 200 ms of silence and bookended with 100 ms of silence. Another 10-ms ramp was applied to the beginning and end of the WAV file for each block using the Pysox package.

For each fNIRS experiment, five additional stimuli were created to be used in attention-keeping trials (i.e., trials eliciting a button-press response upon hearing a tone to ensure the participant was attending to the stimuli). These stimuli were identical to the experimental stimuli, with an additional pure tone overlapping with the speech at a random time. To create these, each of the experimental stimuli was copied and a 500-ms-long 400-Hz pure tone (mean intensity of -18.34 dB relative to maximum stimulus RMS) generated in Praat, was overlaid on the speech using Adobe Audition. The tone's intensity was selected to ensure that the attention task was sufficiently challenging to maintain participants' attention but that the tone was not overly salient or shocking.

Following the intensity normalisation and ramping, an additional 200 ms of silence was appended to the beginning and 100 ms to the end of all sentence stimuli using the Pysox package. This was done to ensure that the stimuli would not be interrupted by buffering delays. Next, 10-ms ramps were applied to the beginning and end of the WAV files.

Matching to presentation level

In a final step, all behavioural and fNIRS stimuli were matched to the target presentation

level. For the behavioural experiment on vocal emotion recognition (Chapter 3), the target level was 60 dBA. Participants reported that subjectively, the stimuli were presented at a comfortable level, but could be presented at a higher level and still be comfortable. As stimulus intensity is positively correlated to the amplitude of auditory haemodynamic responses (Weder et al., 2018), which are intrinsically relatively small (e.g., relative to motor responses), the target level was increased to 70 dBA for the fNIRS experiments (Chapters 4 and 5) and the associated behavioural experiment (Chapter 5), to ensure that reliable haemodynamic responses could be obtained. Sound-level matching was achieved using a white noise generated in Praat, which was played in Presentation through an RME Fireface UC attached to Etymotic ER2 (first behavioural experiment) or ER3 (fNIRS and second behavioural experiment) insert phones (Etymotic Research, Inc., Elk Grove Village, USA). The long-term average sound pressure level (in dBA) of the white noise was measured over ~20 s using a 2-cc ear simulator (RA0045, G.R.A.S., Twinsburg, USA) and a B&K Type 2250 sound level meter (Brüel & Kjær, Nærum, Denmark). The measured level (96 dBA) was used to calculate the RMS of the white noise, which served as the reference RMS. Using this reference, the RMS was calculated for a calibration track consisting of concatenated *intensity+rate* sentences with no silences (the silence threshold was set to -40 dB). Using these RMS values and the reference RMS, the calibration track and the stimuli were then scaled to the target presentation level (i.e., 60 dBA for the behavioural experiment in Chapter 3, and 70 dBA for the fNIRS experiments and associated behavioural experiment in Chapters 4 and 5).

2.1.7 Final stimuli for behavioural and fNIRS experiments

In the behavioural experiment (Chapter 3), the experimental paradigm comprised *angry*, *happy*, *sad*, and *unemotional* sentences, presented in five conditions (*natural, intensity+rate*, *F0*, *intensity+F0*, *rate+F0*). In both fNIRS experiments in this thesis, a reduced number of conditions and/or emotions was presented, with a silent control condition and 10 attention trials, to respect the time constraints imposed by fNIRS recordings. For the first fNIRS experiment (Chapter 4), the block-design paradigm included each *angry*, *happy*, *sad*, and *unemotional* trials (*natural* condition only), the silent control trials, as well as attention and practice trials (total trials=114, duration=50 minutes). In the second fNIRS experiment (Chapter 5), the block-design paradigm included two emotions (*happy* and *sad*), in two speech conditions (*natural* and *F0*), as well as the silent control, attention, and practice trials (total trials=144, duration=50 minutes).

2.1.8 Percepts vs. acoustic measures

As discussed above and illustrated in Figure 2.4, the success of attenuating variations in each acoustic feature was verified and secondary effects of the implemented signal-processing procedures were considered. Still, it is difficult to determine whether the cues associated with acoustic features have been rendered completely perceptually constant, and thus unhelpful, to listeners. Auditory speech perception is flexible, meaning that one acoustic characterisation of a feature can be perceived differently depending on the accompanying acoustic landscape (Liberman 1967). Concretely, this means that despite efforts to limit the richness of the cues in each F0, intensity and speech rate, the associated percepts may be enriched by acoustic information from the unattenuated acoustic features and acoustic features beyond the scope of this thesis, including spectral information (Plack, 2018), and aspects of voice quality, such as vocal effort (Gobl & Ní Chasaide, 2003; Waaramaa et al., 2010) and vocal fry (Kuang & Liberman, 2016). For example, when variations in F0 are attenuated, the harmonic information may be sufficient to create a replacement F0 percept (Terhardt, 1979). Intensity is the acoustic correlate of perceived loudness, which is influenced by F0 and spectral information (Plack, 2018; Yanushevskaya et al., 2013), and can even covary with expectation and meaningfulness (Mershon et al., 1981; Tian et al., 2018). Finally, speech rate impacts the perception of global F0 peaks (Niebuhr & Pfitzinger, 2010), whereas vocal fry induces a lower F0 percept (Kuang & Liberman, 2016). In sum, the relationship between the richness of variations in F0, intensity, and speech rate and the associated percepts is complex.

2.2 Functional near-infrared spectroscopy (fNIRS)

2.2.1 Equipment and laboratory

The fNIRS data presented in this thesis was collected using a NIRScoutX (NIRx Medical Technologies LLC, Berlin, Germany), with 24 sources, 32 detectors (with avalanche photodiode sensors) and one bundle of 8 short detectors. The spectrometer measures oxygenated (HbO) and deoxygenated (HbR) haemoglobin using wavelengths of 760 and 850 nm, and the sampling rate is 62.5 Hz divided by the number of sources used. For the experiments described in this thesis, the sampling rate was 2.6 Hz (62.5 Hz/24 sources). This sampling rate is higher than the rate of cardiac, respiratory and blood pressure (physiological noise) signal components, facilitating the identification of these components (Huppert, 2016). The optodes were secured on the head using flexible mesh caps (Easycap, Herrsching, Germany) pre-marked with International 10/10 system positions (Chatrian et al., 1985; Figure 2.9C). Optode cables were supported by a strain-relief arm for stabilisation and to alleviate any pull on the optodes or the partici-

pant's head (Figure 2.7).

The fNIRS measurements were conducted in a sound-proof booth, with the computers running the recording and experiment presentation software in the adjacent room to minimise ambient sound levels. An RME Fireface UC sound card (Audio AG, Haimhausen, Germany), Etymotic ER3 insert phones (Etymotic Research, Inc., Elk Grove Village, USA) used to present auditory stimuli, and an RB-840 button box (Cedrus Corporation, San Pedro, USA) were inside the booth. Participants were supervised via a video camera (Figure 2.7). The mean ambient sound level, i.e., the long-term average sound pressure level in dBA, measured over ~20 s using a B&K Type 4189 free-field microphone and B&K Type 2250 sound level meter (Brüel & Kjær, Nærum, Denmark) was ~26 dBA. During experiments, ambient light was minimised by turning off ceiling lights. As no visual stimulation was required in the experiments, all computer screens in the booth were also turned off.

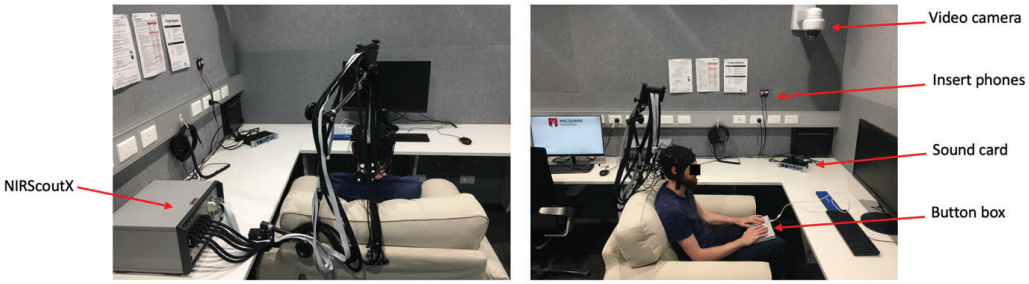


Figure 2.7. fNIRS lab arrangement. Lab arrangement with a comfortable chair, participant wearing an example cap, and ceiling lights turned on. Ceiling lights and computer screens were turned off during fNIRS recordings.

Before recording the data presented in Chapters 4 and 5, the lab setup and functionality of the fNIRS equipment were tested using a finger-tapping experiment, known to evoke reliable haemodynamic responses in individual participants (e.g., Huppert et al., 2006). Motor-evoked haemodynamic responses were recorded from six participants (averaged waveforms in Figure 2.8) with optodes placed over the motor brain regions. Participants tapped the fingers of the cued hand (left and right in separate trials) for 5s with 30 repetitions per hand. A control rest condition was also included (also 30 repetitions). As seen in Figure 2.8, unilateral finger-tapping evoked haemodynamic responses in both HbO and HbR in the bilateral motor regions, with larger responses in the contralateral motor region.

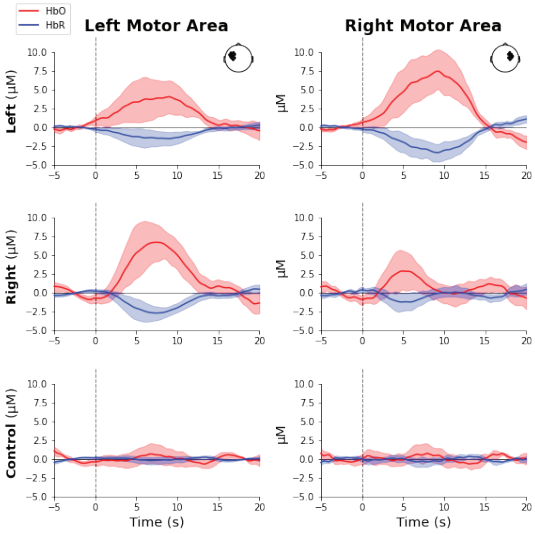


Figure 2.8. Haemodynamic responses measured from motor brain regions evoked by finger-tapping. Grand average waveforms for left-hand tapping, right-hand tapping, and control conditions (rows) and left and right motor brain regions (columns). The position of the optodes in the left and right motor regions is represented in the inset head shape (viewed from above). Solid lines indicate the mean and the shaded area indicates the 95% confidence interval.

2.2.2 Designing of montage

A montage of 24 sources, 17 detectors, and 8 short detectors was used to record from bilateral STG, IFG and MFG. To cover these brain areas, the montage comprised 60 long channels (source-detector pairs ~30 mm apart), and 8 short channels (source-detector pairs 8 mm apart). Bilateral STG and IFG ROIs were covered by 10 long channels each, and bilateral MFG ROIs were covered by 6 long channels each. The short channels were distributed among the ROIs, with one in each IFG and MFG, and two in each STG to account for location-dependent heterogeneity in the extracerebral signals (Brigadoi & Cooper, 2015; Gagnon et al., 2012; Y. Zhang et al., 2015).

The positions of optodes on the scalp were determined using the AAL2 atlas in the fOLD toolbox (Rolls et al., 2015; Tzourio-Mazoyer et al., 2002; Zimeo Morais et al., 2018). The toolbox suggests channels within a selected anatomical region, which are associated with a specificity parameter. 'Specificity' is an estimate of photon sensitivity to the brain regions probed by each channel as a percentage, and the specificity parameter is used to define the minimum specificity of the channels suggested for a given brain area. To define ROIs covering bilateral STG, IFG and MFG, the relevant brain areas were selected using AAL2 nomenclature, with symmetry enforced to have an equal

number of channels over both hemispheres. This included left and right Temporal_Sup for the STG, left and right Frontal_Inf_Oper, Frontal_Inf_Tri, and Frontal_Inf_Orb for the IFG, and left and right Frontal_Mid_2 for the MFG. Subsequently, to maximise the number of channels covering each region, thereby accounting for individual differences in cortical anatomy and functionality (Cooper et al., 2012; D. Wang et al., 2015), the specificity parameter was reduced incrementally. Once the suggested channels included one point of overlap with a neighbouring brain region of interest, the value of the specificity parameter was noted as the specificity cut-off for the selected brain region (Table 2.4). All suggested channels from the International 10/10 system (Chatrian et al., 1985), above the specificity cut-off, were included in the ROI. Additional channels suggested for the IFG and STG according to the International 10/5 system (Oostenveld & Praamstra, 2001), were also included to maximise the use of available optodes. The optodes were mounted onto mesh caps marked with International 10/10 positions (Easycap, Herrsching, Germany) using grommets and spacers to maintain 30-mm separation (NIRx Medical Technologies LLC, Berlin, Germany).

The montage provides dense coverage of the ROIs with variable specificity. Variable specificity is to be expected based on the variable depth of the cortical tissue below the scalp. In the montage, bilateral MFG has the highest specificity, followed by bilateral IFG, and finally STG (Table 2.4). Accordingly, the scalp-brain distance increases from ~13 mm for the MFG and IFG to ~15 mm for the STG (Cui et al., 2011). These scalp-brain distances support the suitability of these ROIs for fNIRS measurements.

Table 2.4. Average specificity of long channels per ROI

ROI	**Hemisphere**	**N channels**		**Long channel specificity %**		
		Long	Short	Cut-off	*M*	*SD*
STG	L	10	2	12	22.55	8.27
	R	10	2	15	39.01	21.78
	Bilateral	-	-	14	-	-
IFG	L	10	1	25	56.86	18.04
	R	10	1	26	50.62	20.87
	Bilateral	-	-	30	-	-
MFG	L	6	1	48	62.78	17.8
	R	6	1	62	69.97	6.69
	Bilateral	-	-	60	-	-

Notes. Average channel specificity, i.e., an estimate of photon sensitivity to the probed brain area as a percentage per ROI, extracted from fOLD toolbox using the AAL2 atlas (Rolls et al., 2015; Zimeo Morais et al., 2018). Cut-off refers to minimum specificity of suggested channels within ROI, where 'bilateral' indicates cut-off values for channels suggested with symmetry between hemispheres enforced.

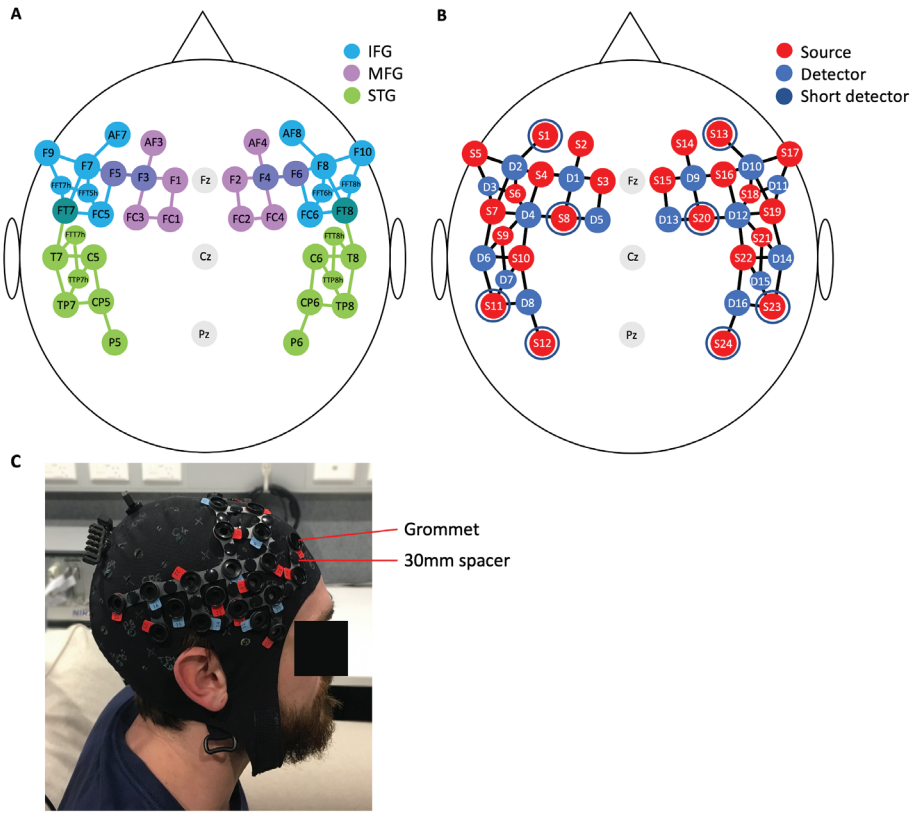


Figure 2.9. Positioning of optodes, channels and the ROIs on the head. A) ROIs coded by colour. STG=superior temporal gyrus, IFG=inferior frontal gyrus, MFG=middle frontal gyrus. IFG and MFG overlap at F3, F5, F4 and F6. IFG and STG overlap at FT7 and FT8. B) Array of sources, detectors and short detectors in the montage. C) Cap positioned on the head; red tags indicate source positions, blue tags indicate detector positions. Tags are attached to grommets that hold optodes and are separated by 30-mm spacers.

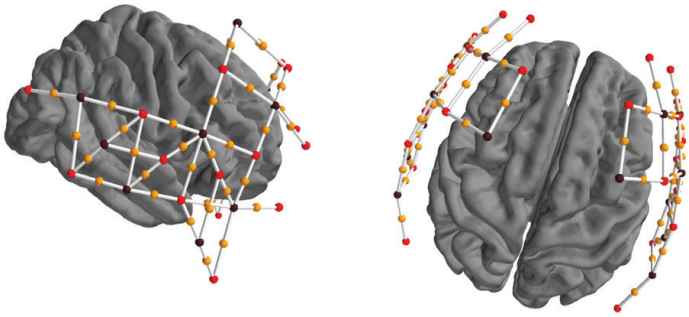


Figure 2.10. Visualisation of montage over brain. Black spheres are detector optodes, red spheres are source optodes, and yellow spheres represent the estimated point of measurement between a source-detector pair. Short-channel detectors are not represented.

2.2.3 Experiment protocol

Before the experiment, the researcher turned on the NIRScoutX (allowing at least 30 minutes to reach stable temperature before recording) and turned on all lab computers. The researcher opened the fNIRS recording software (NIRStar, version 15.3, NIRx Medical Technologies LLC, Glen Head, USA), selected the montage described above and verified that the hardware settings were correct (i.e., corresponded to the correct number of available sources, detectors and short-detectors). During a preview recording, the researcher checked that the experiment was correctly presented (i.e., correct triggers received in NIRStar, correct inter-stimulus intervals, and stimuli were heard through insert phones). Furthermore, the presentation level of the auditory stimuli was verified as described in the stimulus creation section (see Chapter 2.1.6).

After a participant was welcomed to the lab, the researcher explained the experiment and requested informed consent from the participant. Upon receiving informed consent, the participant was seated in the comfortable armchair and instructed to let down hair, if needed. To ensure correct positioning of the cap, a measuring tape was used to position the CZ-marker on the cap at the midpoint between the nasion and inion. Because optodes interfere with measurements between preauricular points, the researcher checked that cap looked centred on the forehead and the participant's ears fit equally into the cap's ear-slits. Optodes were mounted onto the cap, ensuring that hair was moved to the side, allowing each optode to make contact with the scalp. Glasses were positioned on the outside of the cap if they were needed.

Once all optodes were mounted, the signal quality was checked in NIRStar, using a computer inside the booth with the ceiling lights dimmed. NIRStar's signal quality metric combines the gain stage of each source LED, with the measured voltage and a measure of noise for each channel, describing a channel's signal quality as 'excellent', 'acceptable', 'critical' or 'lost' (Figure 2.11). For channels that were not 'excellent', hair was cleared again from under the corresponding optode pair and the signal quality was re-assessed. While the aim was to reach 'excellent' for all channels, this was not always possible (i.e., in cases of very thick, coarse, dark hair). Once satisfied with the signal quality, the researcher turned off the computer screens in the booth, instructed the participant on the task, and invited the participant to ask questions. Next, the researcher placed the button-box on the participant's lap and inserted the insert phones into the participant's ears, with the participant's permission. The researcher left the booth and turned off the ceiling lights. The calibration was repeated from the adjacent room, with the ceiling lights off in the lab, before beginning the fNIRS recording and the experiment. If necessary, the researcher re-entered the booth to remove hair from under 'critical' channels. The calibration values for each channel were noted for each participant, participants with more than two 'critical' and/or 'lost' channels, were excluded from the analysis.

During the experiment, the researcher monitored the participant from the adjacent room via the video camera, as well as the fNIRS signal in NIRStar. After the experiment, the researcher stopped the recording, turned the ceiling lights on to a dim setting, entered the booth and, with the participant's permission, removed insert phones and the fNIRS cap from the participant's head. If a behavioural task was included, participants were offered a break before continuing. When all tasks were completed, the participant was debriefed, thanked, and paid an honorarium or given course credit for their participation. The recorded data was saved in BIDS data format (Gorgolewski et al., 2016). The optodes were then detached from the cap, and subsequently, the optodes, cap and lab were sanitised.

Signal Quality	NScout Gain [10^x]	Level [V]	Noise [%]
Excellent	1 - 6	0.09 - 1.40	< 2.5
Acceptable	7	0.03 - 0.09 1.40 - 2.50	2.5 - 7.5
Critical	0 8	0.01 - 0.03 > 2.50	> 7.5
Lost	-	< 0.01	-

Figure 2.11. Lookup table for NIRStar signal quality metric, all units used were taken from the manual. Gain=gain stage between 0–8, Level=voltage on a logarithmic scale, Noise=coefficient of variation=(standard deviation/mean*100) using voltage data. Table adapted from NIRScout 15.3 User Manual (NIRx Medical Technologies LLC, Glen Head, USA).

2.2.4 Data analysis pipeline

Grand average waveforms for visual inspection

To facilitate visual inspection of the data, grand average waveforms of the HbO and HbR time courses for each combination of condition and ROI, e.g. Figure 2.8, were generated using MNE (version 0.21.2; Gramfort et al., 2013, 2014) and MNE-NIRS (version 0.0.1; Luke et al., 2020). The data were resampled to 3 Hz, to exceed the Nyquist frequency required for later processing steps (e.g., 2 x 1.35 Hz for the scalp coupling index). Then raw intensity is converted to optical density. Absolute raw intensity values were used to avoid miscalculations stemming from negative intensity values, which can be introduced by movement or hardware glitches (Luke et al., 2021). A scalp-coupling index (SCI) was employed as a measure of signal quality. The SCI calculated the correlation of the heart rate signal (~1 Hz) in the HbO and HbR signals for each channel (Pollonini et al., 2014) for frequencies between 0.7–1.35 Hz. Channels with SCI values <0.8 were rejected. To correct motion artefacts in the signal, the tempo-

ral derivative distribution repair (TDDR) algorithm was applied. TDDR identifies and corrects baseline shifts and spikes in the raw signal (Fishburn et al., 2019). Next, the nearest short channel was identified for each long channel, and short-channel regression was applied, effectively isolating the cerebral signal component by regressing out extracerebral and systemic components (R. B. Saager & Berger, 2005; Scholkmann, Klein, et al., 2014). The signal was then converted from optical density to concentrations of HbO and HbR using the Modified Beer-Lambert Law (Delpy et al., 1988; Kocsis et al., 2006) with a partial pathlength factor of 0.1, which accounts for the effective distance that photons travel between the source and detector (Scholkmann, Kleiser, et al., 2014), and the actual proportion of photons that pass through cerebral tissue (Strangman et al., 2003). Cui et al.'s (2010) algorithm that improves signal-to-noise ratio using the negatively correlated dynamics of HbO and HbR was applied. The signal was then bandpass-filtered between 0.01–0.7 Hz to exclude physiological signal components such as heart rate (~1 Hz). Response epochs were trimmed from 5 s before stimulus onset to 20 s post-onset (~12.75 s post-stimulus offset) and linearly detrended, accounting for slow drifts. Epochs with peak-to-peak differences >200 μM, i.e., indicative of motion artefacts, were not included in the grand average time course.

Inferential analysis of haemodynamic response amplitude

First-level analysis. The generalised linear model (GLM) approach was taken to quantify the amplitude of evoked haemodynamic responses. In the first-level (individual-participant-level) analysis estimates of haemodynamic response amplitude for each participant per channel and condition were obtained using MNE, MNE-NIRS, and NiLearn (version 0.70; Abraham et al., 2014). The sampling rate of the recorded signal was reduced from 2.6 to 0.6 Hz (Luke et al., 2021) as the SCI was not calculated, alleviating the need for higher frequencies. The signal was converted from raw intensity to optical density, using absolute raw intensity values. Next, the signal was converted to concentrations of HbO and HbR using the Modified Beer-Lambert Law (Delpy et al., 1988; Kocsis et al., 2006) with a partial pathlength factor of 0.1. As described above, this factor accounts for the effective distance travelled by photons between optodes (Scholkmann, Kleiser, et al., 2014) and the proportion of photons that pass through cerebral tissue as opposed to extracerebral tissue (Strangman et al., 2003). The GLM was only fit to the long-channel data—isolated by rejecting channels <20 mm or >40 mm. The design matrix for the GLM was generated by convolving a 3-s boxcar function at each event-onset time with the canonical haemodynamic response function (Glover, 1999). The boxcar function is shorter than the stimulus duration, informed by Luke et al.'s (2021) demonstration that a 3-s boxcar function maximised true positive rate for haemodynamic responses evoked by 5-s auditory stimuli. The GLM also included all principal components of short-detector channels to account for extracerebral and physi-

ological signal components. Further, to account for slow drifts in the signal, attributable to the warming of the spectrometer or the participant's head, drift orders accounting for signal components up to 0.01 Hz were included as regression factors (Huppert, 2016). The GLM was performed with a lag-1 autoregressive noise model, as recommended by Huppert (2016), to account for the correlated nature of the fNIRS signal components (described in Chapter 1.2.3). The GLM computes a response estimate for each channel per condition per participant. To facilitate the ROI analysis, averages of the estimates for the channels within each ROI were calculated, weighted by the standard error. The ROI estimates were then exported as a CSV data file to be used in the second-level (group) analysis.

Second-level analysis. The second-level (group-level) analysis utilised linear mixed-effects (LME) models performed using the lme4 package (version 1.1.2; Bates et al., 2014) in the R language (version 3.6.3; R Core Team, 2020) within the RStudio IDE (version 1.3.959; RStudio Team, 2019). Models were constructed using a forward stepwise approach (Bates et al., 2015), i.e., sequentially adding predictor terms and testing whether each added term accounted for a significantly greater proportion of the variance using a likelihood ratio test. Separate models were built for each chromophore to address multicollinearity (HbO and HbR are known to be negatively correlated; e.g., Wolf et al., 2002) and to gain insight into the subtleties of the variance structure for each chromophore. The intercept was suppressed to compare each predictor term against zero, i.e., to assess whether the predicted amplitude of a haemodynamic response is significantly different from zero for a given ROI and condition (Santosa et al., 2018). Influential data points were identified using Cook's distance (Cook, 2011) and subsequently excluded. Unstandardised beta coefficients are reported, as the outcome variables are in the same units and thus, no standardisation is needed for comparison of predictors across models. To describe the goodness of fit of each model, marginal and conditional R^2 values are reported together as $R^2_{m/c}$ for each model (Harrison et al., 2018); the marginal R^2 value represents the variance accounted for by the fixed effects and the conditional R^2 value represents the variance accounted for by the random effects. Finally, hypothesis testing was conducted using the emmeans R package (version 1.4.2; Lenth, 2021), with which relevant group-level estimates of haemodynamic response amplitude were contrasted. The contrasts were corrected for multiple comparisons using the false discovery rate procedure (FDR; Benjamini & Hochberg, 1995).

Chapter 3| Withholding emotion: attenuating variations in voice pitch, intensity, and speech rate reduces the accuracy with which listeners recognise vocal emotions

3.1 Abstract

Background. Successful social interactions depend on accurately recognising emotions conveyed in speech, which in turn, relies on the richness of acoustic features such as fundamental frequency (F0), intensity, and speech rate. Listeners may weight these acoustic features based on the usefulness of the cues they provide, allowing listeners to compensate for features with less informative cues (e.g., the altered F0 cues in speech transmitted through a cochlear implant) by making use of acoustic features conveying more informative cues.

Aims. To investigate how normal-hearing (NH) listeners rely on F0, intensity, and speech rate when identifying vocal emotions, behavioural accuracy for recognition of vocal emotions with systematically attenuated variations in F0, intensity and/or speech rate (i.e., one or two acoustic features at a time) was assessed.

Method. 40 NH listeners heard pseudo-English sentences conveying anger, happiness, sadness, or neutrality and were asked to categorise these as *angry*, *happy*, *sad*, or *unemotional* in a 4-alternative forced-choice task. The sentences were presented as a) natural speech or with attenuated variations in the acoustic features that convey emotion in speech, b) intensity and speech rate, c) F0, d) intensity and F0, e) speech rate and F0. Sliding difference contrasts between consecutive conditions (a vs. b, b vs. c, etc.), based partly on a priori knowledge of the relative importance of the acoustic features, were used to examine whether the cue(s) that differed between pairs contributed to successful recognition of vocal emotion, overall and as well as individually for each emotion.

Results. For overall recognition of vocal emotions, the largest significant reduction in accuracy occurred when variations in F0 were attenuated. A smaller, but still significant, reduction was observed when variations in intensity and rate were simultaneously attenuated. For vocal emotions *angry*, *happy* and *sad*, attenuation of variations in F0 significantly reduced the accuracy with which they were identified. This was not the case for *unemotional*.

Conclusions. These findings confirm that to recognise vocal emotions, NH listeners rely on F0 cues most heavily and that, at the group level, they cannot make use of intensity or speech-rate cues, separately or combined, to compensate for uninformative F0 cues.

3.2 Introduction

The ability to recognise emotions in speech is important to successful communication, social interaction, and quality of life (Lindner & Rosén, 2006; Luo et al., 2018; Zinchenko et al., 2018). In face-to-face communication, visual and auditory cues convey information about an interlocuter's emotional state. However, in many common listening situations (e.g., during a telephone conversation or in a poorly lit room) visual information may be limited, and listeners must rely partly or entirely on vocal information to decode their interlocuter's emotional state. Any degradation of this vocal information—or reduced capacity to process it—can impair communication and strain social interaction.

The main acoustic features conveying vocal emotions are the fundamental frequency (F0; the rate of vibration of the vocal folds), intensity (the sound pressure level at which speech is uttered), and speech rate (the number of syllables spoken per second; Frick, 1985; Scherer et al., 2003). Respectively, F0, intensity and speech rate correspond to the percepts of voice pitch, loudness and tempo (Juslin & Laukka, 2001; Scherer et al., 2003). Variations in these acoustic features over time provide cues, which listeners use to extract emotional meaning from vocalisations. In other words, reducing variations in an acoustic feature attenuates the cues with which listeners recognise vocal emotions, rendering the cues, and thereby the acoustic feature, less informative.

As per the weighting-by-reliability hypothesis (Toscano & McMurray, 2010), listeners weight acoustic features according to how useful the conveyed cues are, meaning that when attenuated variations in an acoustic feature render cues uninformative, listeners may rely on other acoustic features with more informative cues. In recognising vocal emotions, listeners rely most heavily upon F0 cues (Banse & Scherer, 1996; Metcalfe, 2017; Patel et al., 2011; Scherer et al., 2003), and the degree to which listeners can make use of intensity and speech-rate cues remains unclear. Metcalfe (2017) demonstrated the primacy of F0 as a stand-alone acoustic feature conveying cues to vocal emotions: NH listeners asked to recognise vocal emotions with variations preserved only in F0, intensity, or speech rate were able to recognise emotions above chance level when only F0 cues were preserved but not when only intensity or only speech-rate cues were preserved.

An initial motivation for this study was that individuals with hearing loss, who use cochlear implants (CIs) or hearing aids (HAs) have difficulty identifying emotions in speech (for CI: Chatterjee et al., 2015; Everhardt et al., 2020; Gilbers et al., 2015; Jiam et al., 2017; Luo et al., 2007; Most & Aviner, 2009; Nakata et al., 2012; Pak & Katz, 2019; Panzeri et al., 2021; Pereira, 2000; Peters, 2006; Ren et al., 2021; Waaramaa et al., 2018; and for HA: Goy et al., 2018; Most & Aviner, 2009; Waaramaa et al., 2018).

CIs and HAs help mitigate sensorineural loss, yet both types of devices can alter spectral, temporal, and intensity information of speech, adding to the challenge of perceiving vocal emotions. The challenge likely arises from a combination of the hearing-loss-induced changes to the auditory perceptual system and the alterations to the speech signal introduced by the hearing device, fit to accommodate the impaired auditory system. In the process of amplifying sounds sufficiently for individuals with severe hearing loss, HAs often compress the intensity range to suit the listeners' reduced dynamic acoustic range. This can alter the shape of the speech spectrum, change the intensity contours, but can also alter temporal envelopes due to time constants of compression. Collectively, these can potentially alter the perceived pitch, as well as intensity contours, of the speech (Goy et al., 2018; Lesica, 2018). CIs divide the acoustic signal into several frequency bands before converting it to an electrical signal. In this electrical signal, the intensity range is compressed and the spectral resolution is reduced, and in each channel the information is carried in temporal envelopes with no temporal fine structure. These collectively weaken the F0 information in the transmitted speech signal (Başkent et al., 2016; Plack, 2018; Wilson & Dorman, 2008). The acoustic signal transmitted by a CI can be approximated using noise-band vocoding (Shannon et al., 1995), and can be presented to NH listeners to investigate the impact of CIs on various aspects of speech perception, including recognition of vocal emotions (e.g., Everhardt et al., 2020). Neither HAs nor CIs alter speech rate per se, meaning that speech-rate cues transmitted by these devices have the potential to be informative cues to vocal emotions.

The weighting-by-reliability hypothesis (Toscano & McMurray, 2010) would suggest that hearing-impaired listeners, faced with reduced F0 cues, would rely more heavily on intensity and speech rate cues to support vocal emotion recognition. Vocal emotion recognition is reduced in HA, CI, and CI-simulation studies, and importantly, substantial variability is observed in individual listeners' accuracy scores (e.g., Chatterjee et al., 2015; Goy et al., 2018; for CI meta-analysis see Everhardt et al., 2020). Just as accuracy scores vary between listeners, the ability to make use of cues conveyed by intact acoustic features to compensate for uninformative F0 cues may differ between listeners (e.g., Luo et al., 2007; Winn et al., 2012). Luo et al. (2007) report that at the group level, NH listeners presented full-spectrum speech or CI-simulations and CI listeners presented full-spectrum speech recognise vocal emotions more accurately when overall intensity cues are intact, i.e., not normalised. Gilbers et al. (2015) suggest that informative overall intensity cues do not improve recognition of vocal emotions perceived through a CI or CI-simulation significantly. While Hegarty & Faulkner (2013) propose that CI listeners make use of speech rate cues to recognise vocal emotions, the majority of studies suggest listeners cannot make use of speech-rate cues to identify emotions conveyed in speech with uninformative F0 cues (Gilbers et al., 2015; Luo, 2016; Van de Velde, 2017). Combined, these studies highlight the lack of consensus regarding listeners'

abilities to make use of intensity or speech-rate cues in the presence of uninformative F0 cues.

Linguistic prosody, i.e., recognition of question/statement intonation and sentence stress, is also conveyed primarily using F0, with intensity and speech rate as secondary acoustic features (P. Warren, 1999). CI and CI-simulation studies of linguistic prosody provide further evidence that listeners can make use of intensity cues when F0 cues are reduced or less informative (Marx et al., 2015; Meister et al., 2011; Peng et al., 2012, 2009). Considering the evidence from vocal emotion recognition and linguistic prosody together, there is stronger evidence for listeners making use of intensity than speech-rate cues to compensate for uninformative F0 cues.

Previous investigations of cue-weighting in vocal emotions have either attenuated variations in individual acoustic features to observe how reducing the information provided by cues impairs accuracy (e.g., Gilbers et al., 2015; Luo et al., 2007) or attenuated variations in pairs of acoustic features, to quantify listeners' abilities to make use of individual acoustic features providing potentially informative cues (e.g., Metcalfe, 2017). Here, I systematically investigate how NH listeners weight the acoustic features that provide cues to vocal emotions by attenuating the variations in F0, intensity and speech rate, one or two features at a time. Based on the evidence of listeners' use of acoustic features with potentially informative cues (i.e., intact variations) to recognise vocal emotions, I hypothesise that the accuracy with which vocal emotions are identified is greatest for vocal emotions in a) natural speech, and decreases as variations are attenuated in the acoustic features that convey emotional prosody. Performance is predicted to fall as potential cues, ordered from least to most impactful, are rendered uninformative in a series of speech conditions: b) intensity and speech-rate cues combined, c) F0 cues alone, d) F0 and intensity cues combined, and e) F0 and speech-rate cues combined. Sliding difference contrasts are employed to compare accuracy in consecutive speech conditions (a vs. b, b vs. c, etc.). The absence of a significant difference between consecutive conditions will indicate that listeners do not make use of the cue(s) that differ(s) between conditions. A reduction in accuracy between consecutive speech conditions will indicate that listeners do indeed make use of the attenuated cue(s) that differ(s) between conditions. This study serves to validate novel vocal emotion stimuli for future neuroimaging studies and to inform the experimental design of these studies.

3.3 Methods

3.3.1 Participants

From fifty-one native speakers of English initially recruited from Macquarie University, 40 participants (50% female; age range: 18–36 years, *M*=24 years; *SD*=6 years) had no known neurological or psychological disorders and pure-tone audiometric thresholds indicating no more than slight hearing losses, as defined by the American Speech Language Hearing Association (≤25 dB HL for all octave frequencies between 250–8000 Hz; Clark, 1981).

Ethical approval was obtained from Macquarie University's ethics committee (Reference number: 5201952978351). Written informed consent was collected from all participants, who received an honorarium for their involvement.

3.3.2 Stimuli

Stimuli comprised six-syllable sentences with rhyming pseudo-words and real function-words, such as "*the ziffox is dorval*" and "*the miffox is borval*" conveying *angry*, *happy*, *sad*, and *unemotional* prosody. These recordings were then manipulated to create five speech conditions, named after the acoustic features within which variations (i.e., cues) were attenuated: *natural*, *intensity+rate*, *F0*, *intensity+F0*, and *rate+F0*. Variations in the relevant acoustic features along the sentence-length trajectory were attenuated as described in Chapter 2.1, rendering the associated acoustic cues to emotional prosody uninformative. In speech conditions where variations in a given acoustic feature were not attenuated, overall differences in that acoustic feature are intact (i.e., F0 in *natural* and *intensity+rate* conditions; intensity in *natural*, *F0*, *rate+F0* conditions; speech rate in *natural*, *F0*, *intensity+F0* conditions). For a more detailed description of the stimuli and the stimulus-creation procedure, refer to Chapter 2.1. The total number of stimuli was 100 (i.e., 5 sentences x 4 emotions x 5 speech conditions).

3.3.3 Procedure

Participants completed a four-alternative forced-choice (4-AFC) emotion recognition task. The 4-AFC task followed a functional near-infrared spectroscopy (fNIRS) pilot recording during which each stimulus sentence was presented 10 times, rendering the participants familiar with the stimuli.

The task consisted of 4 practise trials (*angry*, *happy*, *sad*, *unemotional* in the *natural* condition, each presented once) and 100 test trials (each of 100 stimuli presented once). All auditory stimuli presented diotically (i.e., identical at the two ears) in a randomised order using Presentation® software (version 20.2; NeuroBehavioral Systems Inc., 2020)

via an RME Fireface UC (Audio AG, Haimhausen, Germany) and Etymotic Research ER-2 insert earphones (Etymotic Research, Inc., Elk Grove Village, USA). Stimuli were presented at a mean level of 60 dBA. Participants completed the task in 15 minutes without breaks.

Participants were seated in a sound-treated booth in front of a desktop computer and were instructed to indicate which emotion was conveyed in the speech using the top four buttons on an RB-840 button box (Cedrus Corporation, San Pedro, USA). They were informed that the first four trials were practice trials—which were not scored—and then given a chance to ask questions before beginning the experiment.

While each trial was presented, participants viewed a screen with the question "Which emotion did you hear?" and the alternatives *unemotional*, *happy*, *sad*, and *angry* (Figure 3.1) were provided. The position of the response alternatives on the screen corresponded to the position of the response buttons on the button box. Once the participant pressed a button, there was a 1-s delay before the next trial.

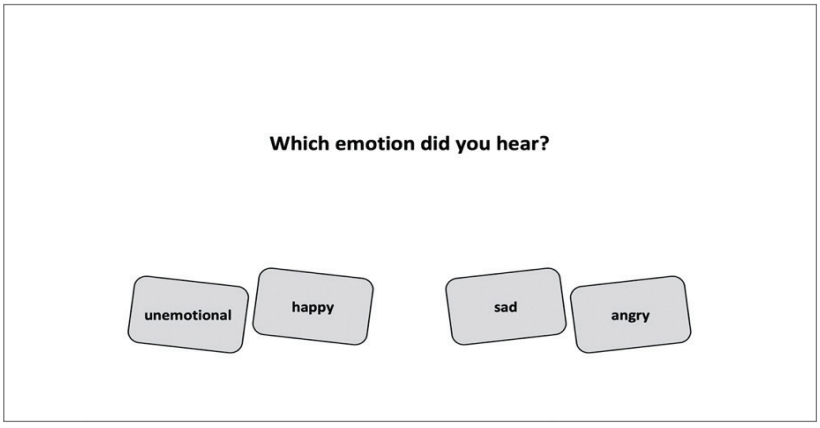


Figure 3.1. Prompt screen displaying response alternatives for vocal emotion recognition task.

3.3.4 Data analysis

The data from the 4-AFC task were used to assess listeners' accuracy in recognising vocal emotions for each speech condition, and for exploring differences in accuracy between consecutive conditions, i.e., in the order a) *natural*, b) *intensity+rate*, c) *F0*, d) *intensity+F0*, e) *rate+F0*. Next, confusion matrices were constructed using the response rate for each alternative per presented emotion and speech condition, and from these,

confusion rates were obtained.

The accuracy data (consisting of 4000 trials, with 2606 correct responses, across all *natural* and manipulated speech conditions) were used to fit generalised linear mixed-effects models (GLMMs) with a logit link function using the lme4 package (1.1.21; Bates et al., 2014) in the R language (version 3.6.3; R Core Team, 2020) within the RStudio IDE (version 1.3.959; RStudio Team, 2019). To predict the difference in overall accuracy between consecutive conditions (the hypotheses defined a priori; see Chapter 3.2), sliding difference contrasts were defined between consecutive conditions (Table 3.1). Each contrast constituted a hypothesis and was therefore included in the final model as a fixed effect (Schad et al., 2020), for which the obtained coefficient represented the difference in accuracy between consecutive conditions. The model-building procedure, therefore, deviated from the convention of building models from the simplest to most complex structure supported by the hypotheses (Bates et al., 2015). As such, there was no theoretical basis for the order in which random terms should be added. To simplify the model before making comparisons, the maximal model suitable for the planned contrast coding was fit first. This comprised Contrasts 1–4, defined in Table 3.1, as fixed effects, with uncorrelated random slopes and intercepts of Contrasts 1–4 for each Participant, Sentence and Emotion. Next, random effects terms explaining very little variance (SD<0.01) were removed from the model. Subsequently, a likelihood ratio test was performed for each remaining random term to confirm that the term explained a significant proportion of the variance, and should therefore be retained. Unstandardised coefficients are reported as all predictor variables input into the model are in the same units. Additionally, goodness-of-fit of the final model is reported with the marginal R^2, the variance accounted for by fixed effects, and conditional R^2, the variance accounted for by both fixed and random effects, in tandem ($R^2_{m/c}$; Nakagawa et al., 2017).

To investigate whether the predicted reduction in accuracy between consecutive conditions was consistent across emotions, four additional models—one per emotion—were generated following the same procedure, each with the same four contrasts between consecutive conditions as fixed effects (Table 3.1) with Sentence and Participant as random effects.

Table 3.1. Contrasts between consecutive speech conditions

Contrast	Comparison
Contrast1	*natural* vs. *intensity+rate*
Contrast2	*intensity+rate* vs. *F0*
Contrast3	*F0* vs. *intensity+F0*
Contrast4	*intensity+F0* vs. *rate+F0*

3.4 Results

3.4.1 Attenuating variations in F0 causes the largest reduction in accuracy

In the *natural* condition, where variations in F0, intensity and speech rate were intact, listeners identified the vocal emotions with very high accuracy (proportion correct *M*=0.94, *SD*=0.24). The accuracy with which listeners identified vocal emotions decreased in the four other conditions, and performance was highly variable (Figure 3.2). When variations in intensity and speech rate were attenuated (*intensity+rate*) accuracy fell to *M*=0.89 (*SD*=0.31). The largest reduction in accuracy occurred when variations in F0 were attenuated in the *F0* condition (*M*=0.51, *SD*=0.50). Mean accuracy when variations were attenuated in intensity and F0 (*intensity+F0*: *M*=0.42, *SD*=0.49), as well as speech rate and F0 (*rate+F0*: *M*=0.50; *SD*=0.50), were similar to the *F0* condition, indicating that concurrent attenuation of variations in F0 and intensity or speech rate had little effect on accuracy relative to F0 alone.

To examine overall accuracy without differentiating between emotions, the following model was employed: *Accuracy* ~ *Contrast1* + *Contrast2* + *Contrast3* + *Contrast4* + (1+ *Contrast2*||*Participant*) + (1|*Sentence*) + (1 + *Contrast2* + *Contrast3* + *Contrast4* ||*Emotion*). In this model, the fixed effects accounted for just over one-third of the variance in the data ($R^2_{m/c}$=0.38/0.55). The inclusion of uncorrelated random intercepts and slopes of *Contrast2* per *Participant* accounted for a significant amount of variance ($\chi^2(1)=33.76$, $p<0.001$), as did random intercepts per *Sentence* ($\chi^2(1)=53.82$, $p<0.001$). Uncorrelated random intercepts and slopes of *Contrast2* ($\chi^2(1)=185.14$, $p<0.001$), *Contrast3* ($\chi^2(1)=17.26$, $p<0.001$), and *Contrast4* ($\chi^2(1)=41.27$, $p<0.001$) per *Emotion* also improved the model fit. Significant effects of *Contrast1* and *Contrast2* were observed (Figure 3.2), indicating that accuracy was reduced when variations in intensity and rate were simultaneously attenuated relative to the *natural* condition (*Contrast1*, β=0.72, *SE*=0.19, *z*=3.73, $p<0.001$), and that attenuated F0 variations led to reduced accuracy relative to simultaneously attenuated variations in intensity and rate (*Contrast2*, β=2.84, *SE*=0.85, *z*=3.33, *p*=0.001). The smaller coefficient for *Contrast1* relative to *Contrast2* provides evidence that while listeners did use intensity and speech rate cues to identify vocal emotions accurately, listeners relied more heavily on F0 cues. *Contrast3* (β=0.47, *SE*=0.47, *z*=0.88, *p*=0.381) and *Contrast4* (β=0.45, *SE*=0.45, *z*=-0.80, *p*=0.424) were not significant, suggesting that attenuating variations in intensity and F0 simultaneously did not reduce accuracy more than attenuating F0 variations, and that attenuating variations in speech rate and F0 concurrently did not reduce accuracy more than attenuating variations in intensity and F0 concurrently. In other words, listeners were not able to make use of intensity cues when F0 cues were uninformative or speech rate cues when

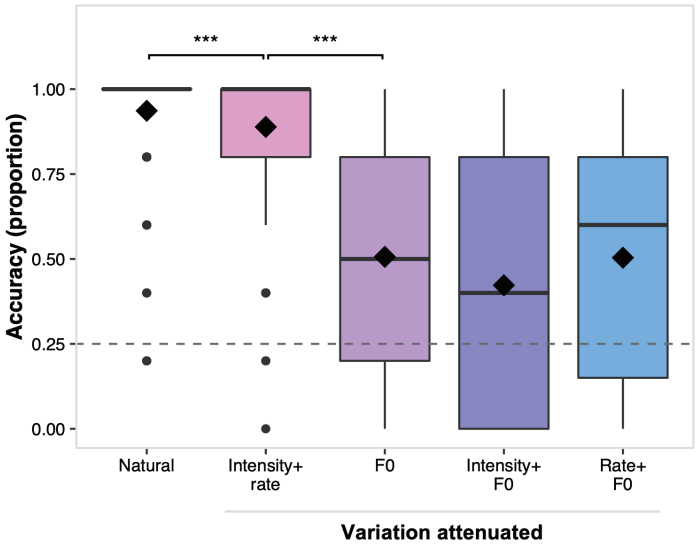


Figure 3.2. Accuracy, as the proportion of correct responses, aggregated across emotions per condition. The solid bar indicates median, the diamond indicates mean, the box includes inter-quartile range (IQR), the whiskers indicate values within 1.5 times the IQR above the 75th or below the 25th percentile, the dots indicate values outside 1.5 times the IQR. The dashed line indicates chance level. Significance for sliding difference contrast, ***p<0.001.

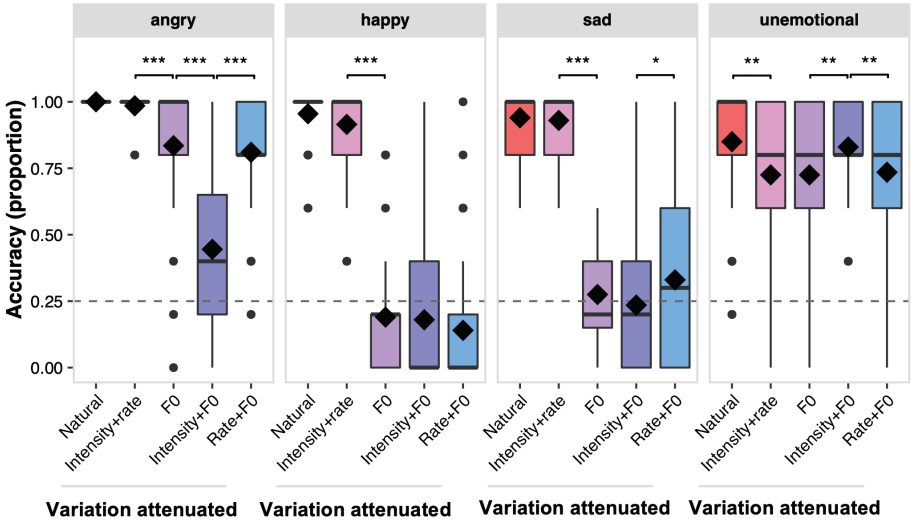


Figure 3.3. Accuracy, as the proportion of correct responses for each emotion in each condition. Solid bar indicates median, diamond indicates mean, the box includes inter-quartile range (IQR), whiskers indicate values within 1.5 times the IQR above the 75th or below the 25th percentile, dots indicate values outside 1.5 times the IQR. Dashed line indicates chance level. Significance for sliding difference contrast, *p<0.05, **p<0.01, ***p<0.001.

both intensity and F0 cues were uninformative. In summary, to identify vocal emotions successfully, listeners afforded most weight to F0 cues. The combination of intact intensity and speech-rate cues supported accurate recognition of vocal emotion only when F0 cues were also intact.

For *angry* speech, the best model was *Accuracy ~ Contrast1 + Contrast2 + Contrast3 + Contrast4 +* (1 + *Contrast2||Participant*). Nearly all the variance in the *angry* data was accounted for by the fixed effects ($R^2_{m/c}$=0.93/0.95). The uncorrelated random slopes and intercepts of *Contrast2* per *Participant* ($\chi^2(1)$=5.00, p=0.025) explained a significant proportion of the variance. *Contrast1* was not significant (β=15.69, *SE*=83.88, z=0.19, p=0.721), suggesting that attenuating variations in intensity and speech rate did not reduce accuracy significantly, i.e., that intact F0 cues alone were sufficient to support accurate recognition of *angry*. Significant effects of *Contrast2*, *Contrast3* and *Contrast4* were observed (Figure 3.3). Listeners' accuracy was reduced when F0 variations were attenuated relative to when variations in intensity and rate were simultaneously attenuated (*Contrast2*, β=3.10, *SE*=0.79, z=3.90, p<0.001), indicating that listeners relied more heavily on F0 cues than intensity and speech-rate cues combined to recognise *angry*. Accuracy was further reduced when intensity and F0 variations were concurrently attenuated relative to when variations in F0 were attenuated (*Contrast3*, β=2.46, *SE*=0.30, z=8.31, p<0.001), providing evidence that listeners made use of intensity cues to recognise *angry* when F0 cues were uninformative. Increased accuracy was observed for the simultaneous attenuation of variations in speech rate and F0 relative to the simultaneous attenuation of variations in intensity and F0 (*Contrast4*, β=-2.24, *SE*=0.29, z=-7.86, p<0.001), indicating that listeners were not able to make use of speech rate cues when F0 and intensity cues were uninformative. Mean accuracy was above chance level in all conditions for *angry*. Listeners afforded most weight to F0 cues in adjudging *angry* speech. When these were attenuated, listeners were able to make use of intensity cues, but not speech-rate cues, to recognise *angry* successfully.

For *happy* speech, (*Accuracy ~ Contrast1 + Contrast2 + Contrast3 + Contrast4 +* (1 + *Contrast2||Participant*) +(1|*Sentence*) was the best model. The fixed effects accounted for approximately two-thirds of the variance in the data ($R^2_{m/c}$=0.63/0.81). A significant proportion of the variance was explained by the uncorrelated random slopes of *Contrast2* per *Participant* ($\chi^2(1)$=9.92, p=0.002) and random intercepts of *Sentence* ($\chi^2(1)$=8.66, p=0.003). *Contrast1* was not significant (β=0.85, S*E*=0.38, z=0.41, p=0.680), meaning that relative to *natural* speech, the ability of listeners to identify vocal emotions was unaffected by the simultaneous attenuation of variations in intensity and speech rate; F0 cues alone appear to be sufficient to support recognition of *happy*. A significant effect of *Contrast2* was observed (Figure 3.3), indicating reduced accuracy when variations in F0 were attenuated relative to when variations in intensity

and speech rate were simultaneously attenuated (β=6.08, *SE*=0.68, *z*=8.90, *p*<0.001), confirming that listeners depended on F0 cues to recognise *happy*. Additionally, *Contrast3* (β=0.10, *SE*=0.32, *z*=0.32, *p*=0.752), and *Contrast4* (β=0.44, *SE*=0.33, *z*=1.31, *p*=0.188) were not significant, suggesting listeners did not make use of intensity cues when F0 cues were uninformative or speech-rate cues when both intensity and F0 cues were uninformative. Recognition of *happy* was at chance level for conditions in which variations in F0 were attenuated; for *happy*, listeners relied heavily on F0 cues and did not make use of intensity or speech-rate cues to compensate for uninformative F0 cues.

For *sad* speech, the best model was *Accuracy* ~ *Contrast1* + *Contrast2* + *Contrast3* + *Contrast4* + (1 + *Contrast2*||*Participant*) + (1|*Sentence*). Slightly over half of the variance in the data was accounted for by the fixed effects ($R^2_{m/c}$=0.56/0.67). The uncorrelated random slopes of *Contrast2* per *Participant* ($\chi^2(1)$=23.82, *p*<0.001) and random intercepts of *Sentence* ($\chi^2(1)$=34.71, *p*<0.001) explained a significant amount of variance. *Contrast1* was not significant (β=0.19, *SE*=0.32, *z*=0.43, *p*=0.666), suggesting the ability to recognise *sad* was not impacted by simultaneous attenuation of variations in intensity and speech rate, i.e., that intact F0 cues alone support the successful recognition of *sad* speech. A significant effect of *Contrast2* (β=4.73, *SE*=0.52, *z*=9.01, *p*<0.001) indicates that, relative to simultaneously attenuated variations in intensity and speech rate, attenuating variations in F0 reduced accuracy. *Contrast3* (β=0.27, *SE*=0.26, *z*=1.04, *p*=0.299) was not significant, indicating that, relative to attenuated variations in F0, accuracy with which vocal emotions are identified was not reduced by simultaneous attenuation of variations in intensity and F0; listeners did not make use of intensity cues when F0 cues were uninformative. A significant effect of *Contrast4* was observed (β=-0.61, *SE*=0.26, *z*=-2.39, *p*=0.017) suggesting that accuracy increased when variations in speech rate and F0 were simultaneously attenuated relative to simultaneously attenuated variations in intensity and F0. For *sad*, accuracy decreased to near chance level in all speech conditions with uninformative F0 cues, indicating that listeners relied mainly on these cues to recognise *sad* speech. However, listeners may have also made use of intensity cues when F0 and speech-rate cues were simultaneously uninformative.

For *sad*, the initial random-effects structure of the model included random slopes of each contrast per *Sentence*, designed to detect possible variation introduced by the signal processing used to create the different speech conditions; notably, when a random slope of *Contrast4* per *Sentence* was included, a non-significant difference is observed for *Contrast4* suggesting that accuracy differs substantially for certain stimulus sentences between *sad* in the *intensity+F0* and *rate+F0* conditions. However, consistent with the stepwise model-fitting approach (see Chapter 3.3.4), the random slope of *Contrast4* per *Sentence* was excluded from the final model. Thus, the significant effect of *Contrast4* for *sad* should be interpreted with caution.

Finally, for *unemotional* speech, the best model was *Accuracy ~ Contrast1 + Contrast2 + Contrast3 + Contrast4* + (1+ *Contrast2||Participant*) +(1|*Sentence*). The uncorrelated slopes of *Contrast2* per *Participant* ($\chi^2(1)=20.69$, $p<0.001$), and random intercepts of *Sentence* ($\chi^2(1)=23.75$, $p<0.001$) accounted for a significant proportion of the variance. The fixed effects accounted for little of the variance in the data ($R^2_{m/c}=0.04/0.42$), meaning that the significant effects should be interpreted cautiously. Significant effects of *Contrast1*, *Contrast3*, and *Contrast4* were observed (Figure 3.3), while *Contrast2* was not significant. The significant effect of *Contrast1* (β=1.84, *SE*=0.34, *z*=5.34, $p<0.001$) indicates that, relative to *natural* speech, simultaneously attenuating variations in intensity and speech rate reduced listeners' ability to recognise *unemotional* speech. The non-significant *Contrast2* (β=-0.01, *SE*=0.37, *z*=-0.01, *p*=0.990) suggests that uninformative F0 cues did not impact listeners' abilities to adjudge *unemotional* speech, as accuracy was similar when variations in intensity and speech-rate were simultaneously attenuated and when variations in F0 were attenuated. The significant effects of *Contrast3* (β=-0.85, *SE*=0.28, *z*=-3.01, *p*=0.003) and *Contrast4* (β=0.78, *SE*=0.28, *z*=2.75, *p*=0.006) reflect reduced accuracy when variations in intensity and F0 were simultaneously attenuated relative to attenuated variations in F0 alone, and reduced accuracy when variations in F0 and speech rate were attenuated simultaneously relative to simultaneous attenuation of F0 and intensity variations. This can be cautiously interpreted as evidence that listeners may have been able to make use of intensity cues when F0 cues were uninformative and speech rate cues when F0 and intensity cues were uninformative. Accuracy was well above chance in all conditions for *unemotional* speech.

3.4.2 Uninformative cues cause listeners to mistake emotional for unemotional speech

The distribution of confusions—incorrect responses—across the response alternatives illustrates how listeners confuse vocal emotions and whether they do so systematically (Figure 3.4). *Angry* speech was recognised well above chance level in all conditions (proportion correct for *natural*=1, *intensity+rate*=0.98, F0=0.84, *intensity+F0*=0.45, *rate+F0*=0.81). Attenuating variations in intensity and speech rate simultaneously did not influence the recognition of *angry* speech. *Angry* was confused with *unemotional* in 0.16 of trials with attenuated F0 variations (*F0*) and attenuated speech-rate and F0 variations (*rate+F0*), and 0.50 of trials with simultaneously attenuated variations in intensity and F0 (*intensity+F0*).

When variations in F0 were intact, listeners recognised *happy* speech almost perfectly (*natural*=0.96, *intensity+rate*=0.92). However, when variations in F0 were attenuated, recognition of *happy* speech fell below chance level (*F0*=0.19, *intensity+F0*=0.18,

rate+F0=0.14). Of the three conditions where F0 variations were attenuated, the most confusions for *happy* speech occurred when variations in F0 and speech rate were attenuated simultaneously, and *unemotional* was the most commonly selected response (0.72), but *sad* (0.12) was also selected with lower probability.

Like *happy* speech, *sad* speech was recognised with high accuracy in conditions where F0 variations were intact (*natural*=0.94, *intensity+rate*=0.93) and near chance level when variations in F0 were attenuated (*F0*=0.28, *intensity+F0*=0.24, *rate+F0*=0.33). In the conditions where variations in F0 were attenuated, *happy* was most commonly confused with *unemotional* (*F0*=0.70, *intensity+F0*=0.74, *rate+F0*=0.64).

For *unemotional* speech, high performance, well above chance level, was observed in all conditions (*natural*=0.85, *intensity+rate*=0.72, *F0*=0.73, *intensity+F0*=0.83, *rate+F0*=0.74). Listeners occasionally mistook *unemotional* stimuli for *sad* in speech conditions with attenuated F0 variations (*natural*=0.09, *intensity+rate*=0.21, *F0*=0.22, *intensity+F0*=0.14, *rate+F0*=0.23). The confusion rates for all emotions and conditions considered together demonstrate that listeners most often confuse emotions conveyed in speech with uninformative acoustic cues as *unemotional*.

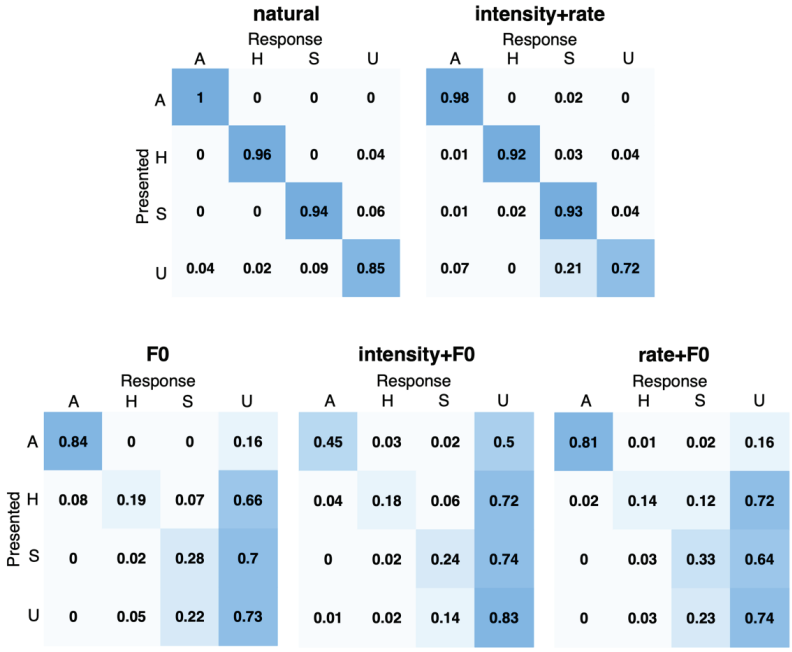


Figure 3.4. Confusion matrices for all conditions. Proportion of responses (columns) per presented emotion (rows). With the exception of natural, conditions labels name the acoustic features within which variations were attenuated. Colour indicates the proportion of responses per response alternative for the presented emotions. Chance level is 0.25. A=*angry*, H=*happy*, S=*sad*, U=*unemotional*.

3.5 Discussion

3.5.1 Uninformative F0 cues reduce listeners' abilities to recognise vocal emotions

The goal of this study was to investigate how NH listeners weight the cues provided by acoustic features (F0, intensity and speech rate) to recognise vocal emotions, with a specific focus on listeners' abilities to exploit intensity and speech-rate cues when F0 cues are uninformative. The data demonstrate that listeners are proficient at recognising emotions in natural speech and that rendering F0 cues uninformative by attenuating variations in F0 reduces the accuracy with which listeners can identify vocal emotions. To a lesser degree, rendering intensity and speech-rate cues simultaneously uninformative also reduces the accuracy with which listeners recognise vocal emotions. The larger reduction in accuracy that arises when F0 cues are uninformative supports the view that listeners weight F0 most heavily in their decisions, consistent with existing evidence for the central role of F0 cues in recognising vocal emotions (Juslin & Laukka, 2001; Scherer et al., 2003). Moreover, this finding is consistent with previous investigations of recognition of vocal emotions by CI listeners and NH listeners presented CI-simulations (Chatterjee et al., 2015; Everhardt et al., 2020; Gilbers et al., 2015; Jiam et al., 2017; Luo et al., 2007; Most & Aviner, 2009; Nakata et al., 2012; Pak & Katz, 2019; Panzeri et al., 2021; Pereira, 2000; Peters, 2006; Ren et al., 2021; Waaramaa et al., 2018) for whom F0 cues are reduced, and thus less informative.

Successful recognition of *happy* speech also depends on informative F0 cues (Figure 3.3): Mean accuracy drops below chance, indicating that the cues provided by variations in intensity and speech rate (and, indeed, other potential acoustic features) are insufficient to support the recognition of *happy* as a vocal emotion. This is consistent with previous reports in which variations in F0 were attenuated (Leitman et al., 2010; Metcalfe, 2017; Pell, 1998), and with reports that F0 offers the most robust characterisation of happy speech (Kamiloğlu et al., 2020).

Listeners depend on F0 cues to recognise *sad* speech, with mean accuracy at, or just above, chance level for all conditions where F0 cues were rendered uninformative. This is inconsistent with the reported behaviour of NH listeners, in which accuracy in identifying *sad* speech remains well above chance when F0 cues are uninformative (Metcalfe, 2017; Pell, 1998). In the present study, the *sad* stimuli recorded from a single female speaker, mean F0 is the only investigated acoustic feature that differs considerably between *sad* and *unemotional* speech (for acoustic analysis see Figure 3.1, and Chapter 2.1). Pell's (1998) *sad* and *unemotional* (neutral) stimuli recorded from a single female, are quite similar in mean F0 and mean intensity information is not reported. As Pell also does not describe normalising intensity, the higher accuracy for *sad* with uninformative F0 cues reported by Pell (1998), relative to the present study, may suggest that Pell's

sad and *unemotional* stimuli differ substantially in mean intensity, increasing discriminability, where *sad* and *unemotional* speech in the present study are quite similar in mean intensity. Metcalfe's (2017) stimuli were recorded from ten speakers with overall intensity normalised. Stimuli recorded from additional speakers are likely required to confirm the source of differences in the accuracy with which *sad* speech was identified between the present and previous studies. Moreover, stimuli recorded from additional speakers are needed to demonstrate whether listeners are truly able to make use of intensity cues to recognise *sad* when both F0 and speech-rate cues are uninformative or whether this finding is mediated by the single recorded speaker's way of conveying sadness rather than a generalisable group-level effect.

When F0 cues are uninformative, mean accuracy for identifying *angry* speech is reduced but remains above chance level, suggesting that while rendering F0 cues uninformative impacts recognition of *angry* speech, it is not sufficient to render *angry* unrecognisable completely. This mirrors Pell's (1998) findings closely but diverges from Metcalfe's (2017) report of accuracy scores of ~33% (with chance level of 20%) when variations in F0 cues are attenuated. One possible explanation for the high accuracy with which *angry* speech was identified is that listeners may have made use of overall sound intensity rather than variations in intensity across the utterance. Variations in intensity were only attenuated for *intensity+rate* and *F0+intensity*, but not for *natural*, *F0* and *rate+F0*. Therefore, the overall intensity of *angry* in the *natural*, *F0* and *rate+F0* conditions was higher than in the *intensity+rate* and *F0+intensity* conditions. Pell (1998) does not describe normalising overall intensity between emotions, whereas Metcalfe (2017) does. This may explain why the accuracy scores for *angry* with attenuated F0 variations are more similar to those described by Pell (1998) than by Metcalfe (2017). While evidence exists to support the notion that normalisation of overall intensity does not influence recognition of vocal emotions (Gilbers et al., 2015; Van Lancker & Sidtis, 1992; Yanushevskaya et al., 2013), resolving the importance of overall intensity in recognising vocal emotions requires both to be assessed. Nevertheless, it is reasonable to conclude accuracy in recognising *angry* is reduced when variations in intensity across utterances are attenuated, compared to conditions where intensity cues (instantaneous and overall) are intact and informative. Translated to listeners' abilities to utilise intensity cues, this means that the present study is unable to disentangle how listeners make use of overall and instantaneous variations in intensity but does support the view that listeners make use of intensity cues to recognise *angry* when F0 cues are uninformative. This suggests that listeners may increase their reliance on intensity cues to compensate for uninformative F0 cues to identify specific emotions in speech.

Another explanation for the high accuracy in *angry* across most speech conditions is that listeners may exploit cues in acoustic features not controlled for in this study. For

example, they might exploit aspects of voice quality, such as soft or strained sounding voice (Laukkanen et al., 1997; Leitman et al., 2010; Patel et al., 2011; Waaramaa et al., 2010; Yanushevskaya et al., 2013), or dissonance cues associated with timbre. Further studies should control for voice quality or include voice quality among the investigated cues.

3.5.2 With uninformative cues, *angry*, *happy*, and *sad* speech sound *unemotional*

A consequence of attenuating variations to the mean value for each acoustic feature is that some acoustic features may be altered more than others, or altered in a different direction than others (i.e., speech rate decreased for *happy* but increased for *angry*). Previous studies assess confusions in identifying vocal emotions to determine which acoustic features can provide misleading cues (e.g., Banse & Scherer, 1996; Pakosz, 1983; Paulmann & Uskul, 2014). In addition to highlighting misleading cues, confusion rates reveal whether or not listeners rely on the relative positions and distances between emotions within each acoustic feature.

When variations in F0 are attenuated, the mean F0 of *sad* speech is changed very little, whereas that of *angry* and *happy* is reduced and the mean F0 of *unemotional* speech is raised (*F0*, *intensity+F0* and *rate+F0* conditions in Figure 2.4). It follows that if mean F0 is the primary cue for emotion recognition, emotions with attenuated variations in F0 would most plausibly be confused with *sad*. However, *unemotional* is the most common response in incorrect trials. Confusions with *sad* are few in conditions where variations in F0 are attenuated, supporting the consensus that listeners make use of variations in F0 rather than mean F0 as cues with which to recognise vocal emotions (Mozziconacci, 1998; Rodero, 2011).

The mean intensity of *happy* speech is closest to the target value to which variations in intensity were attenuated, i.e., the mean intensity across all emotions. Attenuating intensity variations results in an overall reduction in intensity for *angry* and an increase for each *unemotional* and *sad* speech (Figure 2.4, *intensity+rate* and *F0+intensity* conditions). If relative overall intensity were favoured as a cue, then *happy* would likely be the most frequent confusion in conditions with attenuated intensity variations. This was not the case, confirming that listeners cannot make use of variations in intensity alone as cues for vocal emotions (Gilbers et al., 2015; Van Lancker & Sidtis, 1992; Yanushevskaya et al., 2013). Overall, accuracy scores for *angry* were lower when variations in intensity were attenuated simultaneously with F0 than when just F0 variations were attenuated, or when variations in F0 and speech rate were simultaneously attenuated. This demonstrates that listeners do indeed make use of variations in intensity, or possibly overall intensity, as cues to identify *angry* speech.

The mean speech rate across all emotions is closest to that of *sad* and *unemotional* speech. In the process of attenuating variations in speech rate, four of five of both *sad* and *unemotional* utterances become slightly faster, while *angry* speech became substantially faster, and *happy* speech substantially slower (refer *intensity+rate* and *rate+F0* conditions in Figure 2.4). As *unemotional* and *sad* speech are altered least, it might be expected that listeners mislabel *angry* or *happy* speech, with their slower speech rates, as *sad* or *unemotional*. Accordingly, *happy* speech was misidentified as *sad* (13% of the time) when both F0 and speech-rate variations were attenuated. This suggests that when F0 cues are uninformative, some listeners may make use of speech-rate cues to identify vocal emotions.

The main incorrect response to vocal emotions with any uninformative cues was to confuse *happy*, *sad*, and *angry* speech with *unemotional* speech. This is consistent with previous reports (Luo et al., 2007; Metcalfe, 2017), as well as Metcalfe's (2017) suggestion that listeners select 'neutral' or *unemotional* when uninformative cues make recognising emotions more challenging. Metcalfe (2017) also suggests that listeners may use 'neutral' or *unemotional* as a 'not sure' response among the alternatives in a forced-choice task, an interpretation at least consistent with the data presented here. Future research may consider including an 'not sure' alternative, however this should be considered carefully, as factors including motivation and level of education can mediate participants' selection of 'not sure', and the inclusion of an additional response alternative will reduce the statistical power of the analyses (Krosnick et al., 2002).

3.5.3 Relevance to hearing devices

Of the conditions presented in this study, the *F0* and *intensity+F0* conditions are most comparable to the degraded speech signals delivered to hearing-impaired listeners who utilise hearing devices, especially CIs, but also HAs. CIs transmit speech in trains of electrical pulses, whereby the spectrotemporal resolution and dynamic range of intensity of the speech are reduced, and F0 cues are only weakly represented in the delivered speech (Başkent et al., 2016; Chatterjee & Peng, 2008; Everhardt et al., 2020; Plack, 2018). F0 information can be transmitted through HAs, yet the pitch information delivered by a HA can be influenced by changes to the spectrum and temporal envelope of speech and dynamic range of intensity introduced by a listener's personalised amplification settings (tailored to the severity and profile of their hearing loss; Goy et al., 2018; Lesica, 2018). While the reduced F0 information in speech transmitted by a CI renders recognising emotions in speech difficult (Everhardt et al., 2020), the speech signal transmitted by HAs, altered due to hearing loss and HA settings, could impact emotion recognition negatively, although likely to a lesser degree than CIs (Picou et al., 2018). The reduced overall accuracy observed in this study, when F0 cues alone were

uninformative (*F0*=0.51) or intensity and F0 cues were both uninformative (*intensity+F0*=0.42), are consistent with the reduced accuracy in recognising vocal emotions by CI listeners. Further, based on the observation that recognition of *happy* speech by NH listeners is most dependent on informative F0 cues, it is likely that CI listeners struggle disproportionately to identify *happy* speech, and this contention is supported empirically (Hopyan-Misakyan et al., 2009; Luo et al., 2007; Most & Aviner, 2009; Nakata et al., 2012; Z. Zhu et al., 2018).

To compare the current study with investigations of vocal emotion recognition in CI users, the difference in mean accuracy was calculated between the *natural* and *F0* conditions from the present study and between NH and each CI and HA listeners previously reported. Comparisons drawn with previous studies must take into account whether the NH (control) group obtained accuracy scores similar to those observed for *natural* speech in the present study (i.e., all approximately equally far above chance level) to ensure the validity of comparisons. Most & Aviner (2009) offer an ideal point of comparison, as they report accuracy scores for six emotions conveyed with no cues attenuated (i.e., *natural* speech) in several groups, including adolescent NH, early- and late-implanted CI users (implanted before or after age of 6 years, respectively) diagnosed with prelingual hearing loss, and HA listeners with prelingual severe to profound bilateral hearing loss. Accuracy for *angry*, *happy*, and *sad* (all 0.62) in Most & Aviner's NH group is lower than that observed for *natural* speech in the present study (0.94). The stability of the accuracy observed across these emotions for Most & Aviner's NH group suggests that the lower accuracy, relative to the present *natural* condition, may be explained by the inclusion of additional emotions (*surprise*, *fear* and *disgust*) in the forced-choice task. The differences in accuracy between Most & Aviner's NH and CI groups are largest for *happy*, followed by *sad*, then *angry* (accuracy in early-implanted CI users subtracted from accuracy in NH listeners: *happy*=0.42, *sad*=0.32, *angry*=0.14, late-implanted CI subtracted from NH: *happy*=0.46, *sad*=0.34, *angry*=0.28). Comparisons between the present study's *natural* and *F0* conditions (*happy*=0.77, *sad*=0.66, *angry*=0.16) exhibit the same order of decreasing differences in accuracy. The consistent relative magnitude of differences (largest for *happy*, then *sad*, then *angry*) suggests that the present study's *F0* condition and current CI devices render F0 information similarly uninformative.

Pak & Katz (2019) also report accuracy scores obtained for NH listeners and young-adult, mainly late implanted, CI users for emotions conveyed in natural speech; in their study, *happy* does not exhibit the largest difference between the two groups (*happy*=0.25, *sad*=0.46, *angry*=0.10), likely due to the low recognition of *happy* by their NH listeners (0.48), relative to the other emotions (*angry*=0.83, *sad*=0.93). Even so, the difference between the groups in Pak & Katz' (2019) study is smaller for *angry*

than *sad*, as observed when comparing the *natural* and *F0* conditions in the present study. Metcalfe's (2017) NH listeners recognised emotions in *natural* speech with high accuracy (mean accuracy across *angry*, *happy*, and *sad*=0.85) much like NH listeners in the present study's *natural* condition (0.94). Upon subtracting the accuracy reported for Metcalfe's NH listeners in the CI-simulated condition from the natural speech condition (difference as proportions, *happy*=0.54, *sad*=0.19, *angry*=0.22), the differences for *happy* and *angry* are comparable to those observed between the *natural* and *F0* conditions in this study, and the difference for *sad* in Metcalfe (2017) is substantially smaller. This could be explained by the single speaker's way of conveying sadness in the present study, as Metcalfe's *sad* stimuli, recorded from ten speakers, were also well recognised by NH listeners when the variations in F0 were attenuated.

Further, for *happy* and *sad* speech, the accuracy with which CI listeners recognise vocal emotions in natural speech is consistently higher (Luo et al., 2007; Most & Aviner, 2009; Pak & Katz, 2019) than the ability of NH listeners in the present study to recognise emotional speech when F0 cues alone were rendered uninformative. This could, in part, be due to CI listeners being experienced listeners of speech with uninformative F0 cues, a position supported by evidence that listeners' abilities and experience perceiving acoustic cues (Jasmin et al., 2019, 2021), such as F0 and duration, are positively associated with listeners' reliance upon these cues. It is possible, that given more time to gain experience listening to the stimuli, NH listeners' abilities to recognise *happy* and *sad* speech with uninformative F0 cues would improve. However, M. H. Davis et al. (2005) demonstrate that listeners' abilities to adapt to degraded speech is reduced when the speech does not include meaningful lexical items. As such, participants in the present study would likely require substantially more exposure to the pseudo-English speech with attenuated F0 variations before learning to recognise *happy* and *sad* speech.

3.5.4 Strengths and limitations

The systematic attenuation of variations within acoustic features that convey vocal emotions is an important strength of this study. It provides a means of investigating listeners' abilities to recognise emotions conveyed in speech in a variety of degraded listening conditions, as well as demonstrating NH listeners' limited capacity (within the limited duration of exposure to materials during the testing) to compensate for uninformative cues by extracting emotional information from any remaining informative cues. The statistical approach is strengthened through the use of random intercepts per stimulus sentence to account for differences between the five sentences presented in each emotion (sentences described in Chapter 2.1), as well as sentence-specific differences incurred by the signal processing used to create stimuli (i.e., the possibility of acoustic transients generated by attenuating variations in sound intensity; see Chapter

2.1 for further explanation). Moreover, the coding and implementation of sliding difference contrasts, with the planned contrasts defined a priori, allows hypotheses to be tested with increased statistical power, reducing the likelihood of Type II errors (M. J. Davis, 2010). However, there are also some noteworthy limitations. First, stimuli were generated for use in subsequent fNIRS experiments, informed by behavioural accuracy obtained from the present experiment. A key fNIRS constraint is that several presentations (>10) of each stimulus item with long inter-stimulus intervals are required (>10 s) within a relatively short recording session (<1 hour). To respect this constraint, the stimuli were limited to one female speaker, four emotions, and five conditions. Additional speakers of varying voices, ages, and genders would potentially improve the stimuli, as expressions of vocal emotions are known to vary between individuals (Banse & Scherer, 1996). According to Banse & Scherer (1996), the inclusion of more than six emotions would ensure that the task was purely a recognition task, rather than a discrimination task. Moreover, including three further speech conditions in which variations in acoustic features conveying vocal emotions are attenuated, (i.e., speech-rate alone, intensity alone, and F0, intensity and speech-rate combined), would have completed the systematic assessment of listeners' ability to make use of the most potentially informative acoustic cue(s) present in any speech utterance. It is also noted that the signal processing used to attenuate variations within each acoustic feature introduce slight changes to other acoustic features (as described in Chapter 2.1), although this is likely the case for all studies in which acoustic features are manipulated. Measures were taken to minimise these changes where possible. Finally, normalising overall intensity across all stimuli would allow for valid between-emotion comparisons of the effects of attenuating variations in the different acoustic features.

3.6 Conclusions

The present study confirms and extends previous reports of the accuracy of NH listeners in recognising vocal emotions when all acoustic features (i.e., F0, intensity and speech rate) are intact. The data suggest listeners recognise vocal emotions significantly less accurately when F0 cues are uninformative, and that the same is true, although with a smaller reduction in accuracy, when intensity or speech-rate cues are simultaneously rendered uninformative (i.e., only F0 cues intact). Moreover, the data suggest that listeners are unable to make use of intensity and speech-rate cues, separately or together to recognise vocal emotions successfully when F0 cues are uninformative. Using an extended vocal emotion paradigm in which variations in F0, intensity and speech rate are systematically attenuated to render cues uninformative, one or two acoustic features at a time, combined with a robust statistical approach and confusion matrices, the present study contributes benchmark behavioural accuracy scores to inform future neuroimaging studies on vocal emotion recognition.

Chapter 4| Illuminating emotion: Assessing functional near-infrared spectroscopy's sensitivity to discrete vocal emotions

4.1 Abstract

Background. Normal-hearing listeners are skilled at recognising emotions conveyed in speech. Vocal emotion has been the focus of many functional magnetic resonance imaging (fMRI) studies in normal-hearing (NH) listeners that implicate cortical regions including the bilateral—right more strongly than left—superior temporal gyri (STG), the inferior frontal gyri (IFG), and middle frontal gyri (MFG) in vocal emotion processing. Recent functional near-infrared spectroscopy (fNIRS) studies in NH listeners find cortical representations of individual vocal emotions, suggesting that fNIRS may be a promising technique with which to investigate normal and impaired vocal emotion recognition.

Aims. The present study investigated the sensitivity of fNIRS to cortical representations of discrete emotions conveyed in speech.

Method. In a block-design listening task, 21 NH listeners listened to ~7.25-s blocks of *angry*, *happy*, *sad*, and *unemotional* pseudo-English speech and 7.25 s of silence as a control condition. During the task, fNIRS was used to measure changes in cortical blood flow in three regions of interest (ROIs): bilateral STG, IFG, and MFG.

Results. Speech stimuli (*angry*, *happy*, *sad*, and *unemotional* together) evoked increased haemodynamic activity in bilateral STG relative to silence. Moreover, increased haemodynamic activity in right relative to left STG was evoked by emotional speech (*angry*, *happy*, and *sad* together). No significant differences in haemodynamic response amplitude were found when comparing each *angry*, *happy*, *sad* with *unemotional* in any of the ROIs.

Conclusions. Speech stimuli elicited haemodynamic responses bilaterally in STG. The data confirm the involvement of the right hemisphere in processing emotional speech. However, no evidence of cortical representations of discrete emotions was observed, suggesting that fNIRS may not be adequate for investigating the cortical processing of individual vocal emotions.

4.2 Introduction

Healthy, normal-hearing listeners recognise emotions in speech with ease, facilitating successful social interactions (e.g., Cannon & Chatterjee, 2018; Chatterjee et al., 2015; Christensen et al., 2019; Demenescu et al., 2014; House, 1994; Most & Aviner, 2009). In contrast, impaired recognition of vocal emotions is known to present social challenges in those with hearing impairment (Christensen et al., 2019; Everhardt et al., 2020; Goy et al., 2018), of advanced age (Christensen et al., 2019; Demenescu et al., 2014), or suffering a range of disorders such as autism (Philip et al., 2010; Tobe et al., 2016), schizophrenia (Corcoran et al., 2015; Simpson et al., 2013), Parkinson's disease (Péron et al., 2012), or alcoholism (Kornreich et al., 2013). The consequences of impaired recognition of vocal emotions include reduced quality of life (hearing impaired; Luo et al., 2018; Schorr et al., 2009), and increased burden on caregivers for populations with general cognitive impairment (Parkinson's or Alzheimer's disease; Martinez et al., 2018).

An objective measure of recognising vocal emotions, i.e., a neural signature for various discrete emotions conveyed by speech, in particular, would be a valuable addition to subjective and psychoacoustic measures in research and clinical settings (Boehner et al., 2007) and would likely benefit the development of brain-computer interfaces and autonomous listening devices (He et al., 2020; F. Wang et al., 2018). One potential tool is functional near-infrared spectroscopy (fNIRS), a relatively new non-invasive neuroimaging technique that measures changes in the metabolism of cortical oxygen using low levels of near-infrared light. fNIRS is suited to measuring cortical haemodynamic signatures of vocal emotions in a variety of populations, because it is quiet, non-invasive, does not interact with implanted ferrous metal devices, and does not require participants to lie in an enclosed space as in functional magnetic resonance imaging (fMRI; Peelle, 2017; Quaresima & Ferrari, 2019).

The blood oxygen level dependent (BOLD) response measured with fMRI and the haemodynamic responses to vocal emotions measured with fNIRS both capitalise on the phenomenon of neurovascular coupling (Maggioni et al., 2015; Steinbrink et al., 2006); neural activity incurs increased blood flow, leading to an increase in the concentration of oxygenated haemoglobin (HbO) and a reduction in the concentration of deoxygenated haemoglobin (HbR) in the active region (Scholkmann, Kleiser, et al., 2014; Wolf et al., 2002). The substantially reduced spatial resolution of fNIRS (maximum resolution ~10 mm, with depth sensitivity ~15 mm below the scalp) compared to fMRI (maximum resolution ~0.3-mm voxels, sensitive to the whole brain) is mitigated by its improved temporal resolution of up to 10 Hz relative to 1–3 Hz in fMRI (Pinti et al., 2020).

To record changes in cortical oxygen metabolism with fNIRS, light sources and light

detectors (sensors collectively referred to as optodes) are placed in pairs on the scalp to create 'channels'. Sources emit near-infrared light, photons of which penetrate (and are scattered by) the scalp, skull, and cortex. Some of these photons are absorbed as they pass through HbO and HbR (collectively called chromophores) in the vasculature while others are reflected back to the scalp, where they are captured by light detectors. The presence of HbO and HbR in the cortex below a source and detector pair is quantified by the relative difference in intensity of emitted and captured light (Ferrari & Quaresima, 2012; Quaresima & Ferrari, 2019; Scholkmann, Kleiser, et al., 2014). The physical distance between a light source and a light detector on the scalp determines the depth of penetration of the channel (i.e. a 'long' 30-mm separation is suitable for measuring from the cortex, ~15 mm below the scalp, and a 'short' 8-mm separation measures changes in oxygen metabolism in the extracerebral tissue including the scalp; Brigadoi & Cooper, 2015). The target cortical signal recorded with 'long' channels can be isolated by regressing out the extracerebral and physiological signals measured with so-called 'short' channels (Brigadoi & Cooper, 2015; R. Saager & Berger, 2008).

When an NH listener with typical cognition processes vocal emotions, the superior temporal cortex (STC) is active, as well as the frontal cortices, the amygdala, basal ganglia, insula, hippocampus, and cerebellum, as evidenced by neuroimaging (Frühholz, Trost, et al., 2016). According to a model proposed by Frühholz, Trost, et al. (2016), which mainly draws upon fMRI and positron emission tomography (PET) studies, the STC decodes the acoustic signal, extracting acoustic features, while the amygdala appraises the valence of the stimulus, forming an auditory percept. The basal ganglia and thalamus are thought to contribute to the decoding of basic acoustic information. The percept formed by the STC and amygdala is then evaluated and appraised by the medial frontal cortex, with access to episodic memory through the involvement of the hippocampus. Finally, an adaptive response (e.g., an autonomic and/or motor response) is initiated by the inferior frontal gyrus (Frühholz, Trost, et al., 2016).

As summarised by Frühholz, Trost, et al. (2016), fMRI studies reveal that emotions conveyed in speech elicit cortical haemodynamic activity bilaterally in STC (Ethofer et al., 2012; Koch et al., 2018; K. H. Lee & Siegle, 2012; Leitman et al., 2010; Robins et al., 2009; Witteman et al., 2012), and IFG (Leitman et al., 2010; Witteman et al., 2012). With fNIRS, D. Zhang et al. (2018) found emotion-evoked activation in bilateral STG and right IFG, replicating the previous fMRI findings, while the remaining fNIRS studies do not include both STG (Anuardi & Yamazaki, 2019; Gruber et al., 2020) or do not report the simple effect of hearing vocal emotions on cortical activation (Sonkaya & Bayazıt, 2018; Steber et al., 2020; Zhen et al., 2021). Vocal emotions are also commonly reported to evoke greater activity in the right, relative to left, hemisphere (Beaucousin et al., 2007; Kotz et al., 2006; Kreitewolf et al., 2014; Seydell-Greenwald et al., 2020;

von Cramon et al., 2003; Witteman et al., 2012). This right-hemisphere bias, which also holds for studies employing fNIRS (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2017; Zhen et al., 2021), is commonly attributed to a right-hemispheric specialisation for decoding spectral information such as contours of the F0, the fundamental frequency or voice pitch, that unfold over relatively long timescales. This is in contrast to the left hemisphere which decodes rapid spectral changes conveying phonemic information, such as formant transitions (Boemio et al., 2005; Poeppel, 2003; Scott & McGettigan, 2013; Zatorre & Belin, 2001). Moreover, explicit, relative to implicit, attention to the vocal emotions enhances the right-lateralised haemodynamic activity in frontotemporal regions (fMRI: Buchanan et al., 2000; Frühholz et al., 2012; Kotz et al., 2013; Wildgruber et al., 2009; Witteman et al., 2012, fNIRS: Zhen et al., 2021), in much the same way that speech-evoked haemodynamic activity bilaterally in the STC is enhanced by attending to the speech (fMRI: Grady et al., 1997; Sabri et al., 2008; fNIRS: Remijn & Kojima, 2010; M. Zhang et al., 2018).

According to Frühholz, Trost, et al. (2016)'s model, the percept of a vocal emotion is generated in the STC and appraised in inferior and medial frontal brain regions. While fMRI studies consistently implicate the right STC in the processing of vocal emotions, the reported cortical haemodynamic activity evoked by discrete emotions, such as *angry*, *happy*, or *sad*, varies between studies (see Table 4.1 for referenced reports). *Angry* speech has been most thoroughly investigated with fMRI and, compared to *unemotional* or neutral speech, studies have reported haemodynamic activity in each the right or bilateral STC, left, right, or bilateral IFG, left or bilateral MFG. fMRI investigations of *happy* speech, relative to *unemotional*, find increased haemodynamic activity in either the left IFG or right STC, while *sad* speech elicits increased haemodynamic activity in either right or left MFG relative to *unemotional*.

Despite variability in the haemodynamic activity observed with fMRI for *angry*, *happy* and *sad* speech, evidence exists from fMRI studies—with its better spatial resolution compared to fNIRS—of spatially defined haemodynamic signatures of discrete emotions (Ethofer, Van De Ville, et al., 2009). Averaging across ROIs in fMRI data reduces the spatial resolution and renders the obtained signatures statistically unreliable (Ethofer, Van De Ville, et al., 2009). The spatial resolution and depth penetrations of fNIRS are considerably inferior to the fMRI spatial resolution described by Ethofer, Van De Ville, et al. (2009), suggesting that fNIRS is highly unlikely to be sensitive to cortical representations of discrete vocal emotions. Consistent with this, haemodynamic responses to *angry*, *happy*, *fearful* speech did not differ significantly in amplitude from those evoked by *neutral* speech for any channel in fNIRS study by Steber et al. (2020). Further, Westgarth et al.'s (2021) systematic review of emotion perception, comprising mainly studies of the perception of visual emotions, supports this proposition.

Table 4.1. Cortical haemodynamic activity for individual vocal emotions relative to unemotional speech

			Functional neuroimaging technique	
Emotion	**Brain region**	**Side**	**fMRI**	**fNIRS**
Angry	**STC**	R	Ethofer, Kreifelts, et al., 2009 Frühholz & Grandjean, 2012 Grandjean et al., 2005 Korb et al., 2014 Wiethoff et al., 2008	-
		Bilateral	Mothes-Lasch et al., 2011 Quadflieg et al., 2008 Robins et al., 2009	-
	IFG	L	Vytal & Hamann, 2010	-
		R	Korb et al., 2014	-
		Bilateral	Ethofer, Kreifelts, et al., 2009 Quadflieg et al., 2008	Gruber et al., 2020
	MFG	L	Bach et al., 2008	Zhen et al., 2021*
		Bilateral	Kotz et al., 2013	Sonkaya & Bayazıt, 2018
Happy	**STC**	R	Buchanan et al., 2000 Leitman, 2010 Peelen et al., 2010 Vytal & Hamann, 2010 Wiethoff et al., 2008	-
	IFG	L	Frühholz, van der Zwaag, et al., 2016 Kotz et al., 2013	Sonkaya & Bayazıt, 2018 D. Zhang et al., 2018 Zhen et al., 2021*
Sad	**MFG**	L	Kotz et al., 2013 Vytal & Hamann, 2010	-
		R	Buchanan et al., 2000	Anuardi & Yamazaki, 2019

Notes. Summary of findings from studies that compare adult haemodynamic activity evoked by discrete vocal emotions (*angry*, *happy*, *sad*) with *unemotional* (or neutral) or aggregate of emotions as a proxy for speech. The focus is on cortical brain regions including the superior temporal cortex (STC), inferior frontal gyrus (IFG), middle frontal gyrus (MFG). Steber et al. (2020) found no significant activity for any emotion in any ROI and is thus not included. *Zhen et al. (2021) each emotion compared against each *happy* and *fearful* separately but is included for completeness.

Nonetheless, haemodynamic activity is consistently reported in the left IFG in response to *happy*, compared to *unemotional*, speech. Three fNIRS studies investigating the perception of vocal emotions in NH adult listeners report that *happy* speech evokes increased haemodynamic activity in a single channel, described as measuring from pars triangularis (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021). While it is plausible that these findings represent true *happy*-specific haemodynamic responses within a channel approximately covering the region of left IFG, attributing haemodynamic activity to pars triangularis presents a challenge; without structural brain images acquired using MRI or through digitisation of the fNIRS sensor positions on the scalp, variability in optode placement and individual differences in brain anatomy (Peelle, 2017; Westgarth et al., 2021) may reduce the reliability of channel-wise analyses relative to region-of-interest (ROI) analyses (Wiggins et al., 2016). Further, fNIRS is moderately susceptible to Type I errors, i.e., false positives stemming from the large number of comparisons required for channel-wise analyses, as well as its strong sensitivity to signals generated by physiological responses other than cortical haemodynamics (Barker et al., 2013; A. K. Singh & Dan, 2006; Tachtsidis & Scholkmann, 2016). While the three studies reporting haemodynamic activity in left IFG for *happy* speech (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021) use the false detection rate procedure (FDR; Benjamini & Hochberg, 1995) to correct for multiple comparisons, and algorithms to correct for motion artefacts, none report HbR or employ short-channel detectors to account for extraneous physiological signals. Substantial evidence of a cortical signature for *happy* continues to accrue in fNIRS studies of vocal emotions, but a robust assessment reporting both chromophores—one incorporating short-channel detectors—is required to confirm whether or not fNIRS provides for a sensitive measure of cortical signatures of discrete emotions.

Here, I assess whether fNIRS can be used to obtain unique cortical haemodynamic signatures for each *angry*, *happy*, and *sad* speech, with predicted cortical signatures based on brain regions most commonly reported for each given emotion (Table 4.1). I employ an ROI analysis to assess the involvement of bilateral STG, IFG, and MFG using both chromophores (HbO and HbR), with an increased number of channels relative to previous studies, and with the necessary regression of extracerebral and other signal components using short channels, to search for cortical signatures of discrete emotions. If fNIRS is sensitive to cortical haemodynamic signatures of discrete emotions, comparisons with *unemotional* will yield increased haemodynamic activity in right STG for *angry*, right STG and left IFG for *happy*, and bilateral MFG for *sad*. More generally, bilateral STG will exhibit increased haemodynamic activity evoked by speech, as compared to silence, and vocal emotions will evoke greater haemodynamic activity in right, relative to left, STG.

4.3 Method

4.3.1 Participants

27 participants (15 female, 12 male, mean age 26.4 years, SD=4.9 years) were recruited from Macquarie University and via the Macquarie University Cognitive Science Register. Of these, 21 (10 female, 11 male, mean age 27.4 years, SD=4.7 years) met the inclusion criteria. To be included, participants were required to be right-handed native speakers of English, aged 18–36 years, with no known psychological or neurological disorders, and with pure-tone audiometric thresholds within normal limits. These were defined as air-conduction thresholds below 20 dB HL for all octave frequencies between 250 and 8000 Hz. As a criterion of fNIRS signal quality, the signal calibration metric was used to ensure that no more than 2 channels in any ROI were of 'critical' quality, as described in Chapter 2.2.3. Exclusions based on this criterion resulted from extremely thick, coarse, black hair, which impeded contact between optodes and scalp. Of the six excluded participants, two were left-handed, one had elevated hearing thresholds, two had more than 2 'critical' channels in a single ROI, and one did not complete the experiment. Ethical approval was requested and granted by Macquarie University's ethics committee for this study (Reference number: 5201952978351). All participants gave informed consent before participation and received an honorarium or course credit for their involvement.

4.3.2 Stimuli

Stimuli consisted of ~7.25s blocks of *angry*, *happy*, *sad*, and *unemotional* conveyed in natural speech, and a *control* silence (also 7.25s), generated following the procedure described in Chapter 2.1. Five additional blocks of speech with a 500-ms 400-Hz tone (less intense than the speech) overlapping at a random point in the block (n=5, one *happy*, one *sad*, one *angry*, two *unemotional*) were used to ensure participants were attentive. Sound intensity was normalised across all stimuli. See Chapter 2.1 for further detail on stimuli.

4.3.3 Test procedure

Each experiment began with four familiarisation trials to familiarise participants with the pseudo-sentences and the button-press task: One block of each *angry*, *happy*, *sad*, and *unemotional* speech was presented (n=4), whereby two blocks contained the pure tone. Subsequently, the main experiment commenced: blocks of each *angry*, *happy*, *sad*, *unemotional* speech and *control* (silence) blocks were each presented 20 times in a randomised order (n=100). The five stimuli containing the pure tone were pseudo-randomly interspersed between these (n=10). In total, the experiment consisted of 114 trials

separated by inter-stimulus intervals of random lengths in the range 13–23 s (Figure 4.1). The total duration of the recording was 50 minutes.

Participants were instructed that they would hear both speech and silence, and that their task was to sit still throughout the experiment and listen to the speech and silence. They were also informed that they occasionally would hear a tone overlapping with the speech and that they should indicate they heard it by pressing the top left button on an RB-840 button box (Cedrus Corporation, San Pedro, USA) with their left index finger. See Chapter 2.2.3 for further detail on the experimental protocol.

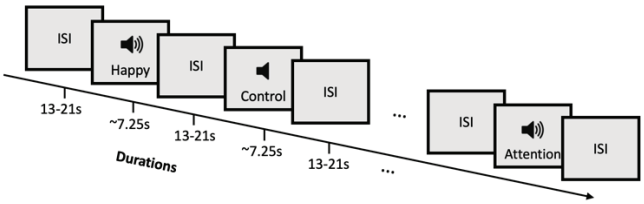


Figure 4.1. Schema of listening experiment. ISI=Inter-stimulus-interval, loud-speaker symbol=listening trial.

4.3.4 Equipment

fNIRS and stimulus presentation equipment described in detail in Chapter 2.2.2.

4.3.5 Data Analysis

Trials for each speech condition (*angry*, *happy*, *sad*, *unemotional*) and *control* were analysed. Trials excluded from the analysis included the familiarisation trials (n=4 per participant), attention trials (n=10 per participant), trials in which a button press was made whether intentionally (n=10 per participant) or mistakenly (n=5, in 4 participants). As a measure of signal quality, a scalp-coupling index, the correlation of the heart rate signal (~1 Hz) in the HbO and HbR signals for each channel (Pollonini et al., 2014), was calculated for frequencies between 0.7–1.35 Hz. 93% of channels had a scalp-coupling index >0.8, indicative of good contact between optodes and the scalp (Figure 4.2).

The generation of grand average waveforms, estimates of haemodynamic response amplitude per condition, ROI, and participant from the first-level analysis, as well as the model building procedures for the second-level (group-level) analysis and hypothesis-testing contrasts, were performed as described in Chapter 2.2.4.

Group-level haemodynamic response amplitudes per condition per ROI. To predict the amplitude of measured haemodynamic responses in response per condition per ROI, linear mixed-effects (LME) models were fit to estimates extracted from the first-level

analysis (N=660 for each HbO and HbR). Models for each HbO and HbR were built in a step-wise fashion, beginning with random intercepts to account for variability between participants. Next, random coefficients (slopes of a categorical variable) of *Condition* per *Participant* were added to account for individual differences between the five levels of *Condition* within each participant. Finally, each *ROI*, *Condition*, and the interaction between *ROI* and *Condition* were added as fixed effects. For additional details on the specification of these models, see Chapter 2.2.4.

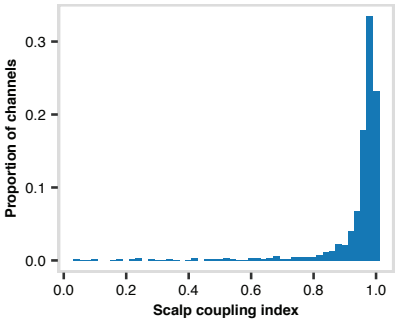


Figure 4.2. Signal quality as measured by scalp-coupling index per channel. Histogram showing the distribution of scalp-coupling indices, calculated per channel, as a proportion of the total number of channels (N=2856).

4.4 Results

4.4.1 fNIRS waveforms reveal speech-evoked haemodynamic responses

The primary goal of this study was to quantify cortical haemodynamic activity in STG, IFG and MFG evoked by individual emotions. Visual inspection of the grand average waveforms per condition per ROI (Figure 4.3) demonstrates that for the *control* condition, both chromophores remained very close to baseline. In response to all presentations of speech, regardless of emotion, STG bilaterally exhibited a concurrent increase in the concentration of HbO and decrease in the concentration of HbR. In both left and right STG, the negatively correlated waveforms for HbO and HbR matched the expected morphology of the haemodynamic response, in that they are both furthest from baseline (i.e., peak for HbO and minimum for HbR) approximately 6 seconds following the onset of the speech stimulus (Scholkmann, Kleiser, et al., 2014). In both left and right IFG, all speech conditions show a trend of reduced concentrations of HbO with concurrently elevated concentrations of HbR. *Happy* speech evoked the largest negative haemodynamic response bilaterally in IFG.

Bilaterally for all speech conditions, MFG showed an increase in HbO (peak at ~3 s

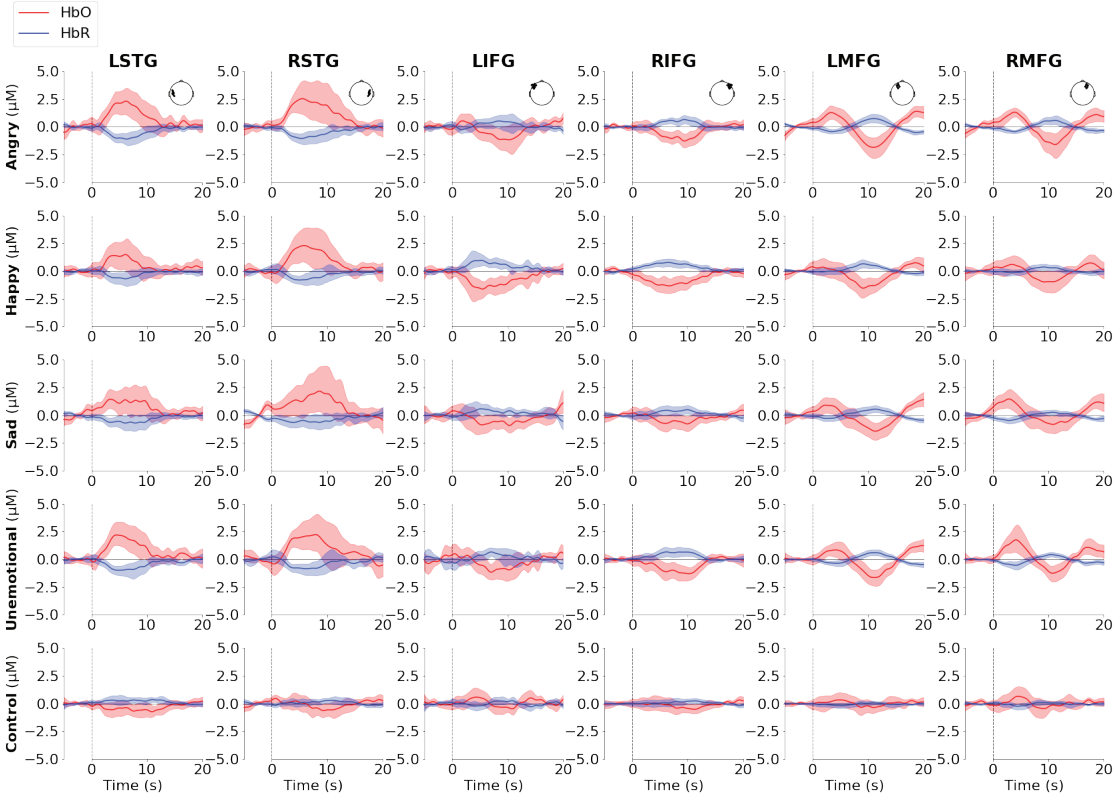


Figure 4.3. Grand average waveforms. Grand average waveforms for each condition (rows) and ROI (columns). The position of the optodes in each ROI is represented in the inset head shape (viewed from above). LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus, LMFG=left middle frontal gyrus, RMFG=right middle frontal gyrus. Solid lines indicate the mean and the shaded area indicates the 95% confidence interval.

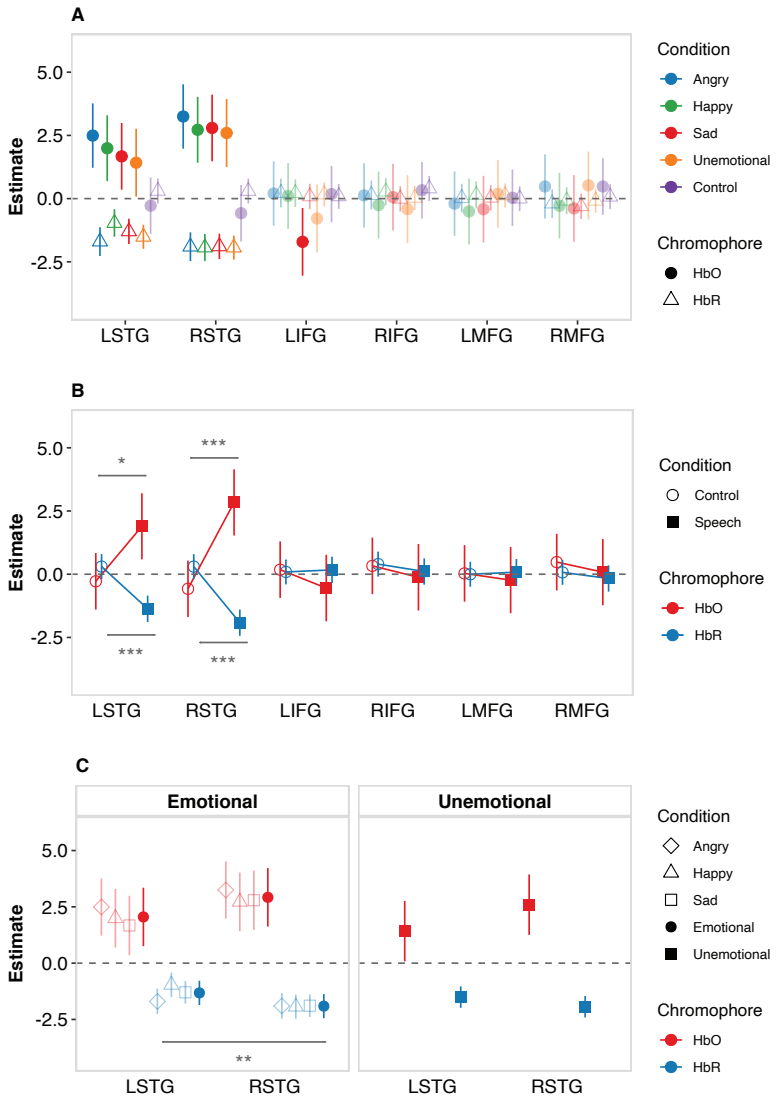


Figure 4.4. Group level response estimates and contrasts. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus, LMFG=left middle frontal gyrus, RMFG=right middle frontal gyrus. A) Group-level response estimates for each ROI, condition and chromophore. Darker symbols indicate a significant difference from zero and transparent symbols indicate a non-significant difference from zero. B) Contrasts between speech and *control* conditions for each *ROI*. C) Contrasts between left and right STG for each emotional and *unemotional* speech; individual emotions included as transparent symbols for reference. Error bars indicate 95% confidence intervals, averaged across aggregated conditions where appropriate in B and C. Significance for contrasts in B and C, *p<0.05, **p<0.01, ***p<0.001.

Table 4.2. Group-level estimates of haemodynamic response amplitude

		HbO				HbR			
	ROI	**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
Angry	LSTG	**2.50**	**0.64**	**3.90**	**<0.001**	**-1.70**	**0.28**	**-5.98**	**<0.001**
	RSTG	**3.25**	**0.64**	**5.07**	**<0.001**	**-1.90**	**0.28**	**-6.68**	**<0.001**
	LIFG	0.20	0.64	0.32	0.753	0.22	0.28	0.78	0.439
	RIFG	0.13	0.64	0.20	0.844	0.14	0.28	0.49	0.628
	LMFG	-0.19	0.64	-0.31	0.758	-0.02	0.28	-0.07	0.947
	RMFG	0.48	0.64	0.75	0.456	-0.20	0.28	-0.69	0.492
Happy	LSTG	**1.99**	**0.66**	**3.05**	**0.003**	**-0.97**	**0.27**	**-3.52**	**0.001**
	RSTG	**2.72**	**0.65**	**4.16**	**<0.001**	**-1.94**	**0.27**	**-7.18**	**<0.001**
	LIFG	0.10	0.65	0.16	0.875	0.22	0.27	0.78	0.435
	RIFG	-0.25	0.67	-0.38	0.703	0.28	0.27	1.02	0.310
	LMFG	-0.51	0.65	-0.77	0.441	0.14	0.27	0.54	0.593
	RMFG	-0.28	0.65	-0.43	0.668	-0.08	0.27	-0.31	0.756
Sad	LSTG	**1.67**	**0.66**	**2.53**	**0.014**	**-1.30**	**0.25**	**-5.13**	**<0.001**
	RSTG	**2.80**	**0.66**	**4.22**	**<0.001**	**-1.89**	**0.25**	**-7.48**	**<0.001**
	LIFG	**-1.71**	**0.67**	**-2.54**	**0.013**	0.09	0.25	0.34	0.736
	RIFG	0.06	0.66	0.09	0.931	-0.01	0.25	-0.03	0.975
	LMFG	-0.42	0.66	-0.63	0.529	0.03	0.25	0.13	0.894
	RMFG	-0.39	0.66	-0.59	0.559	-0.30	0.25	-1.20	0.233
Unemotional	LSTG	**1.42**	**0.67**	**2.11**	**0.038**	**-1.51**	**0.24**	**-6.28**	**<0.001**
	RSTG	**2.60**	**0.67**	**3.85**	**<0.001**	**-1.94**	**0.24**	**-8.05**	**<0.001**
	LIFG	-0.79	0.67	-1.17	0.247	0.16	0.24	0.68	0.499
	RIFG	-0.41	0.67	-0.62	0.540	0.03	0.24	0.11	0.914
	LMFG	0.19	0.67	0.28	0.779	0.13	0.24	0.52	0.603
	RMFG	0.52	0.67	0.78	0.439	-0.09	0.24	-0.36	0.717
Control	LSTG	-0.28	0.57	-0.50	0.622	0.30	0.25	1.21	0.228
	RSTG	-0.58	0.57	-1.02	0.310	0.30	0.25	1.21	0.230
	LIFG	0.18	0.57	0.32	0.752	0.09	0.25	0.36	0.720
	RIFG	0.33	0.57	0.58	0.561	0.40	0.25	1.60	0.111
	LMFG	0.03	0.57	0.06	0.954	0.00	0.25	-0.01	0.996
	RMFG	0.48	0.57	0.85	0.398	0.07	0.25	0.30	0.767

Notes. Significance of contrasts, bold font indicates $p<0.05$. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus, LMFG=left middle frontal gyrus, RMFG=right middle frontal gyrus.

Table 4.3. Planned contrasts with group-level estimates of haemodynamic response amplitude

		HbO				HbR			
Contrast		**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
Speech vs. Control	LSTG	**2.17**	**0.70**	**3.13**	**0.030**	**-1.67**	**0.29**	**-5.78**	**<0.001**
	RSTG	**3.42**	**0.70**	**4.92**	**<0.001**	**-2.22**	**0.29**	**-7.68**	**<0.001**
	LIFG	-0.73	0.70	-1.05	0.848	0.08	0.29	0.28	0.995
	RIFG	-0.45	0.70	-0.65	0.848	-0.29	0.29	-1.01	0.995
	LMFG	-0.41	0.70	-0.59	0.848	-0.01	0.29	-0.03	0.995
	RMFG	-0.25	0.70	-0.36	0.928	-0.57	0.29	-1.97	0.338
Angry vs. Unemotional	LSTG	1.07	0.91	1.18	0.848	-0.19	0.34	-0.56	0.995
	RSTG	0.65	0.91	0.72	0.848	-0.04	0.34	0.11	0.995
	LIFG	0.99	0.91	1.09	0.848	0.06	0.34	0.17	0.995
	RIFG	0.54	0.91	0.60	0.848	0.11	0.34	0.33	0.995
	LMFG	0.59	0.91	0.65	0.848	-0.14	0.34	-0.42	0.995
	RMFG	0.89	0.91	0.98	0.848	-0.22	0.34	-0.65	0.995
Happy vs. Unemotional	LSTG	0.57	0.92	0.62	0.848	0.54	0.35	1.53	0.658
	RSTG	0.13	0.92	0.14	0.928	0.00	0.35	0.01	0.995
	LIFG	0.89	0.92	0.17	0.848	0.05	0.35	0.15	0.995
	RIFG	0.16	0.93	0.17	0.928	0.25	0.35	0.71	0.995
	LMFG	0.28	0.92	0.30	0.928	-0.02	0.35	-0.05	0.995
	RMFG	0.13	0.92	0.14	0.928	-0.11	0.35	-0.31	0.995
Sad vs. Unemotional	LSTG	0.25	0.75	0.34	0.928	0.21	0.32	0.66	0.995
	RSTG	0.20	0.75	0.27	0.928	0.05	0.32	0.15	0.995
	LIFG	-0.92	0.76	-1.22	0.848	-0.08	0.32	-0.24	0.995
	RIFG	0.47	0.75	0.63	0.848	-0.03	0.32	-0.10	0.995
	LMFG	0.37	0.75	0.49	0.902	-0.13	0.32	-0.40	0.995
	RMFG	0.03	0.75	0.04	0.972	-0.33	0.32	-1.01	0.995
RSTG vs. LSTG	Emo.	0.87	0.42	2.08	0.334	**-0.59**	**0.18**	**-3.33**	**0.008**
	Unemo.	1.18	0.73	1.62	0.688	-0.43	0.30	-1.41	0.700

Notes. Significance of contrasts, bold font indicates $p<0.05$. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus, LMFG=left middle frontal gyrus, RMFG=right middle frontal gyrus. Emo=Emotional, Unemo=Unemotional.

post-stimulus-onset) followed by a reduction (minimum at ~10 s post-stimulus-onset), then stabilising slightly above baseline after ~18 s. Deflections in HbR followed the same time course as HbO, but are of inverse polarity and reduced amplitude, and were observed to be introduced by the signal-to-noise improvement algorithm that enhances the negative correlation between chromophores (Cui et al., 2010). In the *control* condition, bilateral MFG showed the same pattern of deflections in both chromophores with greatly reduced amplitudes. Short-channels positioned over the MFG also revealed the same pattern for speech conditions in HbO, although with deflections of greater amplitudes (i.e., peak and minimum of ~5 μM from baseline). The similarity between the short-channel waveforms and MFG waveforms after short-channel regression suggests the regression of short-channels may not account for the entirety of the extracerebral signals in the long-channel data. Thus, the deflections observed in bilateral MFG likely originate from extracerebral sources and are systemic in nature. Extracerebral signal components tend to influence the HbO signal more strongly than HbR (Kirilina et al., 2012), as observed in the MFG. The similarity in the pattern of deflections in the *control* and speech conditions, although with greatly reduced amplitudes in the *control* condition, indicates that the speech stimuli did not evoke the deflections, but rather enhanced them, perhaps reflecting the autonomic nervous system's response to the listening task (Kirilina et al., 2012; Tachtsidis & Scholkmann, 2016).

4.4.2 Haemodynamic response amplitude per ROI per condition

To predict the presence and amplitude of haemodynamic responses evoked by vocal emotions, the following model was fit to each the HbR and HbO data: $\beta + -1 +$ *ROI* $+$ *Condition* $+$ *ROI*:*Condition* $+ (1 +$ *Condition*$|$*Participant*$)$. The model accounted for more of the variance in the HbR data ($R^2_{m/c}$=0.31/0.53) than the HbO data ($R^2_{m/c}$=0.15/0.45). For both HbO and HbR, the inclusion of random intercepts per *Participant* and random coefficients of *Condition* per *Participant* (HbO: $\chi^2(14)=108.96$, $p<0.001$, HbR: $\chi^2(14)=63.42$, $p<0.001$) accounted for a significant amount of variance. The main effects of *ROI* (HbO: $\chi^2(6)=72.51$, $p<0.001$, HbR: $\chi^2(6)=173.33$, $p<0.001$) and *Condition* (HbO: $\chi^2(4)=9.51$, $p=0.050$, HbR: $\chi^2(4)=40.08$, $p<0.001$) accounted for a significant proportion of the variance, as did the interaction between *ROI* and *Condition* (HbO: $\chi^2(20)=43.12$, $p=0.002$, HbR: $\chi^2(20)=71.54$, $p<0.001$).

In both left and right STG, for all four speech conditions (*angry*, *happy*, *sad*, *unemotional*), concentrations of HbO were significantly higher than zero ($p<0.05$) and concentrations of HbR were significantly lower than zero ($p<0.05$; Figure 4.4A, Table 4.2). In left IFG, *sad* evoked a concentration of HbO ($p=0.013$) significantly lower than zero. No other significant changes in the concentration of either chromophore were observed in any other ROI for any speech condition. The *control* condition did not differ significantly from zero in any ROI or either chromophore (Figure 4.4A, Table 4.2).

4.4.3 Speech evokes bilateral STG activation

To address the hypothesis that speech evokes haemodynamic activity bilaterally in STG, contrasts between speech (*angry*, *happy*, *sad*, *unemotional* aggregated) and *control* conditions were employed. The contrasts demonstrate that speech stimuli evoke a higher concentration of HbO in left (p=0.030) and right (p<0.001) STG, relative to the silent *control* condition (Figure 4.4B, Table 4.3). Congruently, concentrations of HbR were significantly lower for speech compared to the *control* condition in left and right STG (p<0.001), confirming that speech evokes haemodynamic activity bilaterally in STG.

4.4.4 Vocal emotions evoke significant right-lateralised HbR responses in STG

The presence of hemispheric lateralisation in STG was investigated by contrasting estimates of response amplitude in left and right STG for emotional (*angry*, *happy*, *sad* speech together) and for *unemotional* speech. For HbO, response estimates for emotional and *unemotional* speech conditions were numerically higher in right, compared to left, STG, but these differences were not statistically significant (p>0.05; Figure 4.4C, Table 4.3). For HbR, a right-lateralisation of the haemodynamic activity was observed for emotional speech with reduced HbR in right, compared to left, STG (p=0.008), but not for *unemotional* speech (p=0.700). Thus the hypothesis that emotional speech evokes haemodynamic activity predominantly in right, compared to left, STG was only partially supported. The trend for larger haemodynamic response amplitudes in right than left STG for *unemotional* speech is unexpected, as activity evoked by non-emotional speech is typically associated with a left-hemisphere bias in STG.

4.4.5 fNIRS insensitive to discrete emotions

To determine whether fNIRS is sensitive to discrete emotions and therefore might yield independent cortical signatures for each emotion, contrasts were made between each of *angry* vs. *unemotional*, *happy* vs. *unemotional*, and *sad* vs. *unemotional*, speech. No significant differences were observed in any ROI for either chromophore for *angry*, *happy*, or *sad* relative to *unemotional*, supporting the null hypothesis that fNIRS is not sensitive to emotion-specific haemodynamic activity (p>0.05; Table 4.3).

4.5 Discussion

This study investigated whether fNIRS imaging could be used to obtain unique neural signatures for *angry*, *happy*, and *sad* speech. Using ROI analyses, the data demonstrate that fNIRS is sensitive to speech-evoked activation of the STG bilaterally, and the right-lateralisation of STG for emotional speech, but it is not sensitive to discrete cortical representations of emotions conveyed in speech.

4.5.1 Haemodynamic activity evoked by speech and emotional speech

Speech-evoked activation of bilateral STG, evidenced by elevated haemodynamic responses relative to silence for both chromophores, is consistent with previous studies employing fNIRS (Mushtaq et al., 2019; Sevy et al., 2010; Wiggins et al., 2016) and fMRI (e.g., Belin et al., 2002; Fecteau et al., 2004; Okada et al., 2010). A right-lateralisation for STG involvement evoked by emotional speech was numerically evident for both emotional and *unemotional* speech in both chromophores but was only significant for emotional speech in HbR. This right-lateralisation for emotional speech is consistent with previous data obtained using fMRI (Beaucousin et al., 2007; Kotz et al., 2006; Kreitewolf et al., 2014; Seydell-Greenwald et al., 2020; von Cramon et al., 2003; Witteman et al., 2012). Right-lateralisation of STG activity to vocal emotions is also consistent with fNIRS investigations of vocal emotions (Table 4.1), of which the majority report only concentrations of the HbO chromophore (Sonkaya & Bayazıt, 2018; Zhang et al., 2017, 2018; Zhen et al., 2021). The speech stimuli employed in the present study are most similar to those of Zhang et al. (2018), which employed six-syllable pseudo-sentences in four emotions, although the present study replaces *fearful* with *sad*, and doubles the number of trials presented per emotion. It seems unlikely that these differences contribute to the right-lateralisation of STG activity for HbO not reaching significance. The greater variability observed in HbO relative to HbR is a more probable cause for the non-significance of the apparent right-lateralisation evoked by vocal emotions in STG for HbO (Figure 4.4C). Increased variability in the amplitude of HbO relative to HbR haemodynamic responses is a common phenomenon (e.g., Defenderfer et al., 2017; Mushtaq et al., 2019; Sato et al., 1999; Wiggins et al., 2016), which in practice, means the differences in amplitudes for HbO estimates need to be proportionately greater than the differences in amplitudes required for HbR to reach significance.

fMRI studies associate language processing with enhanced left relative to right STG activity (Dick et al., 2007; Frost, 1999; Pujol et al., 1999), as do fNIRS studies investigating cortical activations evoked by speech (Mushtaq et al., 2019; Pollonini et al., 2014; Sevy et al., 2010). In the present study, a trend to right-lateralisation of STG activity evoked by *unemotional* speech was observed, which does not align with the expected dominance of left STG in the processing of speech. This may be explained by the increased specificity, i.e., the estimated proportion of emitted photons that pass through a brain area, of the channels covering right, relative to left, STG (Chapter 2.2.2, Table 2.4). In channels with higher specificity, cortical signal components make up a larger portion of the signal, resulting in a higher signal-to-noise ratio, than in channels with lower specificity values (Zimeo Morais et al., 2018). Right STG channels with higher specificity values may yield larger haemodynamic responses than left STG channels

with lower specificity values, potentially explaining the trend to right-lateralisation of the haemodynamic responses in STG evoked by *unemotional* speech.

The measured cortical activity for *angry*, *happy*, and *sad* vocal emotions did not differ significantly from *unemotional* speech in any ROI, indicating that fNIRS, using an ROI-based approach, is insensitive to discrete emotions. This study did not replicate any of the emotion-specific findings described for either fMRI or fNIRS studies referenced in Table 4.1. The inability to replicate the—highly variable—fMRI findings may be explained by fNIRS reduced spatial resolution and depth penetration compared to fMRI (Pinti et al., 2020). Of the existing fNIRS studies, which almost exclusively report channel-wise haemodynamic activity, the present study is most consistent with Steber et al. (2020), who also did not observe any significant haemodynamic activity for discrete emotions conveyed in speech. In contrast to the other fNIRS studies that report significant activity in variable channels for discrete emotions (Table 4.1, Anuardi & Yamazaki, 2019; Gruber et al., 2020; Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021), the present study employed ROI analysis, which reduces fNIRS' spatial resolution, but yields more reliable analyses with a reduced risk of Type I errors (Wiggins et al., 2016).

4.5.2 fNIRS cannot distinguish cortical activity evoked by discrete emotions

Three previous fNIRS studies employing channel-wise analyses reveal *happy*, relative to unemotional, speech evokes elevated haemodynamic activity for HbO in a single channel over left IFG (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2018; Zhen et al., 2021), measuring between F5 and FC5 according to the International 10/10 system (Chatrian et al., 1985). The described channel is included in the left IFG ROI examined in the present study. Visual inspection of the estimates of haemodynamic response amplitude for *happy* in the FC-FC5 channel for each participant (extracted from the first-level GLM analysis) indicated that only one participant showed an HbO estimate significantly higher than zero. Exploratory models (LMEs) fit to these estimates revealed no significant group-level haemodynamic activity evoked by *happy* speech in the FC-FC5 channel (HbO: β=-0.26, SE=1.23, t=-0.21, p=0.838; HbR: β=0.79, SE=0.6, t=1.33, p=0.184), confirming that the present study does not replicate the channel-specific left IFG activity previously reported for *happy* speech.

Angry speech is reported to evoke increased haemodynamic activity in a single channel between AF7 and FP1 according to the International 10/10 system, described by the authors as measuring from left MFG (Zhen et al., 2021). According to the fOLD toolbox and the AAL2 atlas used to define the montage used in the present study (Rolls et al., 2015; Tzourio-Mazoyer et al., 2002; Zimeo Morais et al., 2018), this channel can indeed be classified as corresponding to left MFG. This channel was not included in the

montage used in the present study, because the specificity (i.e., the estimated photon sensitivity per channel) as calculated by the fOLD toolbox for this channel is substantially lower (47%) than the minimum specificity of the other included MFG channels (all>60%). This means that the channels included in the present study are more likely to detect haemodynamic activity in MFG evoked by *angry* speech than the single AF7-FP1 channel reported by Zhen et al. (2021). The increased channel specificity and more reliable ROI analysis approach (Wiggins et al., 2016) used in the present study possibly explain why the present study did not replicate Zhen et al.'s (2021) finding that *angry* speech is processed in left MFG.

In the present study, *sad* speech did not evoke a unique pattern of haemodynamic activity relative to *unemotional* speech. To date, the only report of a unique cortical representation of *sad* speech reported using fNIRS is in right MFG (Anuardi & Yamazaki, 2019), although the exact location of the channel on the scalp for which the response is reported is unclear. The majority of fNIRS studies do not investigate cortical activity evoked by *sad* speech, including *fearful* speech instead (Gruber et al., 2020; Sonkaya & Bayazıt, 2018; Steber et al., 2020; D. Zhang et al., 2018; Zhen et al., 2021). The decision to include *sad* rather than *fearful* speech in the present study was based on the prevalence of behavioural studies consistently reporting higher accuracy for *sad* than *fearful* speech (Metcalfe, 2017; Paulmann et al., 2008; Sauter et al., 2010; Scherer et al., 1991). As such, the null result obtained for *sad* speech in the present study contributes valuable information to the field of vocal emotions as measured with fNIRS.

Differences in haemodynamic response amplitude were not compared between discrete emotions (i.e., *angry* and *happy*) in the present study and would be unlikely to be significant, based on visual inspection of Figure 4.4A. Despite this, the linear reduction in response amplitudes of HbO from *angry* to *happy* and then to *sad* to *unemotional* in left STG suggests some discrete processing might occur. The HbR response amplitude for angry in left STG was lowest, with a linear reduction from *happy* to *sad* to *unemotional*. In right STG, only *angry* showed an elevated response amplitude relative to the other emotions for HbO, and for HbR, response amplitudes of all speech conditions were similar. In bilateral STG for HbO and left STG for HbR, *angry* speech evoked the largest response, consistent with evidence from fMRI that *angry* speech may enhance activity in STG (Frühholz & Ceravolo, 2018). A difference in the arousal, the stimulus-evoked autonomic activation, between *angry* speech and speech conveying other emotions may account for the enhanced response, as vocal emotions with higher arousal are associated with larger haemodynamic response amplitudes bilaterally in STG (Bestelmeyer et al., 2017; Ethofer et al., 2012; Wiethoff et al., 2008). Moreover, a gradual reduction in arousal from *angry* to *happy* to *sad* may explain the gradient in HbR response amplitudes observed in left STG. Alternatively, the greater activation for *angry* speech

observed bilaterally in STG may reflect a difference in perceived loudness (Weder et al., 2020), despite all speech stimuli having been presented at equal intensity. Future fNIRS studies on vocal emotions should account for arousal and perceived loudness.

4.5.3 Comparing waveforms and estimates of haemodynamic response amplitudes

The grand average waveforms of the haemodynamic responses evoked in bilateral STG (Figure 4.3) by vocal emotions match the expected shape of the canonical haemodynamic response function (e.g., Glover, 1999), and their peak amplitudes are consistent with response estimates for left and right STG obtained using the GLM (second-level) analysis (Figure 4.4A). In the IFG bilaterally, the waveforms show concentrations of HbO below zero and HbR above zero, with the closest resemblance to a negative haemodynamic response observed for *happy* speech (Figure 4.3). The second-level analysis revealed a significant negative estimate of haemodynamic response amplitude for HbO for *sad* speech only (Figure 4.4A).

This difference between the grand average waveforms for left IFG and the group-level estimates is likely attributable to the pre-processing steps employed to generate the waveforms and/or the statistical methods applied to estimate amplitudes in each approach. The pre-processing steps to generate waveforms include modifying the recorded signal using algorithms to correct for motion artefacts (Fishburn et al., 2019) and improve the signal-noise-ratio (Cui et al., 2010). Cui et al.'s (2010) algorithm, in particular, enhances the negative correlation between chromophores, and when applied, the negative HbO deflection in left IFG is greater in amplitude and smoother for all speech conditions. Fishburn et al.'s (2019) algorithm corrects spikes and baseline shifts in the signal, and when applied, the negative HbO deflection in left IFG is shallower for all speech conditions. When neither algorithm is applied, *sad* and *unemotional* speech show the largest negative deflections of HbO in left IFG, consistent with the estimates from the GLM approach (Figure 4.4A). The GLM approach does not employ any pre-processing steps, but rather accounts for physiological noise and motion artefacts on the basis of their correlation, using an autoregressive model. As the two approaches employ different mechanisms to isolate the neural signal component from the noise, further investigation would be required to pinpoint the exact origin of the difference observed between the waveforms and second-level estimates of haemodynamic response amplitude in left IFG.

Moreover, the waveform and GLM approaches also employ different statistical methods. Grand average waveforms are obtained by simple averaging of epoched time courses, whereas the GLM approach considers the shape of the time course when calculating the first-level estimates of response amplitude (Huppert, 2016), meaning that

if the time course does not match the haemodynamic response function (e.g., Glover, 1999), the GLM will estimate an amplitude close to zero. Furthermore, in the GLM approach, the second-level LME analysis accounts for differences between participants' response amplitudes overall with random intercepts per participant and for differences between participants' response amplitudes to the investigated conditions with random coefficients (slopes) for each condition per participant. As described, the two approaches are distinct from each other, meaning that while they are complementary, and in the ideal case, the estimates of response amplitude are consistent between the two, it is also possible for certain differences to be observed.

4.5.4 Strengths and limitations

The present study's central strength is the robust experimental design. First, the present study records from the highest, most densely arranged, number of channels used in an fNIRS study of vocal emotions to date, whereby the reliability of the analyses was increased by averaging across channels within an ROI (Wiggins et al., 2016). Second, short-channel detectors were incorporated into the montage to detect and regress out extracerebral changes in oxygenation, reducing the contamination of cerebral signal components by extracerebral components (Brigadoi & Cooper, 2015).

Recent investigations of the influence of ROI size on observations of haemodynamic activity indicate that averaging across large ROIs may obscure more localised haemodynamic activity (Powell et al., 2018; Shader et al., 2021). The present study selected fairly broad ROIs. While this is not assumed to be a severe limitation of the present study, future research should investigate haemodynamic activity evoked by vocal emotions with smaller ROIs, plausibly including smaller ROIs within the ROIs used in the present study. Furthermore, behavioural measures of emotion recognition and participants' emotional state before beginning the experiment may provide valuable insight into individual variability and should be included in future studies.

4.6 Conclusions

The present study confirms that fNIRS is not sensitive to discrete vocal emotions. Evidence was found for fNIRS' sensitivity to speech-evoked haemodynamic activity in bilateral STG. Furthermore, a right-lateralisation of the haemodynamic activity evoked by emotional speech was observed in STG. In light of these findings, fNIRS is not the optimal tool for studying the neural representation of discrete vocal emotions but still holds promise as a valuable technique for studying other aspects of vocal emotion, and more broadly, speech perception in a variety of listeners.

Chapter 5| Deciphering emotion: Cortical responses to vocal emotions with attenuated voice pitch variations as measured by functional near-infrared spectroscopy

5.1 Abstract

Background. Normal-hearing (NH) listeners rely most strongly on the variations of fundamental frequency (F0; the acoustic correlate of voice pitch) to identify emotions conveyed in speech. Without informative F0 cues, listeners' ability to extract emotional meaning from speech is reduced, although substantial variability in performance is observed between listeners. An objective measure of listeners' ability to recognise emotions in speech, in this case, changes in cortical blood flow evoked by cortical neuronal activity measured with functional near-infrared spectroscopy (fNIRS), can help determine the neural underpinnings of successful and poor recognition of vocal emotions resulting from uninformative F0 cues.

Aims. fNIRS was used to investigate cortical haemodynamic responses to vocal emotions conveyed in natural speech and speech with uninformative F0 cues, with a particular focus on the relationship between haemodynamic response amplitude evoked by vocal emotions and behavioural accuracy in recognising vocal emotions with uninformative F0 cues.

Method. NH listeners (N=21) completed a block-design listening task, in which they listened to ~7.25-s blocks of *happy* and *sad* speech, in each natural speech and speech with variations in F0 attenuated (rendering F0 cues uninformative), as well as 7.25 s of silence as a control condition. Regions of interest (ROI) included bilateral superior temporal gyri (STG) and inferior frontal gyri (IFG). To obtain behavioural accuracy scores, listeners adjudged *angry*, *happy*, *sad* and *unemotional* sentences (each in natural speech and speech with variations in F0 attenuated) in a 4-alternative forced-choice emotion recognition task, repeated before and after the fNIRS listening task.

Results. Significant haemodynamic activity in right STG evoked by vocal emotions (vocal emotions in natural speech and speech with attenuated F0 variations aggregated, compared to silence), as well as a significant right-lateralisation of STG activity evoked by vocal emotions in each natural speech and speech with F0 variations attenuated. Response amplitudes did not differ significantly between vocal emotions in natural speech and speech with variations in F0 attenuated in any ROI. Behavioural accuracy in recognising vocal emotions with attenuated variations in F0 before the fNIRS recording was significantly associated with haemodynamic response amplitude in right STG;

decreasing accuracy was associated with increasing response amplitude.

Conclusions. The data confirm the involvement of right STG in the processing of vocal emotions, with and without informative F0 cues. Moreover, they demonstrate that the amplitude of haemodynamic responses in right STG reflects listeners' abilities to recognise emotions conveyed in speech with uninformative F0 cues.

5.2 Introduction

Healthy, normal-hearing (NH) listeners are highly skilled at recognising emotions in speech (e.g., Bezooijen, 1984; Cannon & Chatterjee, 2018; Chatterjee et al., 2015; Christensen et al., 2019; Demenescu et al., 2014; House, 1994; Jiang et al., 2015; Laukka & Laukka, 2004; Luo et al., 2007; Metcalfe, 2017; Tinnemore et al., 2018)—as long as the natural variation in fundamental frequency (F0) is intact. F0, the rate of vibration of the vocal folds during speech production, is the acoustic correlate of voice pitch. NH listeners rely heavily on the F0 contour, i.e., the variation in F0 over time, to recognise vocal emotions (Chapter 3, Banse & Scherer, 1996; Globerson et al., 2013; Hammerschmidt & Jürgens, 2007; Metcalfe, 2017; Mozziconacci, 1998; Patel et al., 2011; Pell, 1998; Rodero, 2011; Scherer et al., 2003).

Accuracy with which vocal emotions are recognised decreases substantially when cues in F0, the primary acoustic feature conveying emotional information in speech, are reduced (e.g., Chapter 3, Everhardt et al., 2020; Jiam et al., 2017; Leitman, 2010). Accordingly, cochlear implant (CI) users, whose CI devices transmit spectrotemporally degraded speech containing reduced F0 cues (Başkent et al., 2016; Wilson & Dorman, 2008) recognise vocal emotions less accurately than NH listeners (Chatterjee et al., 2015; Everhardt et al., 2020; Gilbers et al., 2015; Jiam et al., 2017; Luo et al., 2007; Most & Aviner, 2009; Nakata et al., 2012; Pak & Katz, 2019; Panzeri et al., 2021; Pereira, 2000; Peters, 2006; Ren et al., 2021; Waaramaa et al., 2018). A similar deficit in recognition of vocal emotions can be elicited in NH listeners when variations in F0 are attenuated through noise-vocoded CI simulations (Chatterjee et al., 2015; Gilbers et al., 2015; Luo et al., 2007; Metcalfe, 2017; Pak & Katz, 2019; Ritter & Vongpaisal, 2018), speech synthesis (Johnson et al., 1986; Laukka & Juslin, 2007; Robinson et al., 2019) or targeted reductions of variations in F0 (Chapter 3; Leitman et al., 2010; Metcalfe, 2017). Moreover, the extent to which accuracy in recognising vocal emotions is reduced is correlated with the degree to which variations in F0 are attenuated, i.e., the more severe the attenuation of variations in F0, the less informative the F0 cues, and the less accurately listeners recognise vocal emotions (Breitenstein, Van Lancker, et al., 2001; Chatterjee et al., 2015; Frühholz, van der Zwaag, et al., 2016; Leitman et al., 2010; Pak & Katz, 2019; Ritter & Vongpaisal, 2018). Confusion analyses indicate that emotions in speech with uninformative F0 cues are most commonly confused with unemotional

speech rather than with other emotions (Chapter 3.4.2; Chatterjee et al., 2015; Metcalfe, 2017).

According to the weighting-by-reliability hypothesis, listeners adapt to the intrinsic variability in speech by weighting acoustic cues according to their usefulness (Toscano & McMurray, 2010). From this perspective, when listeners are faced with uninformative F0 cues, they may rely on more informative cues provided by other acoustic features, such as intensity and speech rate, to decode vocal emotions. Evidence exists that increased reliance on intensity cues may support recognition of emotions in the absence of F0 cues to emotional prosody (Chapter 3; Luo et al., 2007), although intensity cues alone are not sufficient for successful recognition of vocal emotions (Chapter 3; Metcalfe, 2017). Other studies suggest that listeners cannot use intensity cues to compensate for uninformative F0 cues when identifying vocal emotions (Gilbers et al., 2015; Luo, 2016; Van de Velde, 2017). Hegarty & Faulkner (2013) posit that listeners may be able to make use of speech-rate cues to recognise emotions in speech in which F0 cues are uninformative, while other studies indicate that whether or not listeners increase their reliance on speech rate when parsing vocal emotions, they are unable to make use of speech-rate cues to recognise emotions in speech with uninformative F0 cues (Chapter 3; Gilbers et al., 2015; Luo, 2016; Van de Velde, 2017). Further, speech rate, of itself, is not sufficient for successful recognition of vocal emotions (Chapter 3; Metcalfe, 2017). Additional evidence from studies of question/statement intonation in speech with uninformative F0 cues suggests that listeners can increase their reliance on intensity cues (Marx et al., 2015; Meister et al., 2011; Peng et al., 2012, 2009) and, in cases of severe spectral degradation, duration cues (Peng et al., 2012), to discriminate between the intonations of questions and statements. The high variability in all these studies, confirmed by Everhardt et al.'s (2020) metanalysis of studies examining recognition of vocal emotions and intonation when variations in F0 are attenuated—specifically in CI listeners and in NH listeners listening to simulated CI-speech—suggests that the ability to reweight acoustic cues on the basis of usefulness and to extract meaningful information from them may be listener-specific.

At the group level, functional neuroimaging of cortical metabolic activity, using functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS) demonstrate that aurally perceived speech evokes haemodynamic activity in bilateral superior temporal gyrus (STG; Table 5.1; for review, Alho et al., 2014; Ferrari & Quaresima, 2012; Harrison & Hartley, 2019; Price, 2012; van de Rijt et al., 2018). Reduced activity in bilateral STG is reported for speech with uninformative F0 cues relative to natural speech, and increased listening challenge is associated with increased left IFG activity, whether due to reduced intelligibility or increased syntactic complexity (Table 5.1).

fNIRS and fMRI studies consistently report that vocal emotions, relative to neutral or unemotional speech, evoke increased haemodynamic activity in right, relative to left STG (see Table 5.1). Right lateralisation of STG activity was also observed using fNIRS (see Chapter 4; and Table 5.1), consistent with the proposed specialisation of right STG for decoding speech melody and changes in pitch (Flinker et al., 2019; Friederici & Gierhan, 2013; Poeppel, 2003). The involvement of bilateral IFG in processing vocal emotions has been demonstrated with fMRI, but not with fNIRS (Chapter 4; and Table 5.1). Using fMRI, Ethofer et al. (2009) demonstrate the importance of high spatial resolution for obtaining signatures of discrete emotions: signatures can be decoded from neural activity recorded in bilateral temporal regions, but are not reflected in estimates of haemodynamic responses averaged across brain regions. Most fMRI and fNIRS studies average over cortical areas, and accordingly, report highly variable cortical activity associated with discrete emotions, including increased activity in left and/or right STG, IFG and MFG for angry, right STG and left IFG for happy, and left or right MFG for sad speech (see Table 4.1 in Chapter 4 for synthesis of fMRI and fNIRS findings). Chapter 4 also confirmed that fNIRS is not sensitive to the cortical representations of discrete emotions.

To date, the accounts of functional cortical activity evoked by vocal emotion (i.e., fMRI and fNIRS studies; Table 5.1 in Chapter 4 and Chapter 4) have focussed on group-level activity evoked by easily recognised emotions. The patterns of haemodynamic activity underlying reduced accuracy of recognition of vocal emotions remain to be investigated for vocal emotions in both natural speech and speech with uninformative F0 cues. Two electroencephalography (EEG) studies have associated reduced accuracy of vocal emotion recognition with right-lateralisation of cortical activity: Kislova & Rusalova (2009) interpreted the right-lateralisation in NH listeners listening to vocal emotions in natural speech as an indication of increased attention to the emotional content of the speech. Cartocci et al. (2021) posit that right-lateralisation in CI listeners may reflect reduced left hemisphere activity resulting from neuroplastic changes induced by prolonged deafness. While these two studies implicate the right cortical hemisphere in processing vocal emotions, techniques with higher spatial resolution than EEG, such as fNIRS and fMRI (Pinti et al., 2020), are needed to link activity patterns to specific cortical generators of that activity.

The present study examines cortical haemodynamic activity evoked in right and left STG and IFG by vocal emotions conveyed in natural speech and speech with attenuated variations in F0, providing a neural context for the behavioural deficit in recognising emotions conveyed in speech with uninformative F0 cues. This study will provide insight into the cortical haemodynamic activity associated with reduced behavioural accuracy in hearing-impaired listeners, with implications for CI users for whom uninfor-

mative (or reduced) F0 cues are common. Recognising vocal emotions when F0 cues are uninformative likely draws on domain-general auditory activity in bilateral STG and increased left IFG activity in challenging listening situations, as well as activity in right STG related to prosody processing.

These different aspects of speech perception and processing can be reconciled through the following hypotheses. In behavioural accuracy tasks, vocal emotions with uninformative F0 cues are rarely recognised accurately and often judged as unemotional speech. Assuming that speech with uninformative F0 cues is processed cortically as unemotional speech (as indicated by behavioural confusion rates, e.g., Chapter 3.4.2), vocal emotions with uninformative F0 cues will evoke decreased haemodynamic activity in bilateral STG, relative to natural speech in which F0 cues to vocal emotions are present. Correspondingly, if processing vocal emotions when F0 cues are uninformative presents a substantial listening challenge, activity will increase in left IFG. Extending these hypotheses, variability in listeners' accuracy in recognising vocal emotions when F0 cues are uninformative will most likely be evident in brain regions associated with challenging listening scenarios and with the processing of vocal emotions; increasing accuracy will correspond with reduced left IFG and increased right STG activity. Further, vocal emotions (i.e., both in natural speech and speech with attenuated variations in F0) will evoke increased haemodynamic activity in bilateral STG relative to silence, whereby the activity observed in the STG will be enhanced in right relative to left STG.

To test these hypotheses, cortical haemodynamic activity will be recorded with fNIRS. fNIRS is a non-invasive, quiet neuroimaging technique that measures changes in cortical oxygenation without interacting with ferrous metal implants (such as hearing devices), or requiring participants to lie in small spaces, as in fMRI (Bortfeld, 2019). To measure changes in cortical oxygenation, fNIRS transmits near-infrared light from an emitting light source, through the scalp, skull, and cortex, back to the scalp where it is received by a light detector (Quaresima et al., 2012). Source-detector pairs, positioned ~30 mm apart, create channels that probe the oxygenation of the cortex ~15 mm below the scalp. The near-infrared light, which penetrates most biological tissue, is absorbed by oxygenated (HbO) and deoxygenated (HbR) haemoglobin, allowing the relative concentrations of these two chromophores (HbO and HbR) to be quantified by the relative change in the intensity of the transmitted light (Ferrari & Quaresima, 2012; Quaresima & Ferrari, 2019; Scholkmann, Kleiser, et al., 2014). Stimulus-evoked haemodynamic neural activity is characterised by an increased concentration of HbO and a concomitantly decreased concentration of HbR, referred to as a haemodynamic response (Scholkmann, Kleiser, et al., 2014). Cortical changes in HbO and HbR evoked by vocal emotions conveyed in natural speech and speech with uninformative F0 cues will be measured from bilateral STG and IFG and examined to gain insight into the

Table 5.1. Cortical haemodynamic activity evoked by natural speech and speech with uninformative F0 cues, as well as challenging and emotional speech

		Functional neuroimaging technique	
Comparison	**ROI**	**fMRI**	**fNIRS**
Speech vs. silence	Bilateral STG; increased	Belin et al., 2002 Benson et al., 2001 Evans et al., 2014 Fecteau et al., 2004 Okada et al., 2010 Sevy et al., 2010	Mushtaq et al., 2019 Sato et al., 1999 Sevy et al., 2010 Wiggins et al., 2016
Emotional vs. unemotional speech	Right STG; increased	Beaucousin et al., 2006 Kotz et al., 2006 Kreitewolf et al., 2014 Seydell-Greenwald et al., 2020 von Cramon et al., 2003 Witteman et al., 2012	Sonkaya & Bayazıt, 2018 Zhang et al., 2018 Zhen et al., 2021
	Bilateral IFG; increased	Leitman et al., 2010 Witteman et al., 2012	-
Speech with uninformative F0 cues vs. natural speech	Bilateral STG; decreased	Davis & Johnsrude, 2003 Pollonini et al., 2014 Wild, Davis, et al., 2012	Lawrence et al., 2018 Olds et al., 2016 Pollonini et al., 2014
	Bilateral STG; no difference	Evans et al., 2014	Wijayasiri et al., 2017
	Right STG; increased	Meyer et al., 2004	-
Challenging listening vs. easy listening	Left IFG; increased	*Reduced speech intelligibility*: Davis & Johnsrude, 2003 Eisner et al., 2010 Evans et al., 2014 Meyer et al., 2004 Vaden et al., 2017 Wild, Davis, et al., 2012 Wild, Yusuf, et al., 2012 Zekveld et al., 2006	*Reduced speech intelligibility:* Lawrence et al., 2018 Wijayasiri et al., 2017
		Syntactic complexity: Alain et al., 2018 Friederici & Gierhan, 2013 Lee et al., 2016	*Syntactic complexity:* Hassanpour et al., 2015

Notes. Summary of findings from adult studies that compare NH listeners' cortical responses to natural speech, speech with uninformative F0 cues, as well as challenging and emotional speech, as recorded with fMRI and fNIRS. The direction of change in cortical activation observed is described for each ROI.

cortical underpinnings of successful and poor recognition of vocal emotion resulting from uninformative F0 cues.

5.3 Method

5.3.1 Participants

21 participants (9 female, 12 male, mean age 27 years, SD=4.6 years) were recruited from Macquarie University. 17 of these participants also participated in the fNIRS experiment described in Chapter 4. Participants were 18–36 years of age, native speakers of English with no known psychological or neurological disorders, and right-handed, with pure-tone audiometric thresholds within normal limits, defined as air-conduction thresholds below 20dB HL for all octave frequencies between 250 and 8000 Hz. All participants met the signal-quality criterion employed, i.e., no more than 2 channels in any ROI of 'critical' quality according to the calibration metric, as described in Chapter 2.2.3. Ethical approval for this study was approved by the Macquarie University Ethics Committee (Reference number: 5201952978351). Informed consent was obtained from all participants before the experimental session; all participants received course credit or an honorarium for their participation.

5.3.2 Stimuli

The behavioural task stimuli (N=40) consisted of *angry*, *happy*, *sad*, and *unemotional* pronunciations of six-syllable pseudo-sentences (i.e., "the biffox is dorval"). For each emotion, five sentences were presented in each natural speech and speech with attenuated variations (i.e., uninformative cues) in F0. Stimuli were generated as detailed in Chapter 2.1. The fNIRS experimental stimuli (N=5) consisted of two ~7.25-s blocks of natural vocal emotions (*happy* and *sad*), the same two blocks of speech with attenuated variations in F0 (*F0_happy* and *F0_sad*), and one 7.25-s block of silence (*control*). Five attention stimuli with a 500-ms, 400-Hz tone superimposed on the speech blocks at a random time (n=5, two *happy*, one *sad*, one *F0_happy*, one *F0_sad*) were included the attention-keeping button-press task. The tone was less intense than the speech to ensure attention to the task. Refer to Chapter 2.1 for more detailed description of stimuli.

5.3.3 Test procedure

All participants completed a behavioural 4-alternative forced-choice emotion recognition task, once before (pre-test) and once after (post-test) fNIRS recordings were made. The task began with 4 practice trials (*angry*, *happy*, *sad*, *unemotional* conveyed in natural speech, once each), followed by 40 test trials (each stimulus once). Participants were instructed to indicate which emotion was conveyed using RB-840 button box (Cedrus

Corporation, San Pedro, USA). The correspondence between response alternatives and buttons was shown on a screen (as in Figure 3.1). Participants completed the task in ~7 minutes without breaks.

Each fNIRS experiment began with four practice trials (*happy*, *sad*, *F0_happy*, and *F0_sad*, n=4). To ensure participants were familiar with the button-press task, two of the practice trials contained the pure-tone. During the main experiment, *happy*, *sad*, *F0_happy*, and *F0_sad*, and *control* (silence) blocks were presented 20 times each in a pseudo-randomised order (n=100). The five attention stimuli containing pure tones were presented twice each at pseudo-random intervals (n=10). The total experiment duration was 50 minutes, in which 114 trials were presented, each separated by a 13 to 23-s inter-stimulus-interval (Figure 4.1).

Participants were instructed that they would hear silence and speech, some of which would sound 'more robotic'. They were told that their task was to sit still throughout the experiment, to listen to the speech and silence, and indicate when they heard a tone overlapping with the speech by pressing the top left button on the button box with their left index finger. Additional details pertaining to the experimental protocol are reported in Chapter 2.2.3.

5.3.4 Equipment

Instrumentation for fNIRS and stimulus presentation described in Chapter 2.2.2.

5.3.5 Data analysis

Behavioural accuracy

Listeners' accuracy in recognising vocal emotions (*angry*, *happy*, *sad*, *unemotional*) conveyed in natural speech and speech with uninformative F0 cues was assessed using data collected during the behavioural task before (pre-test) and after (post-test) the fNIRS recording. The data (consisting of 1680 trials, with 1177 correct responses) were used to fit generalised linear mixed-effects models (GLMMs) with a logit function using lme4 (version 1.1.21; Bates et al., 2015). To test the a priori hypotheses of lower accuracy in recognising vocal emotions conveyed in speech with attenuated variations in F0 relative to emotions in natural speech, and increased accuracy in the post-test relative to the pre-test, treatment contrasts were defined between a) natural speech and speech with attenuated variations in F0, and b) pre-test and post-test.

fNIRS haemodynamic responses

Only vocal emotion (*happy*, *sad*, *F0_happy*, *F0_sad*) and *control* trials were analysed. Familiarisation trials (n=4 per participant), attention trials (n=10 per participant), button presses (n=10 per participant), and trials containing accidental button presses (n=9, in 5 participants) were excluded. As a measure of the fNIRS signal quality, a scalp-coupling index was calculated per channel for frequencies between 0.7–1.35 Hz, providing an index of the correlation of the heart-rate signal in the HbO and HbR signals (Pollonini et al., 2014). 94% of channels had a scalp-coupling index >0.8, indicative of good contact between optodes and the scalp (Figure 5.1).

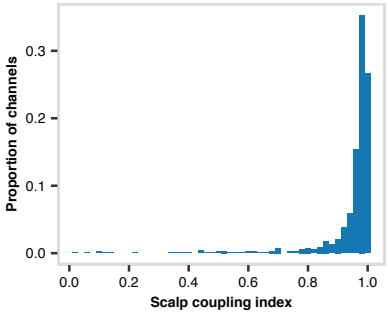


Figure 5.1. Scalp coupling index per channel demonstrating good signal quality. Histogram showing the distribution of scalp-coupling indices, calculated per channel, as a proportion of the total number of channels (N=2856).

The statistical analyses used to generate grand average waveforms, to extract estimates of haemodynamic response amplitude per participant, condition and ROI from first-level analysis, are described in detail in Chapter 2.2.4, as are the procedures used to specify linear mixed-effects (LME) models for second-level (group-level) analyses.

Group-level haemodynamic response amplitudes per condition per ROI. The ROI estimates for right and left STG and IFG were extracted from the first-level analysis, constituting the response variable (N=440 for each chromophore). The model construction, separately for HbO and HbR, began with random intercepts per *Participant*, to allow for overall differences in levels of the measured chromophore between participants. To account for differences between levels of *Condition* within *Participant*, random coefficients (slopes of a categorical variable) of *Condition* per *Participant* were also assessed. Subsequently, fixed effects of *ROI*, *Condition* and the interaction between *ROI* and *Condition* were added.

Relationship between haemodynamic response amplitude and recognition of emotions in speech with attenuated F0 variations. The models were built to include random intercepts and coefficients of *Condition* per *Participant*, and fixed effects of *ROI* and *Accuracy*, as well as the interaction of *ROI* and *Accuracy*. These models were fit to each the HbO and HbR response estimates for all speech conditions (*happy*, *sad*, *F0_happy*, *F0_sad*) extracted from the first-level analysis. Accuracy consisted of a score per participant of the proportion of correctly recognised emotions in speech with attenuated variations in F0, i.e., uninformative F0 cues, before the fNIRS recording.

Exploring the relationship between the magnitude of the HbO-HbR difference and accuracy of emotion recognition in speech with uninformative F0 cues. 'Haemodynamic response magnitude', i.e., the difference between estimated haemodynamic response amplitudes for HbO and HbR, was calculated to synthesise the neural activity measured in the two chromophores into a single metric (called *HbDiff* in previous clinical studies investigating hypo- and hyperoxia, e.g., Kaynezhad et al., 2019; Kolvya et al., 2014; Tachtsidis et al., 2019). This metric serves to address the multicollinearity between HbO and HbR, allowing additional variables (such as accuracy of emotion recognition) to be associated with patterns in the data of both chromophores in a single model. The haemodynamic response magnitude (HRM) metric is the opposite of total haemoglobin (HbT), i.e., the sum of estimated haemodynamic response amplitudes for HbO and HbR, which is used to quantify local cortical blood volume (e.g., Ferrari & Quaresima, 2012; Wolf et al., 2002). Quantifying the magnitude of the difference between the HbO and HbR allows functional cortical responses (both positive and negative) to be isolated, and the influence of their sign to be considered.

For the four speech conditions, the HRM was calculated by subtracting the HbR estimate from the HbO estimate for each participant, ROI and condition. The relationship between HRMs and accuracy per ROI per condition was assessed using an LME, which was built following the procedure described in the planned second-level analysis (i.e., *Magnitude* as the response variable, and stepwise addition of random intercepts and coefficients per participant, with each of *Accuracy* and *ROI* as main effects, and the interaction between *ROI* and *Accuracy*). The intercept was not suppressed, in this instance, as no comparisons against zero were made. Identical models were fit to all HRM values (N=336) and only values calculated from pairs of negatively correlated estimates (i.e., positive HbO and negative HbR estimates, or the inverse, N=158). The pairs of negatively correlated response estimates were examined on the basis that neural-evoked haemodynamic activity is characterised by a negative correlation between the concentrations of HbO and HbR (Wolf et al., 2002) and that the pairs yield valuable insight into the relative proportions of negative and positive haemodynamic responses.

5.4 Results

5.4.1 Behavioural accuracy in recognising vocal emotions

Listeners' accuracy in recognising vocal emotions was assessed as a behavioural measure and investigated as a predictor of the amplitude of haemodynamic responses. As revealed by the behavioural data, listeners recognised the emotions conveyed in natural speech with very high accuracy but exhibited lower accuracy when recognising emotions conveyed in speech in which variations in F0 were attenuated (Figure 5.2). In the behavioural assessment before the fNIRS recordings (pre-test), the mean proportion of correctly recognised emotions in natural speech was close to perfect, i.e., *happy* and sad (M=0.93, SD=0.08), and significantly lower for emotions conveyed in speech with attenuated F0 variations, i.e., *F0_happy* and *F0_sad* (M=0.43, SD=0.17). A similar level of accuracy for both natural and modified speech was evident in the behavioural assessment following the fNIRS recordings (post-test); natural speech (M=0.95, SD=0.08), and attenuated F0 variations (M=0.50, SD=0.14).

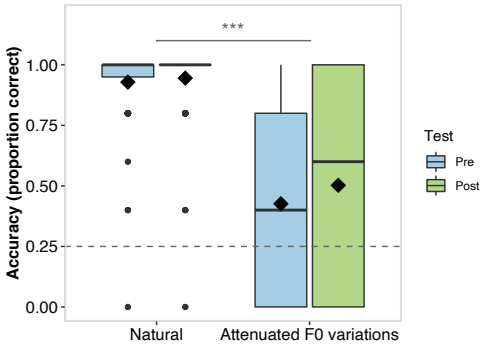


Figure 5.2. The proportion of correct responses aggregated across emotion per condition and test time. Test time refers to whether the assessment was before (pre-test) or after (post-test) the fNIRS recording. Solid bar indicates median, diamond indicates mean, box includes inter-quartile range (IQR), whiskers represent values within 1.5 times the IQR above the 75th or below the 25th percentile, dots indicate values beyond 1.5 times the IQR. Dashed line indicates chance level. Significance for treatment contrast between speech conditions, ***p<0.001.

To predict the accuracy of listeners' recognition of emotions conveyed in natural speech and speech with attenuated variations in F0, a GLMM was developed: *Accuracy* ~ *Condition* + *Test* + (1+*Condition*||*Participant*) + (1|*Stimulus*) + (1+*Condition*+*Test*||*Emotion*). The inclusion of uncorrelated random intercepts and coefficients of *Condition* per *Participant* accounted for a significant proportion of the variance (χ^2(2)=6.96, p=0.008), as did random intercepts of *Stimulus* (χ^2(1)=252.57, p<0.001). The same was true of

uncorrelated random coefficients and intercepts of *Condition* per *Emotion* ($\chi^2(1)=20.73$, $p<0.001$) and *Test* per *Emotion* ($\chi^2(1)=11.54$, $p<0.001$). Significant main effects were observed for *Condition* ($\chi^2(1)=511.35$, $p<0.001$) and *Test* ($\chi^2(1)=6.20$, $p=0.013$), but not for the interaction between *Condition* and *Test* ($\chi^2(1)=0.01$, $p=0.908$). The planned treatment contrasts between levels of *Condition* confirmed that *Accuracy* was significantly lower for emotions conveyed in speech with attenuated variations in F0 than natural speech ($\beta=-4.89$, $SE=0.88$, $z=-5.54$, $p<0.001$), and that *Accuracy* did not differ significantly for the two behavioural assessments ($\beta=0.49$, $SE=0.36$, $z=1.35$, $p=0.178$; Figure 5.2). These data confirm that listeners recognise vocal emotions in natural speech with ease and that uninformative F0 cues to emotions undermine the accuracy of listeners' judgements. They also suggest that listeners' accuracy in recognising emotions with uninformative F0 cues did not improve significantly as a result of the listening task performed during the fNIRS recording.

5.4.2 fNIRS waveforms reveal speech-evoked haemodynamic responses

A primary goal in this study was to determine the effect of attenuating variations in F0 on the cortical haemodynamic responses evoked by vocal emotions. Visual inspection of the grand average waveforms (Figure 5.3) indicates in the control *condition* (i.e., silence) neither the HbO nor HbR waveforms deviated substantially from baseline. Both speech conditions, i.e., natural and with attenuated variations in F0, exhibited elevated concentrations of HbO and reduced concentrations of HbR in bilateral STG, relative to baseline. These concurrent changes in concentration are consistent with the presence of cortical haemodynamic responses (Hong & Santosa, 2016; van de Rijt et al., 2018). Further, the peak of the waveform for HbO, and the minimum for HbR, occurred ~6 s following the onset of the speech stimulus, consistent with the expected morphology of functions describing haemodynamic responses (Scholkmann, Kleiser, et al., 2014). A slight reduction in the concentration of HbO was evident for all four speech conditions bilaterally in IFG, with concentrations of HbR always close to baseline.

Bilaterally in MFG, the time course does not match the morphology of the canonical haemodynamic response function (e.g., Glover, 1999). HbO increased, peaking ~3 s post-stimulus-onset, then dipped below baseline, with a minimum ~10 s post-stimulus-onset before stabilising slightly above baseline around ~18 s post-stimulus-onset, for all speech conditions. Deflections in HbR mirror the time course as HbO but are of reduced amplitude and inverse sign. The time course for the *control* condition in bilateral MFG remains at baseline, suggesting that the deflections are evoked by the speech stimuli. Waveforms for short-channels positioned over the frontal brain regions show a similar time course for HbO, with a peak and dip ~5 μM from baseline (i.e., larger amplitudes than in Figure 5.3). Based on the close resemblance between MFG waveforms

after short-channel regression and the short-channel waveforms, it seems likely that the regression of short-channels may not have entirely accounted for extracerebral signal components, and that traces are still present in the long-channel data. The deflections observed for bilateral MFG are proposed to be of extracerebral and systemic origin, consistent with the stronger presence of the deflections in HbO than HbR (Kirilina et al., 2012). In the present study, the deflections observed in bilateral MFG appear to be stimulus-evoked as they are not apparent in the *control* condition. In Chapter 4, the deflection pattern was observed with greatly reduced amplitudes in the *control* condition, providing evidence that the deflections may be enhanced by the speech stimuli, rather than evoked by them. While further investigation is required to determine whether the deflections in MFG are stimulus-evoked, as observed here, or stimulus-enhanced as in Chapter 4, the MFG response to speech may result from the autonomic nervous system responding to the listening task (Kirilina et al., 2012).

5.4.3 Haemodynamic response amplitude per ROI per condition

To estimate the amplitude of the haemodynamic response in each condition and the four ROIs relevant to the hypotheses (bilateral STG and IFG), the following model was fit to the first-level response estimates for each chromophore separately: $\beta + -1 + ROI + Condition + ROI{:}Condition + (1 + Condition|Participant)$. The model accounted for a similar proportion of the variance in each the HbO ($R^2_{m/c}$=0.16/0.55) and HbR ($R^2_{m/c}$=0.18/0.53) data. For both HbO and HbR, the inclusion of random intercepts per *Participant* and random coefficients of *Condition* per *Participant* (HbO: $\chi^2(14)=54.06$, $p<0.001$, HbR: $\chi^2(14)=63.43$, $p<0.001$) accounted for a significant proportion of variance. The main effects of *ROI* explained a significant amount of variance (HbO: $\chi^2(4)=88.85$, $p<0.001$, HbR: $\chi^2(4)=104.62$, $p<0.001$), whereas *Condition* did not (HbO: $\chi^2(4)=6.99$, $p=0.136$, HbR: $\chi^2(4)=2.16$, $p=0.706$). The *ROI*:*Condition* interaction accounted for significant proportion of the variance (HbO: $\chi^2(12)=25.63$, $p=0.012$, HbR: $\chi^2(12)=35.27$, $p<0.001$).

In left STG, for *sad*, *F0_happy*, and *F0_sad*, concentrations of HbO were significantly higher than zero ($p<0.05$), while that for *happy* was higher than zero, but not significantly so ($p=0.244$; Figure 5.4A). Concentrations of HbR were significantly lower than zero ($p<0.05$) for *happy*, *F0_happy*, and *F0_sad* (Table 5.2). HbR was also lower than zero for *sad*, but this reduction was not significant ($p=0.081$). In right STG, all four speech conditions exhibited significantly higher concentrations of HbO ($p<0.05$) with significantly lower concentrations of HbR ($p<0.05$) relative to zero. The *control* condition did not differ significantly from zero in left or right STG for either chromophore ($p>0.05$; Table 5.2). In left and right IFG, no significant changes in concentrations of either chromophore for any condition were observed ($p>0.05$).

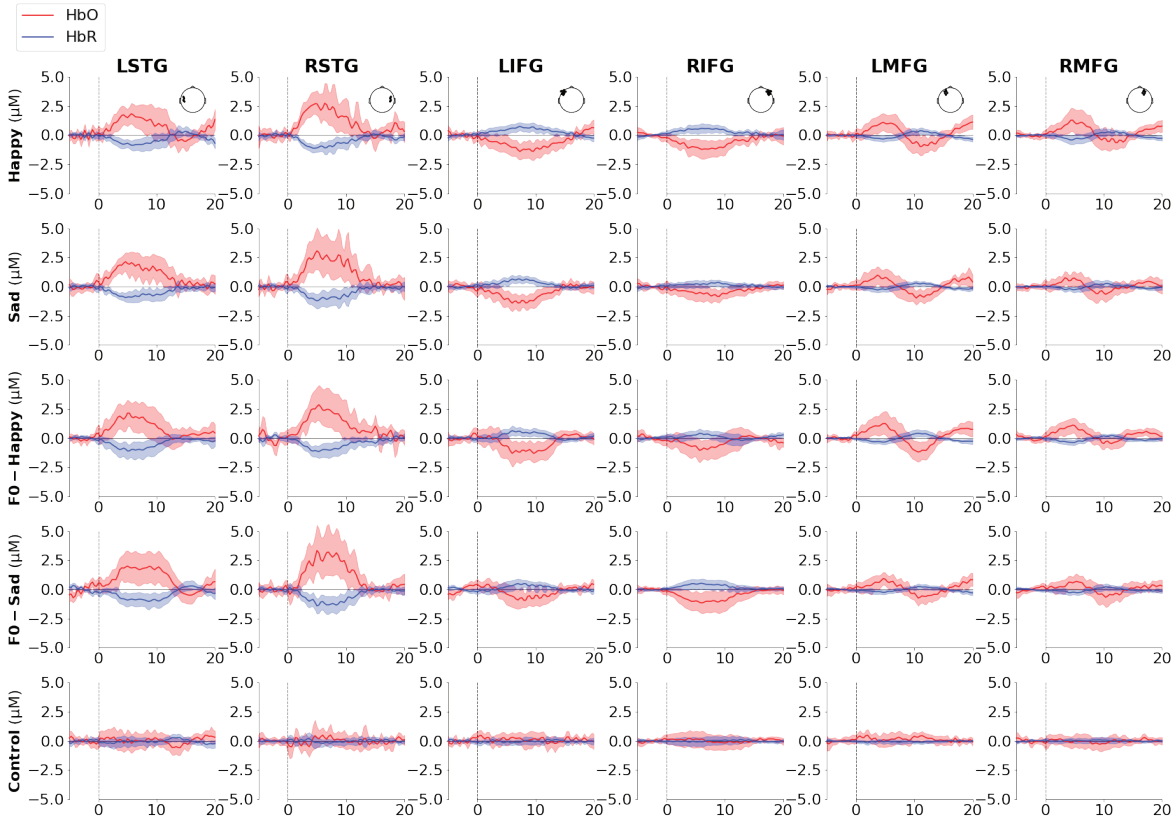


Figure 5.3. Grand average waveforms. Grand average waveforms for all conditions (rows) and ROIs (columns). Solid lines indicate means while shaded areas indicate a 95% confidence interval. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus, LMFG=left middle frontal gyrus, RMFG=right middle frontal gyrus. The inset head shape in the top right corner of each ROI panel (viewed from above) shows the position of the optodes on the head.

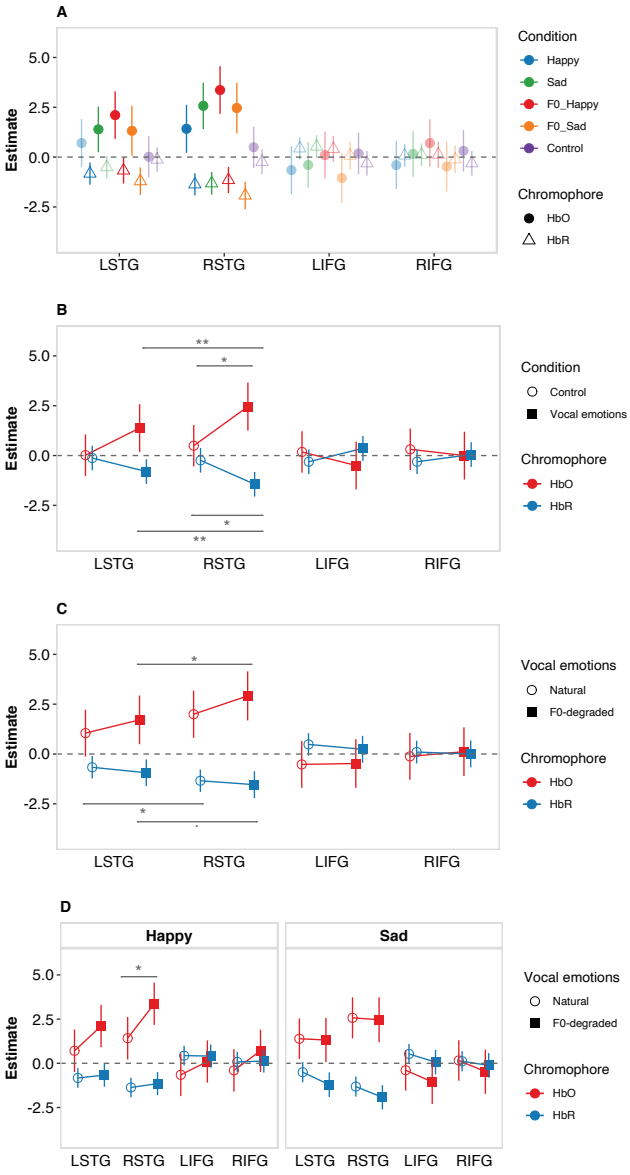


Figure 5.4. Group level estimates and contrasts. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus. A) Group-level response estimates for each ROI, speech condition and chromophore. Darker and transparent symbols indicate significant and non-significant differences from zero, respectively. B) Contrasts *control* vs. aggregated speech conditions for each ROI, as well as left vs. right STG for aggregated speech conditions. C) Contrasts vocal emotions conveyed in natural speech (*happy* + *sad*) vs. speech with uninformative F0 cues (*F0_happy* + *F0_sad*), within each ROI, as well as between left vs. right STG. D) Contrasts between *happy* and *F0_happy*, then *sad* and *F0_sad* for each ROI. Error bars indicate 95%. confidence interval, per response estimate in A, and averaged across aggregated conditions in B, C and D. Significance for contrasts in B, C, and D, *p<0.05, **p<0.01.

Table 5.2. Group-level estimates of haemodynamic response amplitude per condition

		HbO				HbR			
	ROI	**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
Happy	LSTG	0.71	0.60	1.18	0.244	**-0.83**	**0.28**	**-2.97**	**0.004**
	RSTG	**1.42**	**0.60**	**2.37**	**0.022**	**-1.37**	**0.28**	**-4.90**	**<0.001**
	LIFG	-0.65	0.60	-1.06	0.283	0.44	0.28	1.56	0.123
	RIFG	-0.40	0.60	-0.67	0.507	0.08	0.29	0.28	0.782
Sad	LSTG	**1.39**	**0.57**	**2.43**	**0.018**	-0.51	0.29	-1.77	0.081
	RSTG	**2.57**	**0.58**	**4.425**	**<0.001**	**-1.32**	**0.29**	**-4.61**	**<0.001**
	LIFG	-0.39	0.57	-0.69	0.493	0.53	0.29	1.85	0.069
	RIFG	0.16	0.57	0.28	0.778	0.12	0.29	0.40	0.688
F0_happy	LSTG	**2.11**	**0.60**	**3.54**	**0.001**	-0.67	**0.33**	**-2.06**	**0.045**
	RSTG	**3.37**	**0.60**	**5.66**	**<0.001**	-1.15	**0.33**	**-3.53**	**0.001**
	LIFG	0.10	0.60	0.17	0.868	0.41	0.33	1.25	0.216
	RIFG	0.71	0.60	1.19	0.241	0.12	0.33	0.38	0.705
F0_sad	LSTG	**1.32**	**0.62**	**2.12**	**0.040**	**-1.21**	**0.34**	**-3.52**	**0.001**
	RSTG	**2.46**	**0.63**	**3.90**	**<0.001**	**-1.93**	**0.34**	**-5.60**	**<0.001**
	LIFG	-1.05	0.62	-1.69	0.098	0.07	0.34	0.19	0.849
	RIFG	-0.47	0.62	-0.76	0.450	-0.12	0.34	-0.34	0.734
Control	LSTG	0.02	0.52	0.04	0.971	-0.12	0.31	-0.40	0.689
	RSTG	0.49	0.52	0.95	0.348	-0.24	0.31	-0.77	0.442
	LIFG	0.18	0.52	0.35	0.730	-0.31	0.31	-1.01	0.317
	RIFG	0.31	0.52	0.60	0.549	-0.31	0.31	-1.01	0.318

Notes. Significance of contrasts, bold font indicates p<0.05. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus.

Table 5.3. Planned contrasts with group-level estimates of haemodynamic response amplitude

		HbO				HbR			
Contrast		**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
Emotions vs. Control	LSTG	1.36	0.61	2.22	0.095	-0.68	0.37	-1.85	0.235
	RSTG	**1.96**	**0.62**	**3.19**	**0.021**	**-1.20**	**0.37**	**-3.26**	**0.019**
	LIFG	-0.68	0.61	-1.11	0.461	0.67	0.37	1.82	0.235
	RIFG	-0.32	0.61	-0.52	0.755	0.36	0.37	0.98	0.568
Natural vs. F0	LSTG	-0.66	0.50	-1.34	0.351	0.27	0.27	1.03	0.568
	RSTG	-0.92	0.50	-1.83	0.168	0.20	0.27	0.74	0.675
	LIFG	-0.05	0.50	-0.09	0.926	0.24	0.27	0.92	0.568
	RIFG	-0.24	0.50	-0.48	0.755	0.09	0.27	0.35	0.810
Happy_F0 vs. Happy	LSTG	1.40	0.71	1.98	0.137	0.16	0.38	0.42	0.803
	RSTG	**1.94**	**0.71**	**2.75**	**0.034**	0.22	0.38	0.58	0.766
	LIFG	0.75	0.71	1.06	0.461	-0.03	0.38	-0.07	0.943
	RIFG	1.11	0.71	1.57	0.255	0.04	0.38	0.12	0.943
Sad_F0 vs. Sad	LSTG	-0.07	0.77	-0.10	0.926	-0.70	0.46	-1.54	0.353
	RSTG	-0.11	0.78	-0.14	0.926	-0.61	0.46	-1.33	0.449
	LIFG	-0.66	0.77	-0.86	0.555	-0.46	0.46	-1.01	0.568
	RIFG	-0.64	0.77	-0.83	0.555	-0.23	0.46	-0.51	0.776
RSTG vs. LSTG	Emo.	**1.08**	**0.30**	**3.53**	**0.009**	**-0.64**	**0.17**	**-3.81**	**0.003**
	Natural	0.95	0.43	2.20	0.095	**-0.67**	**0.24**	**-2.86**	**0.029**
	F0	**1.20**	**0.43**	**2.79**	**0.034**	-0.60	0.24	-2.53	0.056

Notes. Significance of contrasts, bold font indicates $p<0.05$. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus. Emo=Emotional.

5.4.4 Speech evokes haemodynamic activation in bilateral STG

Contrasts were applied to test the hypothesis that, compared to the silent *control* condition vocal emotions evoke larger haemodynamic responses bilaterally in STG. Contrasts between *control* and vocal emotions (*happy*, *sad*, *F0_happy*, and *F0_sad* aggregated) yielded significant differences in concentrations of both chromophores in right STG, whereby vocal emotions evoked a higher concentration of HbO (p=0.021) and a lower concentration of HbR (p=0.019) relative to the *control* condition (Table 5.3). Left STG exhibited similar numerical trends in both chromophores although contrasts yielded non-significant differences (p>0.05). No significant differences were observed between the *control* and speech in left or right IFG for either chromophore (p>0.05; Table 5.3). The data support the hypothesis of increased haemodynamic activity bilaterally in STG for vocal emotions relative to the *control* condition, with the caveat that the left STG estimates for each chromophore did not differ significantly from the *control* estimates despite being numerically similar to the right STG estimates.

5.4.5 No effect of uninformative F0 cues on the amplitude of haemodynamic responses

To quantify differences between haemodynamic activity generated by vocal emotions in natural speech and speech with uninformative F0 cues, contrasts were computed between the estimates of haemodynamic response amplitude for vocal emotions in each natural speech (*happy* and *sad* aggregated) and speech with attenuated variations in F0 (*F0_happy* and *F0_sad* aggregated). Further contrasts were computed for *happy* vs. *F0_happy* and *sad* vs. *F0_sad*. No significant differences were observed in either chromophore or any ROI for contrasts between vocal emotions conveyed in natural speech vs. speech with attenuated F0 variations (p>0.05; Table 5.3; Figure 5.4C). Contrasts between *happy* and *F0_happy* revealed higher concentrations of HbO in right STG for *F0_happy* relative to *happy* speech (p=0.034); this difference was not observed for HbR in right STG (p=0.766) or any other ROI for either chromophore (p>0.05). No significant differences were found for any ROI or chromophore for *sad* vs. *F0_sad* speech (p>0.05; Table 5.3; Figure 5.4D). Thus, no evidence was found to support the hypotheses that vocal emotions conveyed in speech with uninformative F0 cues evoke a reduction of haemodynamic activity in bilateral STG relative to vocal emotions in natural speech, or that potential increased listening challenge associated with uninformative F0 cues, relative to natural speech, would increase response amplitudes in left IFG.

5.4.6 Vocal emotions evoke right-lateralisation of STG activity

The hypothesis that vocal emotions (*happy*, *sad*, *F0_happy*, and *F0_sad* aggregated) would evoke enhanced haemodynamic activity in right STG relative to left was tested

by contrasting estimates of haemodynamic response amplitude in right STG with left STG (Figure 5.4B). The contrast provided evidence of a right-lateralisation of STG activity evoked by vocal emotions for both HbO (p=0.009) and HbR (p=0.003). For natural speech (*happy* and *sad*), significantly larger haemodynamic response amplitudes were observed for right compared to left STG for HbR (p=0.029), but not HbO (p=0.095), as shown in Figure 5.4C and Table 5.3. For speech with attenuated variations in F0 (*F0_happy* and *F0_sad*) larger-amplitude haemodynamic responses were observed for right compared to left STG for HbO (p=0.034), but the difference only approached significance for HbR (p=0.056). These findings support the hypothesis that the haemodynamic activity evoked by vocal emotions in natural speech is larger in right than left STG and that the same is true for vocal emotions in which F0 cues were rendered uninformative by attenuating variations in F0.

5.4.7 Accuracy of emotion recognition in speech with uninformative F0 cues covaries with haemodynamic response amplitude

The relationship between listeners' abilities to recognise vocal emotions in speech with uninformative F0 cues and the amplitude of haemodynamic responses in each ROI was investigated using a model: β + -1 + *ROI* + *Accuracy* + *ROI*: *Accuracy* + (1 + *Condition*|*Participant*). The model explained a similar proportion of variance for each HbO ($R^2_{m/c}$=0.16/0.55) and HbR ($R^2_{m/c}$=0.22/0.58). The inclusion of random intercepts and coefficients of *Condition* per *Participant* accounts for a significant proportion of the variance (HbO: $\chi^2(9)=28.97$, p=0.001, HbR: $\chi^2(9)=33.96$, $p<0.001$). The main effects of *ROI* explained a significant proportion of the variance for both chromophores (HbO: $\chi^2(4)=87.29$, $p<0.001$, HbR: $\chi^2(4)=117.91$, $p<0.001$), while *Accuracy* did not explain variance significantly (HbO: $\chi^2(1)=2.06$, p=0.151, HbR: $\chi^2(1)=0.42$, p=0.516). The interaction of *ROI*:*Accuracy* accounted for variance significantly for both chromophores (HbO: $\chi^2(3)=8.42$, p=0.038), HbR: $\chi^2(3)=28.24$, $p<0.001$). Likelihood ratio test tests indicated that neither a *Condition:Accuracy or ROI:Condition:Accuracy* interaction explained additional variance for either chromophore (all $p>0.05$), meaning that the above 3-way interaction was not included in the final models.

In right STG, the amplitude of the haemodynamic response decreased significantly with increasing *Accuracy* for both HbO (p=0.030) and HbR (p=0.003). For HbO, the amplitude of the haemodynamic response in left STG (p=0.990) did not covary with *Accuracy* and decreased in IFG bilaterally ($p>0.05$; Table 5.4), though not significantly, with increasing *Accuracy*. Bilaterally in IFG ($p>0.05$), HbR did not covary with *Accuracy*, while a trend of decreasing HbR with increasing *Accuracy* was observed in left STG (p=0.331; Table 5.4; Figure 5.5). In summary, larger amplitude haemodynamic responses in right STG (positive for HbO and negative for HbR) were observed in

listeners who were less accurate in identifying emotions in speech when variations in F0 that convey emotional information were attenuated.

5.4.8 Magnitude of HbO-HbR difference covaries with accuracy of emotion recognition in speech with uninformative F0 cues

Magnitude of the HbO-HbR difference, or haemodynamic response magnitude (HRM), was employed as the response variable to determine whether HRMs can be used to assess the relationship between listeners' abilities to recognise vocal emotions with uninformative F0 cues and the amplitude of their haemodynamic responses in both chromophores in one model while avoiding multicollinearity between the chromophores. HRM values (N=336) were derived from each first-level response estimate pair (calculated as HbO-HbR). Of the 336 derived values, 158 were calculated from negatively correlated HbO and HbR estimate pairs (i.e., theoretically reflecting neural haemodynamic activity). Of these 158 values, 34% were negative, suggesting that approximately one-third of the measured speech-evoked haemodynamic responses were negative, with HbO values below zero (Figure 5.6).

To explore the relationship between the HRMs of negatively correlated pairs of first-level response estimates in each ROI and accuracy, the following model was fit: *Magnitude* ~ +1 + *ROI* + *Accuracy* + *ROI*:*Accuracy* + (1+*Condition*|*Participant*). The fixed and random effects each account a similar proportion of the variance ($R^2_{m/c}$=0.34/0.78). The random intercepts and coefficients of *Condition* per *Participant* account for a significant proportion of the variance ($\chi^2(9)=25.41$, $p=0.003$). The main effect of *ROI* ($\chi^2(3)=84.56$, $p<0.001$) accounted for a significant amount of variance, while *Accuracy* ($\chi^2(1)=1.31$, $p=0.253$) did not. The interaction between *ROI*:*Accuracy* ($\chi^2(3)=26.29$, $p<0.001$) explained a significant proportion of the variance, indicating that the covariance of HRM with accuracy scores differs between ROIs.

This model indicates that *Magnitude* decreased with increasing *Accuracy* in right STG, left and right IFG, and increased with increasing accuracy in left STG (Figure 5.6; Table 5.5). However, the relationship between *Accuracy* and *ROI* was significant only in right STG ($p=0.016$). The same model fit to all evoked HRM values, i.e., both positive and negatively correlated pairs of first-level response estimates, demonstrated the same significant relationship between *Accuracy* and right STG ($p=0.003$), and the same non-significant trends in the other ROIs ($p>0.05$; Table 5.5). These results, therefore, confirm that haemodynamic activity in right STG increases as a listener's ability to extract emotional meaning from speech with uninformative F0 cues decreases. Further, they confirm that HRM is a useful derived metric for investigating the relationship between listeners' abilities to recognise vocal emotions with uninformative F0 cues and the amplitude of their haemodynamic responses.

5.5 Discussion

The primary aims of the present study were 1) to quantify the difference in the amplitude of cortical haemodynamic responses evoked by natural speech and speech in which F0 cues associated with processing vocal emotions are uninformative, and 2) to assess the relationship between haemodynamic responses evoked by vocal emotions and the behavioural accuracy with which listeners recognise emotions in speech with uninformative F0 cues. With regards to the first of these aims, both natural speech and speech with uninformative F0 cues evoked haemodynamic activity bilaterally in STG, each with right-lateralisation of haemodynamic activity observed in one chromophore (HbO or HbR). Second, the accuracy with which listeners recognised emotions in speech with uninformative F0 cues was significantly associated with haemodynamic activity in right STG; analyses of response amplitudes per chromophore and HRM consistently indicated increasing haemodynamic activity in right STG with decreasing accuracy in the behavioural task.

5.5.1 Behavioural accuracy reduced for vocal emotions in speech with uninformative F0 cues relative to natural speech

Listeners recognised emotions (*angry*, *happy*, *sad*, *unemotional*) with approximately equal accuracy before and after they partook in a listening task during fNIRS recording. Emotions in the natural condition were recognised with near-perfect accuracy across all listeners in both assessments, whereas overall recognition of emotions conveyed in speech with uninformative F0 cues increased non-significantly by approximately 7% from before to after the fNIRS task. Nearly half of listeners showed improved accuracy in judging vocal emotions conveyed in speech with uninformative F0 cues between behavioural assessments, with the improvement ranging from 5% to 25%. During fNIRS recordings, listeners heard 20 repetitions of *happy* and *sad* in each natural speech and speech with attenuated F0 variations. While improvement was not possible for natural speech, where accuracy was near perfect, it is possible that the small improvement observed for speech with uninformative F0 cues arose from exposure to the degraded speech signal; listeners may have become more familiar with acoustic characteristics of the speaker's natural voice and/or the distortions introduced by rendering F0 cues uninformative. M. H. Davis et al. (2005) reports that listeners adapt rapidly to noise-vocoded speech containing meaningful words but struggle to adapt to noise-vocoded speech consisting of nonsense words (like the pseudo-words in the stimuli used here). Here, listeners likely had insufficient exposure to the pseudo-word stimuli with attenuated variations in F0 to become optimally acclimatised to the altered speech signal. It is possible that with additional exposure to the pseudo-word stimuli, listeners may learn to recognise emotions more accurately, and that their learning would likely be expedited by using meaningful sentences.

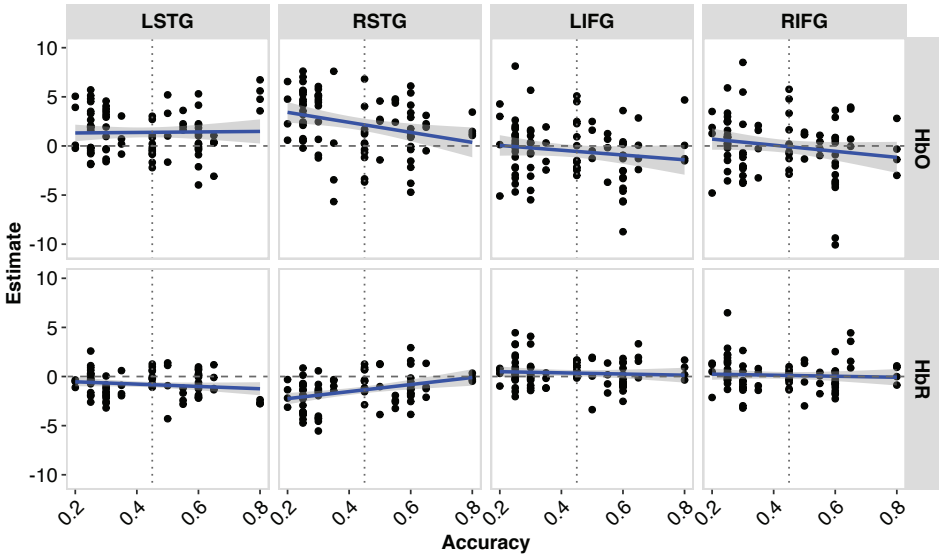


Figure 5.5. First-level estimates as a function of Accuracy for HbO and HbR per ROI. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus. Dots indicate the response estimates (β) extracted from the first-level GLM analysis per ROI, condition, and participant, collapsed across condition. The blue regression line illustrates the linear relationship between beta values for each chromophore and *Accuracy*. Grey shading indicates standard error. The vertical dotted line indicates the median accuracy score (0.45) for vocal emotions with uninformative F0 cues (*F0_happy* and *F0_sad*) assessed before the fNIRS recording (pre-test).

Table 5.4. Relationship between speech-evoked haemodynamic activity and behavioural accuracy of emotion recognition in speech with uninformative F0 cues

	HbO				HbR			
ROI	**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
LSTG	0.03	2.36	0.01	0.990	-1.14	1.16	-0.98	0.331
RSTG	**-5.33**	**2.36**	**-2.26**	**0.030**	**3.63**	**1.16**	**3.14**	**0.003**
LIFG	-2.67	2.36	-1.13	0.265	-0.28	1.16	-0.24	0.811
RIFG	-3.33	2.36	-1.41	0.166	-0.05	1.16	-0.05	0.963

Notes. Slope estimates, bold font indicates $p<0.05$. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus.

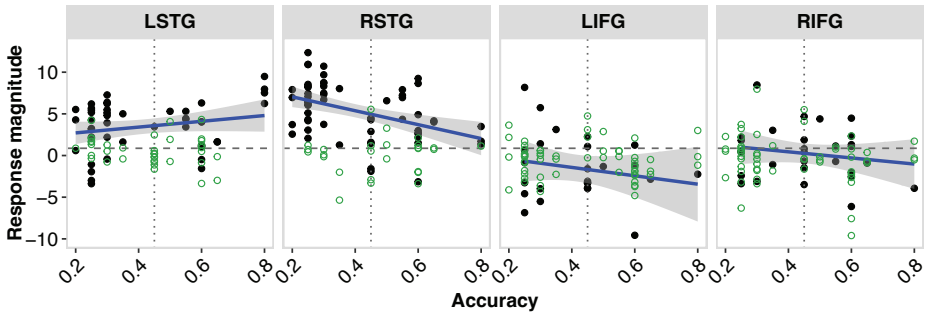


Figure 5.6. Haemodynamic response magnitude (HRM) per ROI interaction with Accuracy. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus. Dots indicate the magnitude of the difference between first-level response estimate pairs (HRM) for HbO and HbR per condition, ROI, and participant. Black dots indicate negatively correlated HbO and HbR values (i.e., one positive and one negative estimate), allowing for positive and negative haemodynamic responses. Open green dots indicate positively correlated HbO and HbR values (i.e., both estimates positive or negative), thus not reflecting functional haemodynamic responses. Negative haemodynamic responses are observed where black dots are below the horizontal dashed line at 0.87 (the mean HRM in the *control* condition of negatively correlated HbO and HbR values). The blue regression line illustrates the relationship between response *Magnitude* and *Accuracy* of the black dots only, with grey shading indicating the standard error. The vertical dotted line indicates the median accuracy score (0.45) for vocal emotions with uninformative F0 cues (*F0_happy* and *F0_sad*) assessed before the fNIRS recording (pre-test),

Table 5.5. Relationship between haemodynamic response magnitude and behavioural accuracy of emotion recognition in speech with uninformative F0 cues

	Negatively correlated pairs				**All pairs**			
ROI	**β**	***SE***	***t***	***p***	**β**	***SE***	***t***	***p***
LSTG	3.12	3.08	1.02	0.319	1.16	2.56	0.45	0.653
RSTG	**-7.87**	**3.06**	**-2.57**	**0.016**	**-8.27**	**2.57**	**-3.22**	**0.003**
LIFG	-2.10	4.04	-0.52	0.604	-2.12	2.56	-0.83	0.413
RIFG	-6.15	3.76	-1.63	0.108	-2.85	2.56	-1.11	0.273

Notes. Slope estimates, bold font indicates $p<0.05$. LSTG=left superior temporal gyrus, RSTG=right superior temporal gyrus, LIFG=left inferior frontal gyrus, RIFG=right inferior frontal gyrus.

5.5.2 Differences between waveforms and estimates of haemodynamic response amplitude

Visual inspection of the morphology of the grand average waveforms per condition per ROI (Figure 5.3) confirmed the presence of haemodynamic responses in left and right STG evoked by all vocal emotions, with response morphology matching that of the canonical haemodynamic response function (e.g., Glover, 1999). From visual inspection, peak values of waveforms and estimates of the GLM group-level response (Figure 5.4A; Table 5.2) are comparable. The morphology of IFG waveforms bilaterally for all speech conditions tends toward negative haemodynamic responses of varying amplitudes (Figure 5.3), and is most obvious for *happy* in IFG bilaterally and sad in left IFG. In the group-level GLM analysis, the values and direction of left IFG responses mirror those observed in the waveforms of IFG responses.

Although bilateral MFG was not included in the group-level GLM analysis, the pattern of deflections in the MFG time course for all vocal emotions matched the pattern described for bilateral MFG in all speech conditions in Chapter 4. In Chapter 4, the amplitude of deflections was present but greatly reduced in the *control* condition relative to the speech conditions, suggesting that the pattern in MFG is not stimulus-evoked and that stimulus-evoked activation in other brain areas may enhance the amplitude of the observed deflections when speech is presented. Here, the MFG response is only observed for speech stimuli, and not in the *control* (silence) condition, which can be interpreted as evidence that the deflections observed in MFG bilaterally are stimulus-evoked. Alternatively, it is possible that deflections in MFG of the small amplitude observed in Chapter 4 in the *control* condition (~0.6 μM), may have been washed out by stronger physiological signal components in the present data. Further investigation is required to determine if the MFG deflections are stimulus-evoked or -enhanced. Nonetheless, the observed MFG response is very similar in Chapter 4 and the present study, with deflections of greater amplitude observed in the short- than long-channel signal in both cases, and more dominant in HbO than HbR in long- and short-channel signals. This indicated that the MFG response most likely reflects extra-cerebral signal components, which were not fully accounted for by the short-channel regression and are likely a response driven by the autonomic nervous system to the listening task (Kirilina et al., 2012; Tachtsidis & Scholkmann, 2016).

5.5.3 Haemodynamic activity evoked by vocal emotions in natural speech and speech with uninformative F0 cues

The amplitudes of haemodynamic responses evoked by vocal emotions in natural speech in STG differed between hemispheres (Figure 5.4A). In right STG, haemodynamic responses were significant for both *happy* and *sad* conveyed in natural speech in

both chromophores. However, in left STG, haemodynamic responses were significant only for HbR but not HbO for *happy* speech, and for HbO but not HbR for *sad*. Speech with uninformative F0 cues evoked significant haemodynamic responses in both hemispheres and both chromophores. In comparison to the response estimates for *happy* speech observed in Chapter 4, where all speech conditions evoked significant haemodynamic responses in STG bilaterally (see Chapter 4.4.2), the current *happy* STG estimates are comparable for HbR but substantially lower for HbO. For *sad*, the current STG estimates are slightly reduced for HbO and substantially closer to zero for HbR than those in Chapter 4.4.2. This may be explicable in terms of the alternation between natural and modified speech resulting from the pseudo-randomised trial order, compared to the constant presentation of natural vocal emotions in Chapter 4. Listeners' predictions about upcoming stimuli dampen the neural activity evoked by expected, but not unexpected, stimuli (e.g., Fishman et al., 2021; Friston, 2005). Speech in which F0 cues to vocal emotions are uninformative may have been less expected than natural speech, as a result of the effect of degraded speech on short-term memory: Retention of words is lower in noise-vocoded CI-simulations, i.e., with uninformative F0 cues, than for natural speech (Bosen & Luckasen, 2019), and natural speech interferes more strongly with working memory than CI-simulated speech (Ellermeier et al., 2015; Wöstmann & Obleser, 2016). Together, the increased salience and better retention of natural speech in short-term memory suggest that vocal emotions conveyed in natural speech are more predictable than those conveyed in speech with uninformative F0 cues, explaining the smaller haemodynamic responses evoked by vocal emotions in natural speech.

The present study predicted that vocal emotions conveyed in speech with uninformative F0 cues would evoke reduced haemodynamic activity in STG bilaterally compared to vocal emotions in natural speech, on the basis of the common misidentification of emotional speech as with *unemotional* when F0 cues are uninformative (Chapter 3.4.2; Metcalfe, 2017). No significant differences were found in any ROI or chromophore between natural speech and speech with uninformative F0 cues—right-lateralisation of haemodynamic activity in STG was observed for both speech conditions. For natural speech, the significantly larger amplitude of haemodynamic responses was observed only in HbR (see also Chapter 4). The right-lateralisation of STG activity, although only significant in one chromophore, is consistent with data reported in fNIRS studies of vocal emotions (Sonkaya & Bayazıt, 2018; D. Zhang et al., 2017, 2018; Zhen et al., 2021), where right-hemispheric lateralisation evoked by vocal emotions in natural speech is thought to reflect that hemisphere's specialisation for decoding acoustic cues that evolve over the span of words or sentences, such as F0 contours (Flinker et al., 2019; Friederici & Gierhan, 2013; Poeppel, 2003).

The right lateralisation observed for vocal emotions in which F0 cues were uninformative, significant for HbO only, is unlikely to reflect the decoding of the F0 contour, which is flat due to the attenuation of variations in F0. Instead, the increased haemodynamic activity observed in right relative to left STG may arise from spectral changes introduced to the speech, listeners attentively seeking to recognise the conveyed emotions, or a combination of these two factors. While no previous studies investigate cortical haemodynamic responses to vocal emotions with uninformative F0 cues, the right-dominant response to speech with uninformative F0 cues is consistent with evidence from one fMRI study of adult participants with non-emotional speech (Meyer et al., 2004), as well as evidence from EEG studies of right-lateralisation of cortical activity in CI listeners (Cartocci et al., 2021), and that spectrally-altered speech evokes increased neural activity in the right hemisphere (Liikkanen et al., 2007; Miettinen et al., 2011, 2012). Additionally, two fNIRS studies (Homae et al., 2007; Wartenburger et al., 2007) examining haemodynamic responses to child-directed speech—characterised by increased variations in F0 and intensity (Fernald & Simon, 1984; Grieser & Kuhl, 1988) report increased activity in right STG activity when variations in F0 are attenuated, consistent with the current findings. However, participants in those studies were 10-month-old infants (Homae et al., 2007) and 4-year-old children (Wartenburger et al., 2007), in whom morphology of cortical haemodynamic responses and sensitivity to vocal emotions are clearly not matured (e.g., Marcar et al., 2004; Morningstar et al., 2018). The right-lateralisation of STG activity may reflect the reduced usefulness of F0 cues in degraded speech, regardless of the emotional valence assigned.

An alternative explanation for the right-lateralisation of STG activity is that the experiment sequence in the present study (i.e., behavioural recognition assessed immediately before and after fNIRS recordings were made), may have led listeners to try to recognise the conveyed emotions explicitly during the listening task. The right-lateralisation of STG activity is consistent with the enhancement of right-hemisphere haemodynamic activity mediated by attention to vocal emotions (e.g., Frühholz et al., 2012; Kotz et al., 2013; Zhen et al., 2021), and accords with Kislova & Rusalova's (2009) suggestion that listeners who struggle to recognise emotions show cortical activation in EEG frequency bands related to increased attention and emotional tension (i.e., gamma and theta, respectively).

5.5.4 Individual behavioural accuracy reflected in amplitude of RSTG haemodynamic responses

The present study suggests the novel finding that the ability to recognise emotions in speech with attenuated variations in F0 is negatively correlated with the amplitude of haemodynamic responses in right STG; as accuracy with which listeners identify

emotions in degraded speech decreased, haemodynamic activity in right STG increased (Figure 5.5 and 5.10). This is consistent with Cartocci et al.'s (2021) report of a negative correlation between the right-lateralisation of the EEG gamma frequency band and accuracy of emotion recognition. One difference is that all of Cartocci et al.'s (2021) participants, both NH and CI listeners, heard full-spectrum speech. Furthermore, the relationship was not assessed separately for each group. It is unclear, therefore, whether the negative correlation between the right-lateralisation and accuracy would be observed in the CI group, whose listening experience might be best compared to normal-hearing listeners listening to speech with uninformative F0 cues.

The increased haemodynamic activity in right STG may reflect the perceptual sensitivity of individual listeners to attenuation of the F0 cues that convey emotions in speech; listeners who show reduced activity in right STG and higher behavioural accuracy may be able to make use of cues extracted from secondary acoustic features, such as intensity, speech rate, or voice quality, more successfully. In Chapter 3, where listeners adjudged emotions conveyed in speech with systematically attenuated variations in F0, intensity and/or speech rate, considerable variability was observed in listeners' abilities to recognise vocal emotions when F0 cues were uninformative. This variability (see Chapter 3.4.1), also observed in the current study, indicates that some listeners are able to overcome uninformative F0 cues by making use of secondary cues. According to the findings of the present study, these listeners would show reduced right STG activity compared to listeners who cannot make use of other informative acoustic cues to recognise vocal emotions.

Individual differences in speech recognition scores for noise-vocoded CI simulations, i.e., speech with uninformative F0 cues, have been suggested to be mediated by the strategies that listeners adopt while learning to decode that speech. For example, taking a 'top-down' strategy, some listeners may seek to decode speech sounds from lexical and semantic information, while others employ a 'bottom-up' strategy, attentive to acoustic-phonetic cues (McGettigan et al., 2014). NH listeners can recognise words in CI-simulation pseudo-sentences, indicating that the listeners can make use of the unattenuated acoustic-phonetic information, i.e., the 'bottom-up' strategy (Loebach et al., 2010). The stimuli in the present experiment were pseudo-words (to control for lexical and semantic information), which rhymed (to control for phonetic information). It is possible that NH listeners in the present study who relied on 'bottom-up' strategies may have been more sensitive to unattenuated acoustic cues, enabling them to recognise vocal emotions more readily in speech with attenuated F0 cues. Listeners' individual abilities to extract emotional meaning from secondary cues, such as intensity and speech rate, are likely mediated by their abilities to perceive these cues, as well as their experience relying on the cues (Jasmin et al., 2019, 2021), and how well acclimatised they are to the altered speech signal (M. H. Davis et al., 2005).

Other predictors of successful speech recognition in spectrally degraded speech include sensitivity to amplitude and spectral modulations (e.g., Erb et al., 2012; Gifford et al., 2018), the size of the vocabulary, working memory, and recognition of visual speech (Rosemann et al., 2017). Volumetric differences in thalamic grey matter (Erb et al., 2012), as well as differences in downregulation of subcortical networks, may also mediate listeners' abilities to adapt to spectrally degraded speech signals (Erb et al., 2013). While the current findings cannot be linked to these predictors, the list of predictors suggests the existence of many potential factors that influence individual differences in the accuracy with which emotions are recognised, and the variable haemodynamic responses in right STG.

Interestingly, the amplitude of the haemodynamic response for HbO in right STG evoked by *happy_F0* was higher than that evoked by *happy*, while the response amplitudes for *sad* and *F0_sad* were very similar (Figure 5.4D). This difference between emotions may be explained by the relative difference in F0 between the original natural and modified speech signals. The mean F0 of the natural vocal emotions was 216.95 Hz for *sad* speech and 243.31 Hz for *happy* speech. To attenuate variations in F0 and render F0 cues uninformative, the F0 of *F0_happy* and *F0_sad* was set to 217 Hz for the duration of the speech. For *sad* speech, an emotion characterised by less variation in F0, F0 variation was attenuated, and the mean F0 increased by 0.05 Hz—less than the 7–12 Hz just-noticeable-difference reported for continuous speech (Rosen & Fourcin, 1986). For *happy* speech which is characterised by wide variations in F0, these variations were attenuated, and the mean F0 was noticeably reduced (i.e., by 26.31 Hz, see Chapter 2.1 for description of stimulus generation). This larger reduction in mean F0 may reduce the ability of listeners to reweight any remaining acoustic cues. In the present study, vocal emotions conveyed in speech with uninformative F0 cues evoke significantly increased activity in right STG relative to natural speech for *happy*, but not *sad*, and in Chapter 3.4.1, behavioural accuracy in the equivalent condition (called *F0*) was lower for *happy* (0.19 correct) than *sad* (0.28 correct). Although the direction of this relationship can be considered consistent with the relationship between behavioural accuracy and activity in right STG in the present study, further investigation of a larger number of emotions is needed to confirm whether differences in behavioural accuracy of individual emotions in speech with uninformative F0 cues correlate with haemodynamic response amplitude in right STG.

5.5.5 Magnitude of HbO-HbR difference–a promising derived metric

Exploratory models examined the relationship between the accuracy with which vocal emotions are recognised when F0 cues are uninformative and the magnitude of the difference between the amplitudes of haemodynamic responses in HbO and HbR. The

exploratory models predicting HRM were consistent with models assessing response amplitude for each chromophore, providing promising evidence that HRM may be a viable composite metric of haemodynamic activity for studies investigating the neural bases of cognitive tasks, in addition to the previous clinical applications (see Kaynezhad et al., 2019; Kolvya et al., 2014; Tachtsidis et al., 2019). HRM values, particularly those including only the negatively correlated pairs of estimates, are conceptually straightforward to interpret, as response amplitudes of HbO and HbR are synthesised into a single value. A compelling advantage of the composite metric is the much-needed oversight it grants into the haemodynamic response patterns observed in individual participants, i.e., the relative proportions of negatively and positively correlated pairs of estimates, beyond simply the positive and negative haemodynamic responses (as visualised in Figure 5.6). Moreover, the HRM metric can facilitate the investigation of relationships between behavioural or cognitive measures and cortical haemodynamic activity by reducing the number of models required, assuming chromophores are modelled separately, or the complexity of models in which chromophores are treated as predictors.

5.5.6 Strengths and limitations

The robust experimental design, as validated in Chapter 4, is an important strength of this study. The ROI-based analysis offers a more reliable approach than channel-wise analysis of haemodynamic data (Wiggins et al., 2016). This approach was complemented by the inclusion of short-detector channels, enabling the regression of extracerebral signal components for the effective isolation of cortically evoked haemodynamic responses (Brigadoi & Cooper, 2015). The second strength of the present study lies in its contribution to understanding the relationship between listeners' abilities to recognise vocal emotions and the amplitude of cortical haemodynamic responses to those emotions. This study contributes novel insight into the enhanced neuro-metabolic cost of the inability to recognise vocal emotions in speech with uninformative F0 cues—a real experience for listeners who use hearing devices. Finally, the exploratory analysis using HRM emphasises the value of synthesising haemodynamic response amplitudes in the two chromophores into a single value. This analysis reveals directional patterns in haemodynamic responses at the level of individual listeners and facilitates analyses involving behavioural measures of listening performance, such as the accuracy of emotion recognition, as investigated here.

The present study also has certain limitations. First, this study could be improved by adding *unemotional* and *F0_unemotional* conditions to facilitate comparisons with neuroimaging studies on noise-vocoded speech. Further, fNIRS data were acquired with fairly broad ROIs, which may reduce the response amplitudes or presence of responses in certain ROIs (Powell et al., 2018; Shader et al., 2021). Future extensions of the

present study might benefit from nesting smaller ROIs within the ROIs analysed here.

5.6 Conclusions

The present study provides novel evidence of an association between the accuracy with which listeners identify vocal emotions when F0 cues to their identification are uninformative and cortical activity in right STG, as determined by haemodynamic responses obtained with fNIRS. Decreasing accuracy in identifying vocal emotions is associated with increasing haemodynamic activity in right STG. The evidence supports the view that vocal emotions—natural and modified through attenuation of F0 cues—evoke significant haemodynamic activity in right STG. No significant differences were observed in the haemodynamic responses of any brain region between vocal emotions conveyed in natural speech or speech in which F0 cues to the identity of those emotions were uninformative. Vocal emotions in natural and modified speech utterances evoked larger haemodynamic response amplitudes in right compared to left STG. Finally, exploratory analyses suggest HRM, i.e., the difference between HbO and HbR estimates of haemodynamic response amplitude, is a useful metric for considering haemodynamic responses at the level of individual listeners and provides a means of correlating cortical neural activity with behavioural measures of listening performance.

Chapter 6| General discussion

6.1 Implications and outlook

In this thesis, I describe a series of studies designed to investigate the relationship between cortical haemodynamic activity in normal-hearing (NH) listeners and their ability to recognise vocal emotions in normal and degraded speech in which cues to vocal emotions were attenuated. Three main observations were made. First, listeners' recognition of emotions conveyed in speech with reduced variations in fundamental frequency (F0), intensity, and/or speech-rate cues was assessed, using speech with attenuated variations in the relevant acoustic feature(s) (Chapter 3). The data demonstrated that listeners' abilities to recognise emotions were most impaired by uninformative F0 cues and more mildly by uninformative intensity and speech-rate cues together. Second, the suitability of functional near-infrared spectroscopy (fNIRS) for obtaining cortical signatures of discrete vocal emotions conveyed in natural speech was investigated (Chapter 4), whereby the data showed that fNIRS cannot differentiate between cortical representations of discrete emotions. Finally, differences in cortical haemodynamic responses to vocal emotions conveyed in natural speech and speech with uninformative F0 cues were assessed at the group level, and the relationship between listeners' abilities to recognise emotions in speech with uninformative F0 cues and their cortical haemodynamic activity while listening to vocal emotions was assessed (Chapter 5). The data showed no difference between patterns of cortical activations to natural speech and speech with uninformative F0 cues, but did reveal that listeners' abilities to recognise emotions in speech with uninformative F0 cues correlate negatively with neural activation to vocal emotions in right superior temporal gyrus (STG); poor accuracy of emotion recognition was associated with increased activity.

Together, these three studies demonstrate that listeners recognise vocal emotions containing uninformative F0 cues with reduced accuracy compared to vocal emotions in unaltered, natural speech. Performance across listeners was highly variable suggesting that listeners differ in their abilities to exploit secondary acoustic cues in the presence of uninformative F0 cues. Listeners' accuracy was negatively correlated with the amplitude of the haemodynamic activity in right STG to vocal emotions conveyed in natural speech and speech with uninformative F0 cues. The experiments also provide insight into a methodological consideration relevant to fNIRS: the difference in amplitude of haemodynamic responses between oxygenated (HbO) and deoxygenated (HbR) haemoglobin can be synthesised into a single metric—'haemodynamic response magnitude' (Chapter 5).

The data presented in this thesis will inform future research on behavioural and cortical aspects of vocal emotion recognition and are pertinent to:

i. Acoustic characterisations of vocal emotions

ii. Processing vocal emotions with hearing devices

iii. fNIRS as a neuroimaging technique for studying vocal emotion, and more generally, auditory processing

iv. Methodological considerations when using fNIRS to assess cortical haemodynamic responses

6.1.1 Acoustic characterisation of vocal emotions

The acoustic analysis of natural speech (Chapter 2) is consistent with previous accounts of the relationships between the acoustic features in natural speech (i.e., F0, intensity, and speech rate) and dimensions used to characterise behavioural and physiological responses to vocal emotions. According to dimensional accounts of vocal emotions (e.g., Goudbeek & Scherer, 2010; Laukka et al., 2005), *angry* and *happy* are higher than *sad* in terms of arousal (dimension ranging from high to low), and *happy* is rated more positive in valence (dimension ranging from negative to positive) than *angry* and *sad*. Applying these arousal and valence ratings to the speech stimuli generated for this thesis (Chapter 2.1), higher arousal was associated with higher mean F0 and intensity, and positive valence was associated with higher speech rates, consistent with previous reports (Bachorowski, 1999; Goudbeek & Scherer, 2010; Laukka et al., 2005).

The role of voice quality, a multifaceted perceptual feature with various acoustic correlates (e.g., the distance between formants and distribution of power in higher to lower frequencies; Gobl & Ní Chasaide, 2003; Scherer, 1986), was not investigated in this thesis. Nonetheless, voice quality can potentially explain two phenomena evoked by angry speech: First, listeners were able to recognise *angry* speech with remarkably high and stable accuracy, despite the attenuation of variations, i.e., cues, in F0, intensity and/or speech rate (Chapter 3). Second, *angry* speech evoked the numerically largest haemodynamic responses of the various vocal emotions (i.e., *angry*, *happy*, *sad*, *unemotional*) in STG bilaterally, although the amplitudes were not found to be significantly different from *unemotional* speech for either chromophore (Chapter 4). The high accuracy and large haemodynamic responses for *angry* speech likely result from the increased vocal effort conveyed in angry speech, which is perceived as a 'tense' sounding voice quality and is associated with increased loudness perception (Gobl & Ní Chasaide, 2003; Yanushevskaya et al., 2013). Listeners make use of voice quality cues to recognise emotions in speech (e.g., Gobl & Ní Chasaide, 2003; Scherer, 1986). As such, voice

quality cues in *angry* speech may have provided cues for successful recognition despite F0, intensity, and/or speech-rate cues having been rendered uninformative. While this explanation involving 'tense' speech pertains specifically to *angry* speech, voice quality likely contributes to the perception of all emotions. While this thesis prioritised using natural speech, future behavioural and neuroimaging studies on vocal emotions should control for voice quality, likely by synthesising speech (e.g., Gobl & Ní Chasaide, 2003) due to the many acoustic features that characterise voice quality.

A limitation of the stimuli used in this thesis is that they were recorded from a single female speaker. Expression of vocal emotions varies between speakers (Laukka et al., 2012; Scherer, 1986), meaning that stimuli recorded from a more diverse set of speakers could improve their generalisability. Further, given the evidence that listeners adapt to altered speech signals more quickly when exposed to meaningful speech (M. H. Davis et al., 2005), future investigations may consider including stimuli with real words rather than pseudo-words.

6.1.2 Processing vocal emotions with hearing devices

Our listeners, all of whom had normal hearing, showed reduced accuracy with high variability, within and between listeners, in recognising emotions conveyed in speech with uninformative F0 cues (Chapters 3 and 5). The high variability is interpreted as evidence that individual listeners differ in their abilities to make use of informative cues, with some listeners successfully assigning more weight to secondary cues to vocal emotions as predicted by Toscano & McMurray's (2010) cue-weighting hypothesis, while other listeners do not adapt their usage of cues to the degraded speech signal. The variability observed between listeners' strategies may be explained by differences in listeners' abilities to perceive the cues provided by each acoustic feature (Jasmin et al., 2019; Kidd et al., 2007) combined with listeners' lack of experience listening to speech with uninformative F0 cues (M. H. Davis et al., 2005; Jasmin et al., 2021). The fNIRS investigation in Chapter 5 revealed that the accuracy with which listeners recognise emotion in speech when F0 cues are uninformative is negatively correlated with haemodynamic activity in right STG, where lower accuracy corresponded to increased activation. This suggests that in NH listeners at least, the ability to compensate for uninformative F0 cues with other informative acoustic cues may be associated with reduced neuro-metabolic demands in right STG.

Therapeutic hearing devices used to mitigate reduced representation of frequency and intensity information in sensorineural hearing loss include hearing aids (HAs), which amplify sounds but may introduce distortions due to microphone placement and front-end signal processing features, and cochlear implants (CIs), which convert acoustic signals to trains of electrical pulses with a heavily compressed intensity range

and limited spectrotemporal resolution (Başkent et al., 2016; Lesica, 2018; Plack, 2018). While CIs transmit F0 information weakly (Başkent et al., 2016; Chatterjee & Peng, 2008; Everhardt et al., 2020; Plack, 2018), F0 can be transmitted through HAs. Nevertheless, in cases of severe hearing loss, a HA user's individual amplification settings, which are matched to the listeners' hearing loss profile, may limit the range of intensity, change the shape of the speech spectrum, and introduce distortions to temporal envelope, thus potentially altering the delivered pitch contours (Goy et al., 2018; Lesica, 2018). For users of both CIs and HAs, the reduced function of the auditory perceptual system in tandem with the altered speech signal provided by the therapeutic device, can influence recognition of vocal emotions. As discussed in Chapter 3.5, the attenuation of variations in F0, used in this thesis to render F0 cues uninformative, and the degradation of F0 cues by CI devices both impact the accuracy with which listeners recognise vocal emotions (e.g., Luo et al., 2007; Most & Aviner, 2009; Pak & Katz, 2019). Consistent with the degradation of the speech signal being less severe in HAs compared to CIs, HA users tend to show a milder reduction in accuracy when recognising vocal emotion (Most & Aviner, 2009). While increasing evidence demonstrates that HA and CI listeners have difficulty recognising vocal emotions due to degraded F0 information (e.g., Everhardt et al., 2020; Goy et al., 2018; Jiam et al., 2017; Most & Aviner, 2009; Waaramaa et al., 2018), little is known about neural mechanisms underpinning poor and successful recognition of vocal emotions in hearing-impaired individuals.

Future investigations should examine whether HA and CI listeners show the same association between cortical activity in right STG and the ability to recognise emotions in speech with uninformative F0 cues evident in NH listeners (Chapter 5). These are needed to assess the viability and reliability of right STG activity as a biomarker for behavioural deficits in recognising emotions conveyed in speech with attenuated F0 cues. Drawing on the stimuli and experiments described in this thesis, the natural-speech stimuli (Chapter 2) could be used to assess the abilities of HA and CI users to recognise emotions in speech, using the behavioural 4-alternative forced-choice experiment (Chapter 3.3). As fNIRS can be used with electrical and ferromagnetic HA or CI components (Saliba et al., 2016), the fNIRS experiment in which participants listened to vocal emotions conveyed in natural speech (Chapter 4.3) is suitable for recording cortical haemodynamic responses in HA or CI listeners. The mode of stimulus presentation would need to be adapted, i.e., the speech stimuli could be presented in the free field or directly to the hearing device, rather than through insert phones. In sum, the stimulus materials and experiments prepared for this thesis are suited to future investigations into the cortical underpinnings of vocal emotion recognition and processing in HA or CI listening.

6.1.3 fNIRS as a neuroimaging technique for studying vocal emotions, and more generally, auditory processing

Using fNIRS, speech-evoked haemodynamic activity was observed bilaterally in STG (Chapters 4 and 5), with the activity evoked by vocal emotions conveyed in natural speech significantly elevated compared to silence (Chapter 4). In Chapter 5, vocal emotions in natural speech and speech with uninformative F0 cues evoked increased haemodynamic activity, compared to silence, in STG bilaterally, although this was only significant for right STG. In left STG, each vocal emotion evoked haemodynamic responses significantly different from baseline in at least one chromophore. Despite the difference in significance observed for left STG between the two studies (see Chapter 5 for discussion), the bilateral activation of cortical areas consisting of the auditory-sensory and surrounding cortices is consistent with previous fNIRS and fMRI findings (e.g., Belin et al., 2002; Evans et al., 2014; Mushtaq et al., 2019; Sevy et al., 2010), strengthening the evidence that fNIRS is robustly sensitive to cortical activity evoked by auditory stimuli such as speech.

Concerning the dual pathway model of auditory processing, the fNIRS montage described in Chapter 2 covers the endpoints of the ventral 'what' pathway, i.e., from the primary auditory regions to the inferior frontal regions (Friederici & Gierhan, 2013). During the processing of vocal emotions, the ventral pathway contributes to the extraction of acoustic features as well as the generation of an auditory percept in STG and appraisal of that percept in IFG (Arnott et al., 2004; Dricu & Frühholz, 2020). In Chapters 4 and 5, vocal emotions evoked cortical haemodynamic activity bilaterally in STG only, likely reflecting feature extraction and percept generation. The absence of cortical activation in inferior and middle frontal regions suggests that fNIRS is not sensitive to the appraisal of vocal emotions in the investigated cortical regions. No evidence for or against the sensitivity of fNIRS to the dorsal stream was presented in this thesis, as the fNIRS montage used in Chapters 4 and 5 did not cover the parietal and prefrontal regions included in the dorsal 'where/how' stream (Dricu & Frühholz, 2020).

fMRI and fNIRS studies commonly report right-lateralisation of cerebral activity evoked by cortical processing of vocal emotions (Frühholz, Trost, et al., 2016; Kreitewolf et al., 2014; Seydell-Greenwald et al., 2020; Witteman et al., 2012). Congruent with these reports, lateralisation of haemodynamic activity evoked by emotional speech to right STG was observed in Chapter 4 (*angry*, *happy* and *sad* speech together) and Chapter 5 (*happy* and *sad* speech together). 17 of 21 participants completed both fNIRS studies, approximately two to three months apart. The consistency of the right-lateralisation of STG activity may be tentatively interpreted as indicating that haemodynamic responses evoked by vocal emotions and recorded with fNIRS are reliable at the group level. Alternatively, fNIRS may be more sensitive to right-hemisphere brain regions (see

higher specificity values described in Chapter 2.2.2). Visualisations of haemodynamic response amplitudes reported in some recent fNIRS studies investigating natural non-emotional speech suggest larger response amplitudes in the right relative to the left temporal lobe (Chapter 4, Lawrence et al., 2018; Luke et al., 2021; Shader et al., 2021; Wiggins et al., 2016), while others do not (Defenderfer et al., 2017; Lawrence et al., 2018; Wijayasiri et al., 2017; Zhou et al., 2018). Brigadoi & Cooper's (2015) model of adult scalp-brain distances does not suggest reduced distances in the right hemisphere compared to the left. Further research is needed to ascertain whether fNIRS is differentially sensitive to right and left cortical regions and whether this potential difference in sensitivity is a pertinent concern in fNIRS investigations of hearing function.

6.1.4 Methodological considerations when using fNIRS to assess cortical haemodynamic responses

Although the focus of this thesis was on behavioural and cortical representations of vocal emotions in NH listeners, exploratory analyses employed the underused metric of the 'haemodynamic response magnitude' (HRM; Chapter 5). In fNIRS studies, best practice is to report analyses of both chromophores, i.e., HbO and HbR (Tachtsidis & Scholkmann, 2016; Yücel et al., 2021), as was performed in Chapters 4 and 5. Some studies also report total haemoglobin (HbT), in which estimates for HbO and HbR obtained in the first-level analysis are summed to reflect total local blood volume (Ferrari & Quaresima, 2012; Wolf et al., 2002). To explore how the inverse of HbT could characterise measured neural activity, HRM values were derived by subtracting HbR from HbO estimates of response amplitude. HRM reflects the difference between the estimates for the two chromophores, whereby a larger difference value represents increased neural activity. Analyses incorporating HRM as the dependent variable, with ROI and accuracy with which emotion are recognised as independent variables demonstrated that this metric provides results consistent with separate analyses of the two chromophores while providing additional insight into the direction of the global haemodynamic response. Negative haemodynamic responses (i.e., reduced HbO and elevated HbR concentrations) are less commonly reported and less well understood than positive haemodynamic responses (elevated HbO and reduced HbR; Howarth et al., 2020; Maggioni et al., 2015; Mullinger et al., 2014). For individual participants, HRM values convey the direction, as well as the potential presence, of the haemodynamic response. This metric, therefore, characterizes haemodynamic responses of individuals more informatively and succinctly than estimates of response amplitude for each HbO and HbR alone, i.e., analysed separately to reduce multicollinearity. In group-level analyses, the HRM embodies the information conveyed in both HbO and HbR, facilitating the association of additional variables, such as measures of listening performance, to the combined haemodynamics of the two chromophores. As demonstrated in Chapter

5, HRM is a valuable metric whose implementation should be further validated in future fNIRS studies.

Finally, the breadth of ROIs has been shown to influence observations of haemodynamic activity (e.g., Powell et al., 2018; Shader et al., 2021). As described in Chapter 2.2.2, relatively broad ROIs were used in Chapters 4 and 5 to increase measurement reliability (Wiggins et al., 2016) and account for individual differences in cortical anatomy and functionality (Cooper et al., 2012; D. Wang et al., 2015). Shader et al. (2021) demonstrate that averaging haemodynamic responses over broader ROIs may obscure haemodynamic activity measured by smaller groups of neighbouring channels. To extend upon the experiments described in Chapters 4 and 5, smaller ROIs within the larger ROIs could be investigated, enabling assessment of responses to discrete emotions (Chapter 4) and emotions in natural speech and speech with uninformative F0 cues (Chapter 5) at finer spatial resolution. The ROIs used in this thesis may be subdivided into smaller ROIs or simply reduced in size by increasing the 'specificity' threshold in the fOLD toolbox (Zimeo Morais et al., 2018), or alternatively, by identifying the functional channels of interest for each participant using an independent sample of data (i.e., unused trials; Powell et al., 2018). The latter approach is of particular interest, as it has the potential to account for individual differences in cortical anatomy and functionality while ensuring that the planned statistical analyses are unbiased by ROI selection (Vul & Kanwisher, 2013).

Chapter 7| Conclusions

This thesis encompasses the generation of novel vocal emotion stimuli, as well as studies investigating the behavioural accuracy with which normal-hearing listeners recognise emotions conveyed in speech, and cortical haemodynamic activity evoked by emotions in speech measured with functional near-infrared spectroscopy (fNIRS). Behavioural accuracy was investigated for vocal emotions conveyed in natural speech and speech with uninformative acoustic cues (i.e., attenuated variations in fundamental frequency (F0), intensity and/or speech rate), and cortical haemodynamic responses to vocal emotions in natural speech and speech with uninformative F0 cues were assessed.

The data confirm that NH listeners rely most heavily on variations in F0 to recognise emotions conveyed in speech and that rendering F0 cues uninformative by attenuating variations in F0 significantly impairs listeners' recognition of vocal emotions. To a lesser degree than F0, rendering intensity and speech rate cues simultaneously uninformative also reduces the accuracy with which listeners recognise vocal emotions. The cortical activity evoked by vocal emotions conveyed in natural speech and speech with uninformative F0 cues did not differ significantly; both evoked cortical activity bilaterally in superior temporal gyrus (STG). Vocal emotions in both speech conditions evoked a right lateralisation of STG activity, and this was negatively correlated with the ability to recognise emotions conveyed in speech with uninformative F0 cues. Together, the data suggest that NH listeners may be able to make use of variations in acoustic features (i.e., intensity and speech rate) when F0 cues are uninformative in order to recognise vocal emotions, and that success in doing so is reflected in haemodynamic responses recorded from right STG using fNIRS. fNIRS is not sensitive to unique cortical signatures evoked by discrete vocal emotions, but rather to the cortical representations of emotional speech and listeners' accuracy of emotion recognition in speech with uninformative F0 cues.

Summary

Problem statement. Normal-hearing (NH) listeners rely heavily on variations over time in the fundamental frequency of speech (F0; the acoustic correlate of voice pitch) to identify vocal emotions. Without reliable F0 cues, such as is the case for individuals who rely on cochlear implants to hear, the ability to extract emotional meaning from speech is reduced. This thesis describes the development of an objective measure for recognising vocal emotions conveyed in speech. A program of three experiments investigates: 1) the ability of NH listeners to use F0, intensity, and speech-rate cues to recognise vocal emotions; 2) a signature of the cortical representation of discrete vocal emotions assessed using functional near-infrared spectroscopy (fNIRS); 3) a characterisation of cortical haemodynamic activity evoked by vocal emotions in natural speech and in speech manipulated to render F0 cues less informative using fNIRS.

Experiment 1. To validate the novel emotional stimuli created for this project and investigate how NH listeners make use of F0, intensity, and speech-rate cues when recognising vocal emotions, accuracy scores were obtained for NH listeners' recognition of *angry*, *happy*, *sad,* and *unemotional* conveyed in natural speech, as well as speech conditions with systematically attenuated variations in F0, intensity and/or speech rate. Listeners identified emotions in natural speech with near-perfect accuracy. Comparisons between speech conditions with less informative cues confirmed that, at the group level, listeners rely most heavily on F0 cues when judging vocal emotions and show limited abilities to use intensity and speech-rate cues to compensate for uninformative F0 cues. High accuracy for emotions in natural speech demonstrates the suitability of these stimuli for investigating cortical representations of discrete emotions in natural speech.

Experiment 2. Cortical haemodynamic responses were recorded from bilateral superior temporal gyri (STG), inferior frontal gyri (IFG), and middle frontal gyri (MFG) using fNIRS while NH listeners heard *angry*, *happy*, *sad*, and *unemotional* speech to determine whether fNIRS can be used to measure cortical signatures of discrete vocal emotions in natural speech. Speech evoked significant haemodynamic responses in bilateral STG, and right-lateralisation of cortical activity evoked by vocal emotions was observed in STG. Comparisons of haemodynamic response amplitude between *unemotional* speech and each vocal emotion suggest fNIRS is not sensitive to cortical representations of discrete emotions. While fNIRS may not be suited to investigating cortical haemodynamic responses to discrete vocal emotions, fNIRS may still enable the investigation of cortical mechanisms supporting speech processing and sensitivity to vocal emotions more generally.

Experiment 3. Using fNIRS, cortical haemodynamic responses were recorded bilaterally from STG and IFG while listeners heard emotional speech with variations in F0 intact

and attenuated to determine if cortical activation evoked by vocal emotions is correlated with NH listeners' abilities to recognise emotions with uninformative F0 cues. Significant haemodynamic activity in right STG was evoked by speech (natural and with uninformative F0 cues). Response amplitudes did not differ significantly between vocal emotions in natural speech and speech with uninformative F0 cues in STG or IFG. The amplitude of the haemodynamic response in right STG was significantly correlated with listeners' abilities to recognise vocal emotions with uninformative F0 cues. The data show fNIRS to be a promising technique with which to obtain an objective measure of listeners' ability to recognise vocal emotions in speech with uninformative F0 cues.

Discussion and conclusions. This series of behavioural and fNIRS experiments confirms that NH listeners rely most heavily on F0 cues and make use of other acoustic information with variable success when judging vocal emotions. They also demonstrate that vocal emotions are processed preferentially in right STG and that the activity in the right STG is positively correlated with listeners' abilities to recognise emotions conveyed in speech with uninformative F0 cues. Future research can use these findings, therefore, to delve into the cortical activity underlying vocal emotion recognition in listeners with aided hearing.

Nederlandse Samenvatting

Probleemstelling. Normaalhorende (NH) luisteraars zijn sterk afhankelijk van variaties in de fundamentele frequentie van spraak (F0; het akoestische equivalent van de toonhoogte van de stem) om vocale emoties te herkennen. Zonder betrouwbare F0-kenmerken, wat het geval is bij mensen die een cochleair implantaat gebruiken om te horen, is het vermogen om emotionele betekenis uit spraak te halen verminderd. Dit proefschrift beschrijft de ontwikkeling van een objectieve maat voor de herkenning van vocale emoties in spraak. Een set van drie experimenten onderzoekt: 1) het vermogen van NH-luisteraars om F0-, luidheid- en spreeksnelheidskenmerken te gebruiken om vocale emoties te herkennen; 2) de corticale representatie van discrete vocale emoties gemeten aan de hand van functionele nabij-infraroodspectroscopie (fNIRS); 3) de corticale hemodynamische activiteit veroorzaakt door vocale emoties in natuurlijke spraak en in spraak waarbij F0-kenmerken gereduceerd werden.

Experiment 1. Dit experiment valideert de emotionele stimuli die gecreëerd werden voor dit project en onderzoekt hoe NH-luisteraars gebruik maken van F0-, luidheid-, en spreeksnelheidskenmerken om vocale emoties te herkennen. Hierbij lag de nadruk op het herkennen van *boosheid, blijdschap, verdriet* en *emotieloosheid,* in zowel natuurlijke spraak als in spraak waarbij variaties in F0, luidheid en/of spreeksnelheid systematisch gereduceerd werden. Nauwkeurigheidsscores tonen aan dat NH-luisteraars emoties in natuurlijke spraak vrijwel perfect kunnen identificeren. Een vergelijking tussen de verschillende spraakcondities bevestigt dat, op groepsniveau, luisteraars het meest afhankelijk zijn van F0-kenmerken bij het beoordelen van vocale emoties. Verder toont dit onderzoek aan dat wanneer er sprake is van niet-informatieve F0-kenmerken, NH-luisteraars moeilijk kunnen compenseren met luidheid- en spreeksnelheid-kenmerken. De hoge nauwkeurigheid voor de herkenning van emoties in natuurlijke spraak bevestigt dat de stimuli geschikt zijn voor verder onderzoek naar corticale representaties van discrete emoties in natuurlijke spraak.

Experiment 2. Het tweede experiment in dit proefschrift onderzoekt of fNIRS corticale kenmerken van discrete vocale emoties in natuurlijk spraak kan opmeten. Corticale hemodynamische reacties werden gemeten terwijl NH-luisteraars naar *boze, blijde, verdrietige* en *emotieloze* spraak luisterden. Activiteit werd geregistreerd ter hoogte van de volgende hersenregio's: bilaterale gyri temporalis superior (STG), gyri frontalis inferior (IFG), en gyri frontalis medius (MFG). De resultaten tonen aan dat spraak significante hemodynamische reacties veroorzaakt in STG, bilateraal. Vocale emoties resulteerde in lateralisatie van de corticale activiteit in de rechter STG. Om het effect van discrete vocale emoties te onderzoeken, werden hemodynamische responsamplitudes tussen *emotieloze* spraak en elke andere vocale emotie met elkaar vergeleken. Geen van deze

vergelijkingen resulteerde in een significant effect. Dit suggereert dat fNIRS niet gevoelig is voor corticale representaties van discrete emoties. fNIRS is daarom niet geschikt voor het onderzoeken van discrete vocale emoties, maar de techniek kan wel verder inzicht verschaffen in spraakverwerking, en gevoeligheid voor emoties in het algemeen.

Experiment 3. Het laatste experiment bestudeert de relatie tussen het vermogen van NH-luisteraars om emoties te herkennen met de corticale activiteit uitgelokt door diezelfde vocale emoties. Hiertoe werd - met behulp van fNIRS - corticale hemodynamische reacties gemeten van bilaterale STG en IFG terwijl NH vrijwilligers naar emotionele spraak luisterden. Twee spraakcondities werden gepresenteerd: natuurlijke spraak en spraak met niet-informatieve F0-kenmerken. Beide spraakcondities lokken significante hemodynamische activiteit uit in de rechter STG. Er was echter geen significant verschil tussen responsamplitudes uitgelokt door de twee spraakcondities. Een belangrijke bevinding van dit experiment is de significante correlatie tussen de responsamplitude in de rechter STG en het vermogen van de luisteraars om vocale emoties zonder informatieve F0-kenmerken te herkennen. fNIRS is dus een veelbelovende techniek die mogelijks het vermogen van luisteraars om vocale emoties in spraak zonder informatieve F0-kenmerken te herkennen objectief in kaart kan brengen.

Discussie en conclusie. Deze reeks gedrags- en fNIRS- experimenten bevestigt dat NH-luisteraars het meest F0-kenmerken benutten en met wisselend succes gebruik maken van andere akoestische informatie tijdens het beoordelen van vocale emoties. Verder tonen deze experimenten aan dat vocale emoties voornamelijk in de rechter STG worden verwerkt, en dat de activiteit in de rechter STG positief gecorreleerd is met het vermogen van luisteraars om emoties te herkennen in spraak zonder informatieve F0-kenmerken. Deze bevindingen kunnen gebruikt worden voor toekomstig onderzoek naar corticale activiteit die ten grondslag ligt aan de herkenning van vocale emotie bij luisteraars die hoortoestellen en/of cochleaire implantaten gebruiken.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014

Agnoli, S., Mancini, G., Pozzoli, T., Baldaro, B., Russo, P. M., & Surcinelli, P. (2012). The interaction between emotional intelligence and cognitive ability in predicting scholastic performance in school-aged children. *Personality and Individual Differences*, *53*(5), 660–665. https://doi.org/10.1016/j.paid.2012.05.020

Agrawal, D., Thorne, J. D., Viola, F. C., Timm, L., Debener, S., Büchner, A., Dengler, R., & Wittfoth, M. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. *NeuroImage: Clinical*, *2*, 229–238. https://doi.org/10.1016/j.nicl.2013.01.001

Agrawal, D., Timm, L., Viola, F. C., Debener, S., Büchner, A., Dengler, R., & Wittfoth, M. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. *BMC Neuroscience*, *13*(113), 1–10. https://doi.org/10.1186/1471-2202-13-113

Al-Radhi, M. S., Csapó, T. G., & Németh, G. (2019). Adaptive refinements of pitch tracking and HNR estimation within a vocoder for statistical parametric speech synthesis. *Applied Sciences*, *9*(12), 2460. https://doi.org/10.3390/app9122460

Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *39*(7), 2695–2709. https://doi.org/10.1002/hbm.24031

Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research*, *307*, 29–41. https://doi.org/10.1016/j.heares.2013.08.001

Anuardi, M. N. A. M., & Yamazaki, A. K. (2019). Effect of emotionally toned Malay language sounds on the brain: a NIRS analysis. *International Journal on Perceptive and Cognitive Computing*, *5*(1), 1–7. https://doi.org/10.31436/ijpcc.v5i1.72

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, *22*(1), 401–408. https://doi.org/10.1016/j.neuroimage.2004.01.014

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*(2), 919–927. https://doi.org/10.1016/j.neuroimage.2008.05.034

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. https://doi.org/10.1111/1467-8721.00013

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Barker, J. W., Aarabi, A., & Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical Optics Express*, *4*(8), 1366–1379. https://doi.org/10.1364/boe.4.001366

Başkent, D., Gaudrain, E., Tamati, T., & Wagner, A. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. Holt, & P. van Dijk (Eds.), *Scientific foundations of audiology: perspectives from physics, biology, modeling, and medicine* (pp. 285–319). Plural Publishing, Inc.

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

Bauernfeind, G., Wriessnegger, S. C., Daly, I., & Müller-Putz, G. R. (2014). Separating heart and brain: on the reduction of physiological noise from multichannel functional near-infrared spectroscopy (fNIRS) signals. *Journal of Neural Engineering*, *11*(5), 056010. https://doi.org/10.1088/1741-2560/11/5/056010

Beaucousin, V., Lacheret, A., Turbelin, M.-R. M. R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex*, *17*(2), 339–352. https://doi.org/10.1093/cercor/bhj151

Beaucousin, V., Zago, L., Hervé, P.-Y., Strelnikov, K., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2011). Sex-dependent modulation of activity in the neural networks engaged during emotional speech comprehension. *Brain Research*, *1390*, 108–117. https://doi.org/10.1016/j.brainres.2011.03.043

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/BRM.40.2.531

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience*, *3*(10), 965–966. https://doi.org/10.1038/79890

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., & Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, *78*(3), 364–396. https://doi.org/10.1006/brln.2001.2484

Bernstein, J. G. W., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *120*, 3929–3945. https://doi.org/10.1121/1.2372452

Bestelmeyer, P. E., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social Cognitive and Affective Neuroscience*, *12*(8), 1351–1358. https://doi.org/10.1093/scan/nsx059

Bezooijen, R. van. (1984). *Characteristics and recognizability of vocal expressions of emotion*. De Gruyter Mouton. https://doi.org/10.1515/9783110850390

Bittner, R. M., Humphrey, E., & Bello, J. P. (2016). Pysox: Leveraging the Audio Signal Processing Power of Sox in Python. *17th International Society for Music In- Formation Retrieval Conference*, 4–6. https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/08/bittner-pysox.pdf

Boden, S., Obrig, H., Köhncke, C., Benav, H., Koch, S. P., & Steinbrink, J. (2007). The oxygenation response to functional stimulation: Is there a physiological meaning to the lag between parameters? *NeuroImage*, *36*(1), 100–107. https://doi.org/10.1016/j.neuroimage.2007.01.045

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291. https://doi.org/10.1016/j.ijhcs.2006.11.016

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. https://doi.org/10.1038/nn1409

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. In *Version* (6.0.37). http://www.praat.org/

Bortfeld, H. (2019). Functional near-infrared spectroscopy as a tool for assessing speech and spoken language processing in pediatric and adult cochlear implant users. *Developmental Psychobiology*, *61*(3), 430–443. https://doi.org/10.1002/dev.21818

Bosen, A. K., & Luckasen, M. C. (2019). Interactions between item set and vocoding in serial recall. *Ear and Hearing*, *40*(6), 1404–1417. https://doi.org/10.1097/AUD.0000000000000718

Breitenstein, C., Lancker, D. Van, & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition & Emotion*, *15*(1), 57–79. https://doi.org/10.1080/02699930126095

Breitenstein, C., Van Lancker, D., Daum, I., & Waters, C. H. (2001). Impaired perception of vocal emotions in Parkinson's disease: Influence of speech time processing and executive functioning. *Brain and Cognition*, *45*(2), 277–314. https://doi.org/10.1006/brcg.2000.1246

Brigadoi, S., & Cooper, R. J. (2015). How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. *Neurophotonics*, *2*(2), 025005. https://doi.org/10.1117/1.nph.2.2.025005

Brosgole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience*, *82*(3–4), 169–189. https://doi.org/10.3109/00207459508999800

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage*, *58*(1), 259–268. https://doi.org/10.1016/j.neuroimage.2011.06.005

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1–2), 135–148. https://doi.org/10.1163/156770908X289242

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research*, *9*(3), 227–238. https://doi.org/10.1016/s0926-6410(99)00060-9

Buxton, R. (2010). Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. *Frontiers in Neuroenergetics*, *2*, 1–16. https://doi.org/10.3389/fnene.2010.00008

Buxton, S. L., MacDonald, L., & Tippett, L. J. (2013). Impaired recognition of prosody and subtle emotional facial expressions in Parkinson's disease. *Behavioral Neuroscience*, *127*(2), 193–203. https://doi.org/10.1037/a0032013

Cannon, S. A., & Chatterjee, M. (2019). Voice emotion recognition by children with mild-to-moderate hearing loss. *Ear and Hearing*, *40*(3), 477–492. https://doi.org/10.1097/AUD.0000000000000637

Carlson, R., Granström, B., & Nord, L. (1992). Experiments with emotive speech-acted utterances and synthesized replicas. *Second International Conference on Spoken Language Processing*.

Cartocci, G., Giorgi, A., Inguscio, B. M. S., Scorpecci, A., Giannantonio, S., De Lucia, A., Garofalo, S., Grassia, R., Leone, C. A., Longo, P., Freni, F., Malerba, P., & Babiloni, F. (2021). Higher right hemisphere gamma band lateralization and suggestion of a sensitive period for vocal auditory emotional stimuli recognition in unilateral cochlear implant children: An EEG study. *Frontiers in Neuroscience*, *15*, 1–10. https://doi.org/10.3389/fnins.2021.608156

Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, *23*(1), 91–100. https://doi.org/10.1023/A:1021339410262

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*, *25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163

Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, *235*(1–2), 143–156. https://doi.org/10.1016/j.heares.2007.11.004

Chatterjee, M., Zion, D. J., Deroche, M. L. D., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, *322*, 151–162. https://doi.org/10.1016/j.heares.2014.10.003

Christensen, J. A., Sis, J., Kulkarni, A. M., & Chatterjee, M. (2019). Effects of age and hearing loss on the recognition of emotions in speech. *Ear and Hearing*, *40*(5), 1069–1083. https://doi.org/10.1097/AUD.0000000000000694

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *American Speech-Language-Hearing Association*, *23*(7), 493–500.

Cook, R. D. (2011). Cook's distance. In M. Lovric (Ed.), *International encyclopedia of statistical science*. Springer. https://doi.org/10.1007/978-3-642-04898-2_189

Cooper, R. J., Caffini, M., Dubb, J., Fang, Q., Custo, A., Tsuzuki, D., Fischl, B., Wells, W., Dan, I., & Boas, D. A. (2012). Validating atlas-guided DOT: A comparison of diffuse optical tomography informed by atlas and subject-specific anatomies. *NeuroImage*, *62*(3), 1999–2006. https://doi.org/10.1016/j.neuroimage.2012.05.031

Corcoran, C. M., Keilp, J. G., Kayser, J., Klim, C., Butler, P. D., Bruder, G. E., Gur, R. C., & Javitt, D. C. (2015). Emotion recognition deficits as predictors of transition in individuals at clinical high risk for schizophrenia: A neurodevelopmental perspective. *Psychological Medicine*, *45*(14), 2959–2973. https://doi.org/10.1017/S0033291715000902

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, *54*(4), 2808–2821. https://doi.org/10.1016/j.neuroimage.2010.10.069

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage*, *49*(4), 3039–3046. https://doi.org/10.1016/j.neuroimage.2009.11.050

Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. Avon Books.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, *23*(8), 3423–3431. https://doi.org/10.1523/jneurosci.23-08-03423.2003

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of Data Science*, *8*, 61–73.

Defenderfer, J., Kerr-German, A., Hedrick, M., & Buss, A. T. (2017). Investigating the role of temporal lobe activation in speech perception accuracy with normal hearing adults: An event-related fNIRS study. *Neuropsychologia*, *106*, 31–41. https://doi.org/10.1016/j.neuropsychologia.2017.09.004

Delpy, D. T., Cope, M., Van Der Zee, P., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, *33*(12), 1433–1442. https://doi.org/10.1088/0031-9155/33/12/008

Demenescu, L. R., Mathiak, K., & Mathiak, K. A. (2014). Age-and gender-related variations of emotion recognition in pseudowords and faces. *Experimental Aging Research*, *40*(2), 187–207. https://doi.org/10.1080/0361073X.2014.882210

Deroche, M. L. D., Felezeu, M., Paquette, S., Zeitouni, A., & Lehmann, A. (2019). Neurophysiological differences in emotional processing by cochlear implant users, extending beyond the realm of speech. *Ear and Hearing*, *40*(5), 1197–1209. https://doi.org/10.1097/AUD.0000000000000701

Dick, F., Saygin, A. P., Galat, G., Pitzalis, S., Bentrovato, S., D'Amico, S., Wilson, S., Bates, E., & Pizzamiglio, L. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: Evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799–816. https://doi.org/10.1162/jocn.2007.19.5.799

Dick, F., Taylor Tierney, A., Lutti, A., Josephs, O., Sereno, M. I., & Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *Journal of Neuroscience*, *32*(46), 16095–16105. https://doi.org/10.1523/JNEUROSCI.1712-12.2012

Dricu, M., & Frühholz, S. (2020). A neurocognitive model of perceptual decision-making on emotional signals. *Human Brain Mapping*, *41*(6), 1532–1556. https://doi.org/10.1002/hbm.24893

Drugman, T., Huybrechts, G., Klimkov, V., & Moinet, A. (2018). Traditional machine learning for pitch detection. *IEEE Signal Processing Letters*, *25*(11), 1745–1749. https://doi.org/10.1109/LSP.2018.2874155

Duncan, A., Meek, J. H., Clemence, M., Elwell, C. E., Tyszczuk, L., Cope, M., & Delpy, D. (1995). Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy. *Physics in Medicine and Biology*, *40*(2), 295–304. https://doi.org/10.1088/0031-9155/40/2/007

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*(21), 7179–7186. https://doi.org/10.1523/JNEUROSCI.4040-09.2010

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America*, *138*(3), 1561–1569. https://doi.org/10.1121/1.4928954

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia*, *50*(9), 2154–2164. https://doi.org/10.1016/j.neuropsychologia.2012.05.013

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, *33*(26), 10688–10697. https://doi.org/10.1523/JNEUROSCI.4596-12.2013

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–1268. https://doi.org/10.1162/jocn.2009.21099

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of Emotional Information in Voice-Sensitive Cortices. *Current Biology*, *19*(12), 1028–1033. https://doi.org/10.1016/j.cub.2009.04.054

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., Mourão-Miranda, J., Wise, R. J. S., & Scott, S. K. (2014). The pathways for intelligible speech: Multivariate and univariate perspectives. *Cerebral Cortex*, *24*(9), 2350–2361. https://doi.org/10.1093/cercor/bht083

Everhardt, M. K., Sarampalis, A., Coler, M., Başkent, D., & Lowie, W. (2020). Meta-analysis on the identification of linguistic and emotional prosody in cochlear implant users and vocoder simulations. *Ear & Hearing*, *41*(5), 1092–1102. https://doi.org/10.1097/AUD.0000000000000863

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage*, *23*(3), 840–848. https://doi.org/10.1016/j.neuroimage.2004.09.019

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, *63*(2), 921–935. https://doi.org/10.1016/j.neuroimage.2012.03.049

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. *NeuroImage*, *184*, 171–179. https://doi.org/10.1016/j.neuroimage.2018.09.025

Fishman, Y. I., Lee, W. W., & Sussman, E. (2021). Learning to predict: Neuronal signatures of auditory expectancy in human event-related potentials. *NeuroImage*, *225*, 117472. https://doi.org/10.1016/j.neuroimage.2020.117472

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, *3*(4), 393–405. https://doi.org/10.1038/s41562-019-0548-z

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412–429. https://doi.org/10.1037/0033-2909.97.3.412

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*(1), 30–40. https://doi.org/10.1006/nimg.1997.0306

Frost, J. A. (1999). Language processing is strongly left lateralized in both sexes: Evidence from functional MRI. *Brain*, *122*(2), 199–208. https://doi.org/10.1093/brain/122.2.199

Frühholz, S., & Ceravolo, L. (2018). The neural network underlying the processing of affective vocalizations. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 430–458). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.19

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, *22*(5), 1107–1117. https://doi.org/10.1093/cercor/bhr184

Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *NeuroImage*, *62*(3), 1658–1666. https://doi.org/10.1016/j.neuroimage.2012.06.015

Frühholz, S., & Grandjean, D. (2013a). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(1), 24–35. https://doi.org/10.1016/j.neubiorev.2012.11.002

Frühholz, S., & Grandjean, D. (2013b). Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2847–2855. https://doi.org/10.1016/j.neubiorev.2013.10.007

Frühholz, S., & Staib, M. (2017). Neurocircuitry of impaired affective sound processing: A clinical disorders perspective. *Neuroscience and Biobehavioral Reviews*, *83*, 516–524. https://doi.org/10.1016/j.neubiorev.2017.09.009

Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, *123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Frühholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A. K., Vuilleumier, P., & Grandjean, D. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective valence. *Social Cognitive and Affective Neuroscience*, *11*(10), 1638–1649. https://doi.org/10.1093/scan/nsw066

Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N., & Boas, D. A. (2012). Short separation channel location impacts the performance of short channel regression in NIRS. *NeuroImage*, *59*(3), 2518–2528. https://doi.org/10.1016/j.neuroimage.2011.08.095

Gifford, R. H., Noble, J. H., Camarata, S. M., Sunderhaus, L. W., Dwyer, R. T., Dawant, B. M., Dietrich, M. S., & Labadie, R. F. (2018). The relationship between spectral modulation detection and speech recognition: Adult versus pediatric cochlear implant recipients. *Trends in Hearing*, *22*, 1–14. https://doi.org/10.1177/2331216518771176

Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., & Başkent, D. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *I-Perception*, *6*(5), 1–19. https://doi.org/10.1177/0301006615599139

Giordano, B. L., Whiting, C., Kriegeskorte, N., Kotz, S. A., Gross, J., & Belin, P. (2021). The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01073-0

Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, and Psychophysics*, *75*, 1799–1810. https://doi.org/10.3758/s13414-013-0518-x

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*(4), 416–429. https://doi.org/10.1006/nimg.1998.0419

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1–2), 189–212. https://doi.org/10.1016/S0167-6393(02)00082-1

Goerlich-Dobre, K. S., Witteman, J., Schiller, N. O., van Heuven, V. J. P., Aleman, A., Martens, S., Heuven, V. J. P. Van, Martens, S., van Heuven, V. J. P., Aleman, A., & Martens, S. (2014). Blunted feelings: Alexithymia is associated with a diminished neural response to speech prosody. *Social Cognitive and Affective Neuroscience*, *9*(8), 1108–1117. https://doi.org/10.1093/scan/nst075

Gold, R., Butler, P., Revheim, N., Leitman, D. I., Hansen, J. A., Gur, R. C., Kantrowitz, J. T., Laukka, P., Juslin, P. N., Silipo, G. S., & Javitt, D. C. (2012). Auditory emotion recognition impairments in schizophrenia: Relationship to acoustic features and cognition. *American Journal of Psychiatry*, *169*(4), 424–432. https://doi.org/10.1176/appi.ajp.2011.11081230

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044. https://doi.org/10.1038/sdata.2016.44

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322. https://doi.org/10.1121/1.3466853

Goy, H., Pichora-Fuller, M. K., Singh, G., & Russo, F. A. (2018). Hearing aids benefit recognition of words in emotional speech but not emotion identification. *Trends in Hearing*, *22*, 1–16. https://doi.org/10.1177/2331216518801736

Grady, C. L., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., & Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport*, *8*(11), 2511–2516. https://doi.org/10.1097/00001756-199707280-00019

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 1–13. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–146. https://doi.org/10.1038/nn1392

Grieser, D. A. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*(1), 14–20. https://doi.org/10.1037/0012-1649.24.1.14

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science*, *71*(5), 901–911. https://doi.org/10.1086/425944

Gruber, T., Debracque, C., Ceravolo, L., Igloi, K., Marin Bosch, B., Frühholz, S., & Grandjean, D. (2020). Human discrimination and categorization of emotions in voices: a functional Near-Infrared Spectroscopy (fNIRS) study. *Frontiers in Neuroscience*, *14*, 1–14. https://doi.org/10.3389/fnins.2020.00570

Haeussinger, F. B., Heinzel, S., Hahn, T., Schecklmann, M., Ehlis, A. C., & Fallgatter, A. J. (2011). Simulation of near-infrared light absorption considering individual head and prefrontal cortex anatomy: Implications for optical neuroimaging. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0026377

Halberstadt, A. G., & Hall, J. A. (1980). Who's getting the message? Children's nonverbal skill and their evaluation by teachers. *Developmental Psychology*, *16*(6), 564–573. https://doi.org/10.1037/0012-1649.16.6.564

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior*, *33*(3), 149–180. https://doi.org/10.1007/s10919-009-0070-5

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice*, *21*(5), 531–540. https://doi.org/10.1016/j.jvoice.2006.03.002

Harrison, S. C., & Hartley, D. E. H. (2019). Shedding light on the human auditory cortex: A review of the advances in near infrared spectroscopy (NIRS). *Reports in Medical Imaging*, *12*, 31–42. https://doi.org/10.2147/RMI.S174633

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6:e4794. https://doi.org/10.7717/peerj.4794

Hassanpour, M. S., Eggebrecht, A. T., Culver, J. P., & Peelle, J. E. (2015). Mapping cortical responses to speech using high-density diffuse optical tomography. *NeuroImage*, *117*, 319–326. https://doi.org/10.1016/j.neuroimage.2015.05.058

He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain–computer interfaces. *Brain Sciences*, *10*(10), 687. https://doi.org/10.3390/brainsci10100687

Hegarty, L., & Faulkner, A. (2013). The perception of stress and intonation in children with a cochlear implant and a hearing aid. *Cochlear Implants International*, *14*(sup4), 35–39. https://doi.org/10.1179/1467010013z.000000000132

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42–46. https://doi.org/10.1111/j.1467-8721.2008.00545.x

Homae, F., Watanabe, H., Nakano, T., & Taga, G. (2007). Prosodic processing in the developing brain. *Neuroscience Research*, *59*(1), 29–39. https://doi.org/10.1016/j.neures.2007.05.005

Hong, K. S., & Nguyen, H.-D. (2014). State-space models of impulse hemodynamic responses over motor, somatosensory, and visual cortices. *Biomedical Optics Express*, *5*(6), 1778–1798. https://doi.org/10.1364/boe.5.001778

Hong, K. S., & Santosa, H. (2016). Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hearing Research*, *333*, 157–166. https://doi.org/10.1016/j.heares.2016.01.009

Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, *15*(2), 136–146. https://doi.org/10.1080/09297040802403682

House, A. S., & House. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, *33*(9), 1174–1178. https://doi.org/10.1121/1.1908941

House, D. (1994). Perception and production of mood in speech by cochlear impant users. *Third International Conference on Spoken Language Processing*, 491–494.

Howarth, Mishra, & Hall. (2020). More than just summed neuronal activity: how multiple cell types shape the BOLD response. *Philosophical Transactions of the Royal Society B: Biological Sciences*. http://eprints.whiterose.ac.uk/166097/

Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, *3*(1), 010401. https://doi.org/10.1117/1.nph.3.1.010401

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382. https://doi.org/10.1016/j.neuroimage.2005.08.065

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K. K., Poldrack, R. A., & Boas, D. A. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *NeuroImage*, *20*(1), 479–488. https://doi.org/10.1016/S1053-8119(03)00311-2

Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2019). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, *149*(5), 914–934. https://doi.org/10.1037/xge0000688

Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481. https://doi.org/10.1016/j.cognition.2020.104481

Jiam, N. T., Caldwell, M., Deroche, M. L., Chatterjee, M., & Limb, C. J. (2017). Voice emotion perception and production in cochlear implant users. *Hearing Research*, *352*, 30–39. https://doi.org/10.1016/j.heares.2017.01.006

Jiang, X., Robin, J., Pell, M. D., Paulmann, S., Robin, J., & Pell, M. D. (2015). More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 597–612. https://doi.org/10.1037/xhp0000043

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of Emotion From Vocal Cues. *Archives of General Psychiatry*, *43*(3), 280. https://doi.org/10.1001/archpsyc.1986.01800030098011

Julien, C. (2006). The enigma of Mayer waves: Facts and models. *Cardiovascular Research*, *70*(1), 12–21. https://doi.org/10.1016/j.cardiores.2005.11.008

Jürgens, R., Drolet, M., Pirow, R., Scheiner, E., & Fischer, J. (2013). Encoding conditions affect recognition of vocally expressed emotions across cultures. *Frontiers in Psychology*, *4*, 1–10. https://doi.org/10.3389/fpsyg.2013.00111

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381–412. https://doi.org/10.1037/1528-3542.1.4.381

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-019-01701-x

Kaynezhad, P., Mitra, S., Bale, G., Bauer, C., Lingam, I., Meehan, C., … Tachtsidis, I. (2019). Quantification of the severity of hypoxic-ischemic brain injury in a neonatal preclinical model using measurements of cytochrome-c-oxidase from a miniature broadband-near-infrared spectroscopy system. *Neurophotonics*, *6*(04), 045009. https://doi.org/10.1117/1.nph.6.4.045009

Khan, M. N. A., Bhutta, M. R., & Hong, K. S. (2020). Task-specific stimulation duration for fNIRS brain-computer interface. *IEEE Access*, *8*, 89093–89105. https://doi.org/10.1109/ACCESS.2020.2993620

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*, 418–435. https://doi.org/10.1121/1.2743154

Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A. M., & Tachtsidis, I. (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *NeuroImage*, *61*(1), 70–81. https://doi.org/10.1016/j.neuroimage.2012.02.074

Kislova, O. O., & Rusalova, M. N. (2009). EEG asymmetry in humans: Relationship with success in recognizing emotions in the voice. *Neuroscience and Behavioral Physiology*, *39*, 825–831. https://doi.org/10.1007/s11055-009-9213-8

Koch, K., Stegmaier, S., Schwarz, L., Erb, M., Reinl, M., Scheffler, K., Wildgruber, D., & Ethofer, T. (2018). Neural correlates of processing emotional prosody in unipolar depression. *Human Brain Mapping*, *39*, 3419–3427. https://doi.org/10.1002/hbm.24185

Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer-Lambert law revisited. *Physics in Medicine and Biology*, *51*(5), N91–N98. https://doi.org/10.1088/0031-9155/51/5/N02

Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology*, *4*, 105. https://doi.org/10.3389/fpsyg.2013.00105

Kolyva, C., Ghosh, A., Tachtsidis, I., Highton, D., Cooper, C. E., Smith, M., & Elwell, C. E. (2014). Cytochrome c oxidase response to changes in cerebral oxygen delivery in the adult brain shows higher brain-specificity than haemoglobin. *NeuroImage*, *85*, 234–244. https://doi.org/10.1016/j.neuroimage.2013.05.070

Korb, S., Frühholz, S., & Grandjean, D. (2014). Reappraising the voices of wrath. *Social Cognitive and Affective Neuroscience*, *10*(12), 1644–1660. https://doi.org/10.1093/scan/nsv051

Kornreich, C., Brevers, D., Canivet, D., Ermer, E., Naranjo, C., Constant, E., Verbanck, P., Campanella, S., & Noël, X. (2013). Impaired processing of emotion in music, faces and voices supports a generalized emotional decoding deficit in alcoholism. *Addiction*, *108*(1), 80–88. https://doi.org/10.1111/j.1360-0443.2012.03995.x

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, *34*(8), 1971–1981. https://doi.org/10.1002/hbm.22041

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Progress in Brain Research*, *156*, 285–294. https://doi.org/10.1016/S0079-6123(06)56015-7

Kranefeld, I., & Blickle, G. (2021). Emotion recognition ability and career success: Assessing the roles of GMA and conscientiousness. *Personality and Individual Differences*, *168*, 110370. https://doi.org/10.1016/j.paid.2020.110370

Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage*, *102*, 332–344. https://doi.org/10.1016/j.neuroimage.2014.07.038

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*(5), 341–372. https://doi.org/10.1016/j.pneurobio.2004.03.006

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The impact of 'no opinion' response options on data quality. Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, *66*(3), 371–403. https://doi.org/10.1086/341394

Kuang, J., & Liberman, M. (2016). The effect of vocal fry on pitch perception. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5260–5264. https://doi.org/10.1109/ICASSP.2016.7472681

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia*, *41*(8), 1082–1090. https://doi.org/10.1016/S0028-3932(02)00294-4

Lambrecht, L., Kreifelts, B., & Wildgruber, D. (2012). Age-related decrease in recognition of emotional facial and prosodic expressions. *Emotion*, *12*(3), 529–539. https://doi.org/10.1037/a0026827

Laukka, P. (2004). *Vocal Expression of Emotion: Discrete-emotions and Dimensional Accounts* [Doctoral thesis, University of Uppsala]. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A165425&dswid=69

Laukka, P., Audibert, N., & Aubergé, V. (2012). Exploring the determinants of the graded structure of vocal emotion expressions. *Cognition and Emotion*, *26*(4), 710–719. https://doi.org/10.1080/02699931.2011.602047

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., Iraki, F. K., Rockstuhl, T., & Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizationscross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, *4*, 1–8. https://doi.org/10.3389/fpsyg.2013.00353

Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, *31*, 182–191. https://doi.org/10.1007/s11031-007-9063-z

Laukka, P., Juslin, P. P. N. P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, *19*(5), 633–653. https://doi.org/10.1080/02699930441000445

Laukkanen, A. M., Vilkman, E., Alku, P., & Oksanen, H. (1997). On the perception of emotions in speech: The role of voice quality. *Logopedics Phoniatrics Vocology*, *22*(4), 157–168. https://doi.org/10.3109/14015439709075330

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64. https://doi.org/10.1016/j.heares.2018.09.005

Lee, K. H., & Siegle, G. J. (2012). Common and distinct brain networks underlying explicit emotional evaluation: A meta-analytic study. *Social Cognitive and Affective Neuroscience*, *7*(5), 521–534. https://doi.org/10.1093/scan/nsp001

Lee, Y. S., Min, N. E., Wingfield, A., Grossman, M., & Peelle, J. E. (2016). Acoustic richness modulates the neural networks supporting intelligible speech processing. *Hearing Research*, *333*, 108–117. https://doi.org/10.1016/j.heares.2015.12.008

Leff, D. R., Orihuela-Espina, F., Elwell, C. E., Athanasiou, T., Delpy, D. T., Darzi, A. W., & Yang, G. Z. (2011). Assessment of the cerebral cortex during motor task behaviours in adults: A systematic review of functional near infrared spectroscopy (fNIRS) studies. *NeuroImage*, *54*(4), 2922–2936. https://doi.org/10.1016/j.neuroimage.2010.10.058

Leitman, D. I. D. I., Wolf, D. H. D. H., Ragland, J. D. D., Laukka, P., Loughead, J., Valdez, J. N. J. N., Javitt, D. C. D. C., Turetsky, B. I., & Gur, R. C. (2010). "It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *4*, 19. https://doi.org/10.3389/fnhum.2010.00019

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means* (1.4.2).

Lesica, N. A. (2018). Why do hearing aids fail to restore normal auditory perception? *Trends in Neurosciences*, *41*(4), 174–185. https://doi.org/10.1016/j.tins.2018.01.008

Liikkanen, L. A., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S., & May, P. J. C. (2007). The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds. *NeuroReport*, *18*(6), 601–605. https://doi.org/10.1097/WNR.0b013e3280b07bde

Lin, F. R., Niparko, J. K., & Ferrucci, L. (2011). Hearing loss prevalence in the United States. *Archives of Internal Medicine*, *171*(20), 1851. https://doi.org/10.1001/archinternmed.2011.506

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *36*(6), 769–777. https://doi.org/10.1007/s10803-006-0105-2

Liu, E. Y., Haist, F., Dubowitz, D. J., & Buxton, R. (2019). Cerebral blood volume changes during the BOLD post-stimulus undershoot measured with a combined normoxia/hyperoxia method. *NeuroImage*, *185*, 154–163. https://doi.org/10.1016/j.neuroimage.2018.10.032

Liu, H. L., Pu, Y., Nickerson, L. D., Liu, Y., Fox, P. T., & Gao, J. H. (2000). Comparison of the temporal response in perfusion and BOLD-based event- related functional MRI. *Magnetic Resonance in Medicine*, *43*(5), 768–772. https://doi.org/10.1002/(SICI)1522-2594(200005)43:5<768::AID-MRM22>3.0.CO;2-8

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 224–234. https://doi.org/10.1037/a0017609

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, *8*(2), 025008. https://doi.org/10.1117/1.nph.8.2.025008

Luo, X. (2016). Talker variability effects on vocal emotion recognition in acoustic and simulated electric hearing. *The Journal of the Acoustical Society of America*, *140*(6), EL497–EL503. https://doi.org/10.1121/1.4971758

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, *11*(4), 301–315. https://doi.org/10.1177/1084713807305301

Luo, X., Kern, A., & Pulling, K. R. (2018). Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. *The Journal of the Acoustical Society of America*, *144*(5), EL429–EL435. https://doi.org/10.1121/1.5079575

Maggioni, E., Molteni, E., Zucca, C., Reni, G., Cerutti, S., Triulzi, F. M., Arrigoni, F., & Bianchi, A. M. (2015). Investigation of negative BOLD responses in human brain through NIRS technique. A visual stimulation study. *NeuroImage*, *108*, 410–422. https://doi.org/10.1016/j.neuroimage.2014.12.074

Marcar, V. L., Strässle, A. E., Loenneker, T., Schwarz, U., & Martin, E. (2004). The influence of cortical maturation on the BOLD response: An fMRI study of visual cortex in children. *Pediatric Research*, *56*(6), 967–974. https://doi.org/10.1203/01.PDR.0000145296.24669.A5

Martinez, M., Multani, N., Anor, C. J., Misquitta, K., Tang-Wai, D. F., Keren, R., Fox, S., Lang, A. E., Marras, C., & Tartaglia, M. C. (2018). Emotion detection deficits and decreased empathy in patients with Alzheimer's disease and Parkinson's disease affect caregiver mood and burden. *Frontiers in Aging Neuroscience*, *10*, 1–9. https://doi.org/10.3389/fnagi.2018.00120

Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., & Deguine, O. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. *Ear and Hearing*, *36*(2), 239–248. https://doi.org/10.1097/AUD.0000000000000105

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's syndrome and high-functioning autism: The importance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders*, *37*(6), 1086–1095. https://doi.org/10.1007/s10803-006-0251-6

McClure, E. B., & Nowicki, S. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior*, *25*, 3–19. https://doi.org/10.1023/A:1006753006870

McGettigan, C., Rosen, S., & Scott, S. K. (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00018

Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M. (2011). The perception of sentence stress in cochlear implant recipients. *Ear and Hearing*, *32*(4), 459–467. https://doi.org/10.1097/AUD.0b013e3182064882

Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, T. L., & Mills, J. T. (1981). Perceived loudness and visually-determined auditory distance. *Perception*, *10*(5), 531–543. https://doi.org/10.1068/p100531

Metcalfe, T. (2017). *Perception of speech, music and emotion by hearing-impaired listeners* [Doctoral thesis, University of Sheffield]. https://etheses.whiterose.ac.uk/19151/

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & Von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, *89*(2), 277–289. https://doi.org/10.1016/S0093-934X(03)00350-X

Miettinen, I., Alku, P., Salminen, N., May, P. J. C., & Tiitinen, H. (2011). Responsiveness of the human auditory cortex to degraded speech sounds: Reduction of amplitude resolution vs. additive noise. *Brain Research*, *1367*, 298–309. https://doi.org/10.1016/j.brainres.2010.10.037

Miettinen, I., Alku, P., Yrttiaho, S., May, P. J. C., & Tiitinen, H. (2012). Cortical processing of degraded speech sounds: Effects of distortion type and continuity. *NeuroImage*, *60*(2), 1036–1045. https://doi.org/10.1016/j.neuroimage.2012.01.085

Mohammadi-Nejad, A.-R., Mahmoudzadeh, M., Hassanpour, M. S., Wallois, F., Muzik, O., Papadelis, C., Hansen, A., Soltanian-Zadeh, H., Gelovani, J., & Nasiriavanaki, M. (2018). Neonatal brain resting-state functional connectivity imaging modalities. *Photoacoustics*, *10*, 1–19. https://doi.org/10.1016/j.pacs.2018.01.003

Momm, T., Blickle, G., Liu, Y., Wihler, A., Kholin, M., & Menges, J. I. (2015). It pays to have an eye for emotions: Emotion recognition ability indirectly predicts annual income. *Journal of Organizational Behavior*, *36*(1), 147–163. https://doi.org/10.1002/job.1975

Moore, B. (1995). *Perceptual consequences of cochlear damage*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198523307.001.0001

Morise, M. (2015). CheapTrick, a spectral envelope estimator for high-quality speech synthesis. *Speech Communication*, *67*, 1–7. https://doi.org/10.1016/j.specom.2014.09.003

Morise, M. (2016). D4C, a band-aperiodicity estimator for high-quality speech synthesis. *Speech Communication*, *84*, 57–65. https://doi.org/10.1016/j.specom.2016.09.001

Morise, M., Yokomori, F., & Ozawa, K. (2016). WORLD: A vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Transactions on Information and Systems*, *E99.D*(7), 1877–1884. https://doi.org/10.1587/transinf.2015EDP7457

Morningstar, M., Nelson, E. E., & Dirks, M. A. (2018). Maturation of vocal emotion recognition: Insights from the developmental and neuroimaging literature. *Neuroscience and Biobehavioral Reviews*, *90*, 221–230. https://doi.org/10.1016/j.neubiorev.2018.04.019

Most, T., & Aviner, C. (2009). Auditory, visual, and auditory - Visual perception of emotions by individuals with cochlear implants, hearing aids, and normal hearing. *Journal of Deaf Studies and Deaf Education*, *14*(4), 449–464. https://doi.org/10.1093/deafed/enp007

Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *Journal of Neuroscience*, *31*(26), 9594–9598. https://doi.org/10.1523/JNEUROSCI.6665-10.2011

Mozziconacci, S. (1998). *Speech variability and emotion: Production and perception* [Doctoral thesis, Technische Universiteit Eindhoven]. https://doi.org/10.6100/IR516785

Mozziconacci, S., & Hermes, D. J. (2000). Expression of emotion and attitude through temporal speech variations. *6th International Conference on Spoken Language Processing*.

Muges. (2017). *AudioTSM*. https://github.com/Muges/audiotsm/

Muirhead, H., & Perutz, M. F. (1963). Structure of reduced human hemoglobin. *Cold Spring Harbor Symposia on Quantitative Biology*, 451–459.

Mullinger, K. J., Mayhew, S. D., Bagshaw, A. P., Bowtell, R., & Francis, S. T. (2014). Evidence that the negative BOLD response is neuronal in origin: A simultaneous EEG-BOLD-CBF study in humans. *NeuroImage*, *94*, 263–274. https://doi.org/10.1016/j.neuroimage.2014.02.029

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, *93*(2), 1097–1108. https://doi.org/10.1121/1.405558

Mushtaq, F., Wiggins, I. M., Kitterick, P. T., Anderson, C. A., & Hartley, D. E. H. (2019). Evaluating time-reversed speech and signal-correlated noise as auditory baselines for isolating speech-specific processing using fNIRS. *PLoS ONE*, *14*(7), e0219927. https://doi.org/10.1371/journal.pone.0219927

Nagels, L., Gaudrain, E., Vickers, D., Matos Lopes, M., Hendriks, P., & Başkent, D. (2020). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ*, *8*, e8773. https://doi.org/10.7717/peerj.8773

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America*, *131*(2), 1307–1314. https://doi.org/10.1121/1.3672697

Netten, A. P., Rieffe, C., Theunissen, S. C. P. M., Soede, W., Dirks, E., Briaire, J. J., & Frijns, J. H. M. (2015). Low empathy in deaf and hard of hearing (pre)adolescents compared to normal hearing controls. *PLoS ONE*, *10*(4), e0124102. https://doi.org/10.1371/journal.pone.0124102

NeuroBehavioral Systems Inc. (2020). *Presentation* (20.2).

Niebuhr, O., & Pfitzinger, H. R. (2010). On pitch-accent identification – The role of syllable duration and intensity. In *Speech Prosody-5th International Conference* (pp. 1–4).

O'Connell, K., Marsh, A. A., Edwards, D. F., Dromerick, A. W., & Seydell-Greenwald, A. (2021). Emotion recognition impairments and social well-being following right-hemisphere stroke. *Neuropsychological Rehabilitation*. https://doi.org/10.1080/09602011.2021.1888756

Ohnishi, M., Kusakawa, N., Masaki, S., Honda, K., Hayashi, N., Shimada, Y., Fujimoto, I., & Hirao, and K. (1997). Measurement of hemodynamics of auditory cortex using magnetoencephalography and near infrared spectroscopy. *Acta Oto-Laryngologica*, *117*, 129–131. https://doi.org/10.3109/00016489709126161

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H. H., Saberi, K., Serences, J. T., & Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495. https://doi.org/10.1093/cercor/bhp318

Olds, C., Pollonini, L., Abaya, H., Larky, J., Loy, M., Bortfeld, H., Beauchamp, M. S., Oghalai, J. S., Beauschamp, M., & Oghalai, J. S. (2016). Cortical activation patterns correlate with speech understanding after cochlear implantation. In *Ear and Hearing* (Vol. 37, Issue 3). https://doi.org/10.1097/AUD.0000000000000258

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Orihuela-Espina, F., Leff, D. R., James, D. R. C., Darzi, A. W., & Yang, G. Z. (2010). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Physics in Medicine and Biology*, *55*(13), 3701–3724. https://doi.org/10.1088/0031-9155/55/13/009

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. https://doi.org/https://doi.org/10.1037/0033-295X.97.3.315

Pak, C. L., & Katz, W. F. (2019). Recognition of emotional prosody by Mandarin-speaking adults with cochlear implants. *The Journal of the Acoustical Society of America*, *146*(2), EL165–EL171. https://doi.org/10.1121/1.5122192

Pakosz, M. (1983). Attitudinal judgments in intonation: Some evidence for a theory. *Journal of Psycholinguistic Research*, *12*(3), 311–326. https://link.springer.com/content/pdf/10.1007/BF01067673.pdf

Panzeri, F., Cavicchiolo, S., Giustolisi, B., Di Berardino, F., Ajmone, P. F., Vizziello, P., Donnini, V., & Zanetti, D. (2021). Irony comprehension in children with cochlear implants: The role of language competence, theory of mind, and prosody recognition. *Journal of Speech, Language, and Hearing Research*, 1–18. https://doi.org/10.1044/2021_JSLHR-20-00671

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*(2), 209–213. https://doi.org/10.1097/WNR.0b013e3282f454db

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, *104*(3), 262–269. https://doi.org/10.1016/j.bandl.2007.03.002

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, *28*(2), 230–244. https://doi.org/10.1080/02699931.2013.812033

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, *30*(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Peelle, J. E. (2017). Optical neuroimaging of spoken language. *Language, Cognition and Neuroscience*, *32*(7), 847–854. https://doi.org/10.1080/23273798.2017.1290810

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, *36*(8), 701–715. https://doi.org/10.1016/S0028-3932(98)00008-6

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435. https://doi.org/10.1016/j.wocn.2009.07.005

Peng, Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends in Amplification*, *16*(2), 67–82. https://doi.org/10.1177/1084713812451159

Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology*, *14*(5), 327–337. https://doi.org/10.1159/000212112

Pereira, C. (2000). Dimensions of emotional meaning in speech. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, *September*, 93–96. https://www.isca-speech.org/archive_open/speech_emotion/spem_025.html

Péron, J., Dondaine, T., Le Jeune, F., Grandjean, D., & Vérin, M. (2012). Emotional processing in parkinson's disease: A systematic review. *Movement Disorders*, *27*(2), 186–199. https://doi.org/10.1002/mds.24025

Peters, K. P. (2006). *Emotion perception in speech: Discrimination, identification, and the effects of talker and sentence varability* [Masters thesis, Washington University]. https://digitalcommons.wustl.edu/cgi/viewcontent.cgi?article=1112&context=pacs_capstones

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., Atkinson, A. P., Calder, A. J., Johnstone, E. C., Lawrie, S. M., & Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine*, *40*(11), 1919–1929. https://doi.org/10.1017/S0033291709992364

Picou, E. M., Singh, G., Goy, H., Russo, F., Hickson, L., Oxenham, A. J., Buono, G. H., Ricketts, T. A., & Launer, S. (2018). Hearing, emotion, amplification, research, and training workshop: Current understanding of hearing loss and emotion perception and priorities for future research. *Trends in Hearing*, *22*, 1–24. https://doi.org/10.1177/2331216518803215

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*, 5–29. https://doi.org/10.1111/nyas.13948

Plack, C. J. (2018). *The sense of hearing* (3rd ed.). Routledge.

Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., & Fallgatter, A. J. (2011). Auditory cortex activation is modulated by emotion: A functional near-infrared spectroscopy (fNIRS) study. *NeuroImage*, *55*(3), 1200–1207. https://doi.org/10.1016/j.neuroimage.2011.01.011

Plichta, M. M., Herrmann, M. J., Ehlis, A. C., Baehne, C. G., Richter, M. M., & Fallgatter, A. J. (2006). Event-related visual versus blocked motor task: Detection of specific cortical activation patterns with functional near-infrared spectroscopy. *Neuropsychobiology*, *53*(2), 77–82. https://doi.org/10.1159/000091723

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350. https://www.jstor.org/stable/27857503

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Pollermann, B. Z., & Archinard, M. (2002). Acoustic patterns of emotions. In & M. H. E. Keller. G. Bailly, A. Monaghan, J. Terken (Ed.), *Improvements in speech synthesis*. J. Wiley. https://doi.org/10.1002/0470845945.ch23

Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., & Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hearing Research*, *309*, 84–93. https://doi.org/10.1016/j.heares.2013.11.007

Powell, L. J., Deen, B., & Saxe, R. (2018). Using individual functional channels of interest to study cortical development with fNIRS. *Developmental Science*, *21*(4), e12595. https://doi.org/10.1111/desc.12595

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Pujol, J., Deus, J., Losilla, J. M., & Capdevila, A. (1999). Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology*, *52*(5), 1038–1038. https://doi.org/10.1212/WNL.52.5.1038

Quadflieg, S., Mohr, A., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. *Biological Psychology*, *78*(2), 129–137. https://doi.org/10.1016/j.biopsycho.2008.01.014

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*(2), 79–89. https://doi.org/10.1016/j.bandl.2011.03.009

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods*, *22*(1), 46–68. https://doi.org/10.1177/1094428116658959

R Core Team. (2020). *R: A language and environment for statistical computing* (3.6.3). R Foundation for Statistical Computing. http://www.r-project.org/

Rahimpour, A., Pollonini, L., Comstock, D., Balasubramaniam, R., & Bortfeld, H. (2020). Tracking differential activation of primary and supplementary motor cortex across timing tasks: An fNIRS validation study. *Journal of Neuroscience Methods*, *341*(May), 108790. https://doi.org/10.1016/j.jneumeth.2020.108790

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, *4*, 5–8. https://doi.org/10.3389/fnevo.2012.00007

Remijn, G. B., & Kojima, H. (2010). Active versus passive listening to auditory streaming stimuli: a near-infrared spectroscopy study. *Journal of Biomedical Optics*, *15*(3), 037006. https://doi.org/10.1117/1.3431104

Ren, L., Zhang, Y., Zhang, J., Qin, Y., Zhang, Z., Chen, Z., Wei, C., & Liu, Y. (2021). Voice emotion recognition by Mandarin-speaking children with cochlear implants. *Ear and Hearing*, 1–16. https://doi.org/10.1097/AUD.0000000000001085

Rieffe, C., & Terwogt, M. M. (2000). Deaf children's understanding of emotions: Desires take precedence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 601–608. https://doi.org/10.1017/S0021963099005843

Rilliard, A., d'Alessandro, C., & Evrard, M. (2018). Paradigmatic variation of vowels in expressive speech: Acoustic description and dimensional analysis. *The Journal of the Acoustical Society of America*, *143*(1), 109–122. https://doi.org/10.1121/1.5018433

Ritter, C., & Vongpaisal, T. (2018). Multimodal and spectral degradation effects on speech and emotion recognition in adult listeners. *Trends in Hearing*, *22*, 1–17. https://doi.org/10.1177/2331216518804966

Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition*, *69*(2), 269–278. https://doi.org/10.1016/j.bandc.2008.08.007

Robinson, C., Obin, N., & Roebel, A. (2019). Sequence-to-sequence modelling of F0 for speech emotion conversion. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 6830–6834. https://doi.org/10.1109/ICASSP.2019.8683865

Rodero, E. (2011). Intonation and emotion: Influence of pitch levels and contour type on creating emotions. *Journal of Voice*, *25*(1), e25–e34. https://doi.org/10.1016/j.jvoice.2010.02.002

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075

Rosemann, S., Gießing, C., Özyurt, J., Carroll, R., Puschmann, S., & Thiel, C. M. (2017). The contribution of cognitive factors to individual differences in understanding noise-vocoded speech in young and older adults. *Frontiers in Human Neuroscience*, *11*, 1–13. https://doi.org/10.3389/fnhum.2017.00294

Rosen, S., & Fourcin, A. (1986). Frequency selectivity and the perception of speech. In B. Moore (Ed.), *Frequency selectivity in hearing*. Academic Press. http://www.phon.ucl.ac.uk/home/stuart/pubs/RosenFourcin1986i.pdf

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, Inc. http://www.rstudio.com/

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294. https://doi.org/10.1016/0092-6566(77)90037-X

Saager, R. B., & Berger, A. J. (2005). Direct characterization and removal of interfering absorption trends in two-layer turbid media. *Journal of the Optical Society of America A*, *22*(9), 1874–1882. https://doi.org/10.1364/JOSAA.22.001874

Saager, R., & Berger, A. (2008). Measurement of layer-like hemodynamic trends in scalp and cortex: implications for physiological baseline suppression in functional near-infrared spectroscopy. *Journal of Biomedical Optics*, *13*(3), 034017. https://doi.org/10.1117/1.2940587

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *NeuroImage*, *39*(3), 1444–1456. https://doi.org/10.1016/j.neuroimage.2007.09.052

Sachs, M. E., Habibi, A., Damasio, A., & Kaplan, J. T. (2018). Decoding the neural signatures of emotions expressed through sound. *NeuroImage*, *174*, 1–10. https://doi.org/10.1016/j.neuroimage.2018.02.058

Saliba, J., Bortfeld, H., Levitin, D. J., & Oghalai, J. S. (2016). Functional near-infrared spectroscopy for neuroimaging in cochlear implant recipients. *Hearing Research*, *338*, 64–75. https://doi.org/10.1016/j.heares.2016.02.005

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. J. (2018). The NIRS Brain AnalyzIR Toolbox. *Algorithms*, *11*(5), 73. https://doi.org/10.3390/a11050073

Sato, H., Takeuchi, T., & Sakai, K. L. (1999). Temporal cortex activation during speech recognition: an optical topography study. *Cognition*, *73*(3), B55–B66. https://doi.org/10.1016/S0010-0277(99)00060-8

Sato, T., Ito, M., Suto, T., Kameyama, M., Suda, M., Yamagishi, Y., Ohshima, A., Uehara, T., Fukuda, M., & Mikuni, M. (2007). Time courses of brain activation and their implications for function: A multichannel near-infrared spectroscopy study during finger tapping. *Neuroscience Research*, *58*(3), 297–304. https://doi.org/10.1016/j.neures.2007.03.014

Sauter, D. A. (2013). The role of motivation and cultural dialects in the in-group advantage for emotional vocalizations. *Frontiers in Psychology*, *4*, 1–9. https://doi.org/10.3389/fpsyg.2013.00814

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92. https://doi.org/10.1177/0022022101032001009

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, *15*(2), 123–148. https://doi.org/10.1007/BF00995674

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*, 331–346. https://doi.org/10.1007/BF00992539

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition index. *Journal of Nonverbal Behavior*, *35*, 305–326. https://doi.org/10.1007/s10919-011-0115-4

Schmid Mast, M., & Hall, J. A. (2018). The impact of interpersonal accuracy on behavioral outcomes. *Current Directions in Psychological Science*, *27*(5), 309–314. https://doi.org/10.1177/0963721418758437

Schmidt, J., Janse, E., & Scharenborg, O. (2016). Perception of emotion in conversational speech by younger and older listeners. *Frontiers in Psychology*, *7*, 1–11. https://doi.org/10.3389/fpsyg.2016.00781

Scholkmann, F., Klein, S. D., Gerber, U., Wolf, M., & Wolf, U. (2014). Cerebral hemodynamic and oxygenation changes induced by inner and heard speech: a study combining functional near-infrared spectroscopy and capnography. *Journal of Biomedical Optics*, *19*(1), 017002. https://doi.org/10.1117/1.jbo.19.1.017002

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85*, 6–27. https://doi.org/10.1016/j.neuroimage.2013.05.004

Scholkmann, F., Spichtig, S., Muehlemann, T., & Wolf, M. (2010). How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiological Measurement*, *31*(5), 649–662. https://doi.org/10.1088/0967-3334/31/5/004

Scholkmann, F., & Wolf, M. (2013). General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. *Journal of Biomedical Optics*, *18*(10), 105004. https://doi.org/10.1117/1.jbo.18.10.105004

Schorr, E. A., Roth, F. P., & Fox, N. A. (2009). Quality of life for children with cochlear implants: Perceived benefits and problems and the perception of single words and emotional sounds. *Journal of Speech, Language, and Hearing Research*, *52*(1), 141–152. https://doi.org/10.1044/1092-4388(2008/07-0213)

Schwartze, M., & Kotz, S. A. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2587–2596. https://doi.org/10.1016/j.neubiorev.2013.08.005

Scott, S. K., & McGettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? *Brain and Language*, *127*(1), 36–45. https://doi.org/10.1016/j.bandl.2013.07.006

Sevy, A. B. G., Bortfeld, H., Huppert, T. J., Beauchamp, M. S., Tonini, R. E., & Oghalai, J. S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hearing Research*, *270*(1–2), 39–47. https://doi.org/10.1016/j.heares.2010.09.010

Seydell-Greenwald, A., Chambers, C. E., Ferrara, K., & Newport, E. L. (2020). What you say versus how you say it: Comparing sentence comprehension and emotional prosody processing using fMRI. *NeuroImage*, *209*, 116509. https://doi.org/10.1016/j.neuroimage.2019.116509

Shader, M. J., Luke, R., Gouailhardou, N., & Mckay, C. M. (2021). The use of broad vs restricted regions of interest in functional near-infrared spectroscopy for measuring cortical activation to auditory-only and visual-only speech. *Hearing Research*, *406*, 108256. https://doi.org/10.1016/j.heares.2021.108256

Shannon, R. V, Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporary cues. *Science*, *270*, 303–304. https://doi.org/10.1086/275028

Shimokawa, A., Yatomi, N., Anamizu, S., Torii, S., Isono, H., Sugai, Y., & Kohno, M. (2001). Influence of deteriorating ability of emotional comprehension on interpersonal behavior in Alzheimer-type dementia. *Brain and Cognition*, *47*(3), 423–433. https://doi.org/10.1006/brcg.2001.1318

Simcock, G., McLoughlin, L. T., De Regt, T., Broadhouse, K. M., Beaudequin, D., Lagopoulos, J., & Hermens, D. F. (2020). Associations between facial emotion recognition and mental health in early adolescence. *International Journal of Environmental Research and Public Health*, *17*(1). https://doi.org/10.3390/ijerph17010330

Simpson, C., Pinkham, A. E., Kelsven, S., & Sasson, N. J. (2013). Emotion recognition abilities across stimulus modalities in schizophrenia and the role of visual attention. *Schizophrenia Research*, *151*(1–3), 102–106. https://doi.org/10.1016/j.schres.2013.09.026

Singh, A. K., & Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *NeuroImage*, *33*(2), 542–549. https://doi.org/10.1016/j.neuroimage.2006.06.047

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*(5), 107. https://doi.org/10.4103/0972-2327.128676

Smith, C. A., Ellsworth, P. C., & Hall, J. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Sonkaya, A. R., & Bayazıt, Z. Z. (2018). A neurolinguistic investigation of emotional prosody and verbal components of speech. *NeuroQuantology*, *16*(12), 50–56. https://doi.org/10.14704/nq.2018.16.12.1877

Steber, S., König, N., Stephan, F., & Rossi, S. (2020). Uncovering electrophysiological and vascular signatures of implicit emotional prosody. *Scientific Reports*, *10*(1), 5807. https://doi.org/10.1038/s41598-020-62761-x

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., & Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI-fNIRS studies. *Magnetic Resonance Imaging*, *24*(4), 495–505. https://doi.org/10.1016/j.mri.2005.12.034

Strangman, G. E., Franceschini, M. A., & Boas, D. A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *NeuroImage*, *18*(4), 865–879. https://doi.org/10.1016/S1053-8119(03)00021-1

Strangman, G. E., Li, Z., & Zhang, Q. (2013). Depth sensitivity and source-detector separations for near infrared spectroscopy based on the Colin27 brain template. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0066319

Tachtsidis, I., Tisdall, M. M., Leung, T. S., Pritchard, C., Cooper, C. E., Smith, M., & E., C. E. (2009). Relationship between brain tissue haemodynamics, oxygenation and metabolism in the healthy human adult brain during hyperoxia and hypercapnea. In P. Liss, P. Hansell, D. F. Bruley, & D. Harrison (Eds.), *Oxygen transport to tissue XXX*. Springer. https://doi.org/10.1007/978-0-387-85998-9_47

Tachtsidis, I., & Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, *3*(3), 031405. https://doi.org/10.1117/1.nph.3.3.031405

Tak, S., & Ye, J. C. (2014). Statistical analysis of fNIRS data: A comprehensive review. *NeuroImage*, *85*, 72–91. https://doi.org/10.1016/j.neuroimage.2013.06.016

Terhardt, E. (1979). Calculating virtual pitch. *Carpathian Mathematical Publications*, *1*(2), 155–182. https://doi.org/https://doi.org/10.1016/0378-5955(79)90025-X

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*(2006), 407–424. https://doi.org/10.1515/SEM.2006.017

Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, *2*(3), 225–234. https://doi.org/10.1038/s41562-018-0305-8

Tinnemore, A. R., Zion, D. J., Kulkarni, A. M., & Chatterjee, M. (2018). Children's recognition of emotional prosody in spectrally degraded speech is predicted by their age and cognitive status. *Ear and Hearing*, *39*(5), 874–880. https://doi.org/10.1097/AUD.0000000000000546

Tobe, R. H., Corcoran, C. M., Breland, M., MacKay-Brandt, A., Klim, C., Colcombe, S. J., Leventhal, B. L., & Javitt, D. C. (2016). Differential profiles in auditory social cognition deficits between adults with autism and schizophrenia spectrum disorders: A preliminary analysis. *Journal of Psychiatric Research*, *79*, 21–27. https://doi.org/10.1016/j.jpsychires.2016.04.005

Toronov, V. Y., Zhang, X., & Webb, A. G. (2007). A spatial and temporal comparison of hemodynamic signals measured using optical and functional magnetic resonance imaging during activation in the human primary visual cortex. *NeuroImage*, *34*(3), 1136–1148. https://doi.org/10.1016/j.neuroimage.2006.08.048

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Undurraga, J. A., Van Yper, L., Bance, M., McAlpine, D., & Vickers, D. (2021). Characterizing cochlear implant artefact removal from EEG recordings using a real human model. *MethodsX*, *8*, 101369. https://doi.org/10.1016/j.mex.2021.101369

Vaden, K. I., Teubner-Rhodes, S., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2017). Cingulo-opercular activity affects incidental memory encoding for speech in noise. *NeuroImage*, *157*, 381–387. https://doi.org/10.1016/j.neuroimage.2017.06.028

Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology*, *14*(4), 387–406. https://doi.org/10.1177/0022002183014004001

van de Rijt, L. P. H., van Wanrooij, M. M., Snik, A. F. M., Mylanus, E. A. M., van Opstal, A. J., & Roye, A. (2018). Measuring cortical activity during auditory processing with functional near-infrared spectroscopy. *Journal of Hearing Science*, *8*(4), 9–18. https://doi.org/10.17430/1003278

Van de Velde, D. J. (2017). *The processing of Dutch prosody with cochlear implants and vocoder simulations* [Doctoral thesis, University of Leiden]. https://scholarlypublications.universiteitleiden.nl/handle/1887/50406

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, *35*(5), 963–970. https://doi.org/10.1044/jshr.3505.963

Velmurugan, J., Sinha, S., & Satishchandra, P. (2014). Magnetoencephalography recording and analysis. *Annals of Indian Academy of Neurology*, *17*(5), 113. https://doi.org/10.4103/0972-2327.128678

Verhelst, W., & Roelands, M. (1993). Overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*, *2*, 554–557. https://doi.org/10.1109/icassp.1993.319366

Vicenik, C. (n.d.-a). *Intensity-neutralizer*. http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html

Vicenik, C. (n.d.-b). *Intensity-scaler*. http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html

von Cramon, D. Y. Y., Kotz, S. A., Besson, M., Meyer, M., Friederici, A. D., Alter, K., Besson, M., von Cramon, D. Y. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*(3), 366–376. https://doi.org/10.1016/s0093-934x(02)00532-1

Vul, E., & Kanwisher, N. (2013). Begging the question: The nonindependence error in fMRI data analysis. In S. J. Hanson & M. Bunzl (Eds.), *Foundational Issues in Human Brain Mapping* (pp. 71–91). The MIT Press. https://doi.org/10.7551/mitpress/9780262014021.003.0007

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885. https://doi.org/10.1162/jocn.2009.21366

Waaramaa, T., Kukkonen, T., Stoltz, M., & Geneid, A. (2018). Hearing impairment and emotion identification from auditory and visual stimuli. *International Journal of Listening*, *32*(3), 150–162. https://doi.org/10.1080/10904018.2016.1250633

Waaramaa, T., Laukkanen, A. M., Airas, M., & Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. *Journal of Voice*, *24*(1), 30–38. https://doi.org/10.1016/j.jvoice.2008.04.004

Waaramaa, T., & Leisiö, T. (2013). Perception of emotionally loaded vocal expressions and its connection to responses to music. A cross-cultural investigation: Estonia, Finland, Sweden, Russia, and the USA. *Frontiers in Psychology*, *4*, 1–13. https://doi.org/10.3389/fpsyg.2013.00344

Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., Baker, J. T., Stufflebeam, S. M., Wang, K., Wang, X., Hong, B., & Liu, H. (2015). Parcellating cortical functional networks in individuals. *Nature Neuroscience*, *18*(12), 1853–1860. https://doi.org/10.1038/nn.4164

Wang, F., Mao, M., Duan, L., Huang, Y., Li, Z., & Zhu, C. (2018). Intersession instability in fNIRS-based emotion recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(7), 1324–1333. https://doi.org/10.1109/TNSRE.2018.2842464

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., Rosen, S., & Scott, S. K. (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *Journal of Neuroscience*, *26*(50), 13067–13075. https://doi.org/10.1523/JNEUROSCI.3907-06.2006

Warren, P. (1999). Prosody and language processing. In S. Garrod & P. Martin (Eds.), *Language processing*. Psychology Press.

Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A. D., & Obrig, H. (2007). The processing of prosody: Evidence of interhemispheric specialization at the age of four. *NeuroImage*, *34*(1), 416–425. https://doi.org/10.1016/j.neuroimage.2006.09.009

Watanabe, H., Homae, F., & Taga, G. (2012). Activation and deactivation in response to visual stimulation in the occipital cortex of 6-month-old human infants. *Developmental Psychobiology*, *54*(1), 1–15. https://doi.org/10.1002/dev.20569

Watson, D., & Tellegen, A. (1999). Issues in dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, *125*(5), 601–610. https://doi.org/10.1037/0033-2909.125.5.601

Weder, S., Shoushtarian, M., Olivares, V., Zhou, X., Innes-brown, H., & McKay, C. (2020). Cortical fNIRS responses can be better explained by loudness percept than sound intensity. *Ear and Hearing*, *41*(5), 1187–1195. https://doi.org/10.1097/AUD.0000000000000836

Weder, S., Zhou, X., Shoushtarian, M., Innes-brown, H., McKay, C., Olivares, V., Zhou, X., Innes-brown, H., & McKay, C. (2018). Cortical processing related to intensity of a modulated noise stimulus—a functional near-infrared study. *Journal of the Association for Research in Otolaryngology*, *19*(3), 273–286. https://doi.org/10.1007/s10162-018-0661-0

Westgarth, M. M. P., Hogan, C. A., Neumann, D. L., & Shum, D. H. K. (2021). A systematic review of studies that used NIRS to measure neural activation during emotion processing in healthy individuals. *Social Cognitive and Affective Neuroscience*, *16*(4), 345–369. https://doi.org/10.1093/scan/nsab017

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, *93*(4), 2152–2159. https://doi.org/10.1121/1.406678

Wiefferink, C. H., Rieffe, C., Ketelaar, L., & Frijns, J. H. M. (2012). Predicting social functioning in children with a cochlear implant and in normal-hearing children: The role of emotion regulation. *International Journal of Pediatric Otorhinolaryngology*, *76*(6), 883–889. https://doi.org/10.1016/j.ijporl.2012.02.065

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody-influence of acoustic parameters and arousal. *NeuroImage*, *39*(2), 885–893. https://doi.org/10.1016/j.neuroimage.2007.09.028

Wiggins, I. M., Anderson, C. A., Kitterick, P. T., & Hartley, D. E. H. (2016). Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? *Hearing Research*, *339*, 142–154. https://doi.org/10.1016/j.heares.2016.07.007

Wijayasiri, P., Hartley, D. E. H., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, *351*, 55–67. https://doi.org/10.1016/j.heares.2017.05.010

Wijeakumar, S., Shahani, U., Simpson, W. A., & McCulloch, D. L. (2012). Localization of hemodynamic responses to simple visual stimulation: An fNIRS study. *Investigative Ophthalmology and Visual Science*, *53*(4), 2266–2273. https://doi.org/10.1167/iovs.11-8680

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, *60*(2), 1490–1502. https://doi.org/10.1016/j.neuroimage.2012.01.035

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage*, *15*(4), 856–869. https://doi.org/10.1006/nimg.2001.0998

Wildgruber, Dirk, Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277–281. https://doi.org/10.1080/17549500902943043

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research*, *242*(1–2), 3–21. https://doi.org/10.1016/j.heares.2008.06.005

Wilson, B. S., & Tucci, D. L. (2021). Addressing the global burden of hearing loss. *The Lancet*, *397*(10278), 945–947. https://doi.org/10.1016/S0140-6736(21)00522-5

Wilting, J., Krahmer, E., & Swerts, M. (2006). Real vs. acted emotional speech. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2*, 805–808.
Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, *131*(2), 1465–1479. https://doi.org/10.1121/1.3672705
Witteman, J., Van Heuven, V. J. P., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(12), 2752–2763. https://doi.org/10.1016/j.neuropsychologia.2012.07.026
Wolf, M., Wolf, U., Toronov, V., Michalos, A., Paunescu, L. A., Choi, J. H., & Gratton, E. (2002). Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: A near-infrared spectroscopy study. *NeuroImage*, *16*(3), 704–712. https://doi.org/10.1006/nimg.2002.1128

Wöstmann, M., & Obleser, J. (2016). Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. *Frontiers in Human Neuroscience*, *10*, 1–9. https://doi.org/10.3389/fnhum.2016.00538
Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2013). Voice quality in affect cueing: does loudness matter? *Frontiers in Psychology*, *4*, 1–14. https://doi.org/10.3389/fpsyg.2013.00335
Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*, *44*(2), 428–447. https://doi.org/10.1016/j.neuroimage.2008.08.036
Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., & Narayanan, S. (2004). An acoustic study of emotions expressed in speech. *8th International Conference on Spoken Language Processing*, 2193–2196.
Yousry, T. A., Schmid, U. D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., & Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain*, *120*(1), 141–157. https://doi.org/10.1093/brain/120.1.141
Yücel, M. A., Lühmann, A. v., Scholkmann, F., Gervain, J., Dan, I., Ayaz, H., Boas, D., Cooper, R. J., Culver, J., Elwell, C. E., Eggebrecht, A., Franceschini, M. A., Grova, C., Homae, F., Lesage, F., Obrig, H., Tachtsidis, I., Tak, S., Tong, Y., … Wolf, M. (2021). Best practices for fNIRS publications. *Neurophotonics*, *8*(1), 1–34. https://doi.org/10.1117/1.nph.8.1.012101
Yücel, M. A., Selb, J., Cooper, R. J., & Boas, D. A. (2014). Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, *7*(2), 1–8. https://doi.org/10.1142/S1793545813500661
Yücel, M. A., Selb, J. J., Huppert, T. J., Franceschini, M. A., Boas, D. A., Angela, M., Boas, D. A., Selb, J. J., Franceschini, M. A., Yücel, M. A., & Boas, D. A. (2017). Functional near infrared spectroscopy: Enabling routine functional brain imaging. *Current Opinion in Biomedical Engineering*, *4*, 78–86. https://doi.org/10.1016/j.cobme.2017.09.011
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946–953. https://doi.org/10.1093/cercor/11.10.946
Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836. https://doi.org/10.1016/j.neuroimage.2006.04.199
Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters*, *658*, 62–66. https://doi.org/10.1016/j.neulet.2017.08.047
Zhang, D., Zhou, Y., & Yuan, J. (2018). Speech prosodies of different emotional categories activate different brain regions in adult cortex: An fNIRS study. *Scientific Reports*, *8*, 1–11. https://doi.org/10.1038/s41598-017-18683-2
Zhang, M., Ying, Y. M., & Ihlefeld, A. (2018). Spatial release from informational masking: evidence from functional near infrared spectroscopy. *Trends in Hearing*, *22*, 1–12. https://doi.org/10.1177/2331216518817464
Zhang, Y., Tan, F., Xu, X., Duan, L., Liu, H., Tian, F., & Zhu, C.-Z. (2015). Multiregional functional near-infrared spectroscopy reveals globally symmetrical and frequency-specific patterns of superficial interference. *Biomedical Optics Express*, *6*(8), 2786. https://doi.org/10.1364/boe.6.002786
Zhen, L. E. I., Rong, B. I., Licheng, M. O., Wenwen, Y. U., & Dandan, Z. (2021). The brain mechanism of explicit and implicit processing of affective prosodies : An fNIRS study. *Acta Psychologica Sinica*, *53*(1), 15–25. https://doi.org/https://dx.doi.org/10.3724/SP.J.1041.2021.00015

Zhou, X., Seghouane, A.-K. K., Shah, A., Innes-Brown, H., Cross, W., Litovsky, R., & McKay, C. M. (2018). Cortical speech processing in postlingually deaf adult cochlear implant users, as revealed by functional near-infrared spectroscopy. *Trends in Hearing*, *22*, 1–18. https://doi.org/10.1177/2331216518786850

Zhu, T., Faulkner, S., Madaan, T., Bainbridge, A., Price, D., Thomas, D., Cady, E., Robertson, N., Golay, X., & Tachtsidis, I. (2012). Optimal wavelength combinations for resolving in-vivo changes of haemoglobin and cytochrome-c-oxidase concentrations with NIRS. *Digital Holography and 3-D Imaging, OSA Technical Digest*, JM3A.6. https://doi.org/10.1364/biomed.2012.jm3a.6

Zhu, Z., Miyauchi, R., Araki, Y., & Unoki, M. (2018). Contributions of temporal cue on the perception of speaker individuality and vocal emotion for noise-vocoded speech. *Acoustical Science and Technology*, *39*(3), 234–242. https://doi.org/10.1250/ast.39.234

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). FNIRS Optodes' Location Decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, *8*, 1–11. https://doi.org/10.1038/s41598-018-21716-z

Zinchenko, A., Kanske, P., Obermeier, C., Schröger, E., Villringer, A., & Kotz, S. A. (2018). Modulation of cognitive and emotional control in age-related mild-to-moderate hearing loss. *Frontiers in Neurology*, *9*, 1–16. https://doi.org/10.3389/fneur.2018.00783

Groningen dissertations in linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.
12. Ton van der Wouden (1994). *Negative Contexts*.
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar*.
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar*.
15. Maarten de Wind (1995). *Inversion in French*.
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. H. Wee (1999). *Definite Focus.*
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority substitutions.*
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement*.
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*.
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures*.
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*.
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology*.
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia*.
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision*.
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*

51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin*.
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis*.
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax.*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency*.
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality*.
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships*.
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning*.
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*
88. Jelena Prokić (2010). *Families and Resemblances*.
89. Radek Šimík (2011). *Modal existential wh-constructions*.
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*
92. Aletta Kwant (2011).*Geraakt door prentenboeken*. *Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*

93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Cagri Coltekin (2011).*Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*
99. Herman Heringa (2012). *Appositional constructions.*
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia*.
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*.
104. Gisi Cannizzaro (2012). *Early word order and animacy*.
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model.*
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*
107. Veerle Baaijen (2012). *The development of understanding through writing.*
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*
109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance dependencies.*
110. Hanneke Loerts (2012).*Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender.*
111. Marjoleine Sloos (2013). *Frequency and phonological grammar: An integrated approach. Evidence from German, Indonesian, and Japanese*.
112. Aysa Arylova. (2013) *Possession in the Russian clause. Towards dynamicity in syntax.*
113. Daniël de Kok (2013). Reversible Stochastic Attribute-Value Grammars.
114. Gideon Kotzé (2013). Complementary approaches to tree alignment: Combining statistical and rule-based methods.
115. Fridah Katushemererwe (2013). Computational Morphology and Bantu Language Learning: an Implementation for Runyakitara.
116. Ryan C. Taylor (2013). *Tracking Referents: Markedness, World Knowledge and Pronoun Resolution.*
117. Hana Smiskova-Gustafsson (2013). *Chunks in L2 Development: A Usage-based Perspective.*
118. Milada Walková (2013). *The aspectual function of particles in phrasal verbs.*
119. Tom O. Abuom (2013). *Verb and Word Order Deficits in Swahili-English bilingual agrammatic speakers.*
120. Gülsen Yılmaz (2013). *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands.*
121. Trevor Benjamin (2013). *Signaling Trouble: On the linguistic design of other-initiation of repair in English conversation.*
122. Nguyen Hong Thi Phuong (2013). *A Dynamic Usage-based Approach to Second Language Teaching.*
123. Harm Brouwer (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model.*
124. Kendall Decker (2014). *Orthography Development for Creole Languages.*
125. Laura S. Bos (2015). *The Brain, Verbs, and the Past: Neurolinguistic Studies on Time Reference.*
126. Rimke Groenewold (2015). *Direct and indirect speech in aphasia: Studies of spoken discourse production and comprehension.*
127. Huiping Chan (2015). *A Dynamic Approach to the Development of Lexicon and Syntax in a Second Language.*
128. James Griffiths (2015). *On appositives*.
129. Pavel Rudnev (2015). *Dependency and discourse-configurationality: A study of Avar.*
130. Kirsten Kolstrup (2015). *Opportunities to speak. A qualitative study of a second language in use.*
131. Güliz Güneş (2015). *Deriving Prosodic structures.*
132. Cornelia Lahmann (2015). *Beyond barriers. Complexity, accuracy, and fluency in long-term L2 speakers' speech.*
133. Sri Wachyunni (2015). *Scaffolding and Cooperative Learning: Effects on Reading Comprehension and Vocabulary Knowledge in English as a Foreign Language.*
134. Albert Walsweer (2015). *Ruimte voor leren. Een etnografisch onderzoek naar het verloop van een interventie gericht op versterking van het taalgebruik in een knowledge building environment op kleine Friese*

basisscholen.

135. Aleyda Lizeth Linares Calix (2015). *Raising Metacognitive Genre Awareness in L2 Academic Readers and Writers.*
136. Fathima Mufeeda Irshad (2015). *Second Language Development through the Lens of a Dynamic Usage-Based Approach.*
137. Oscar Strik (2015). *Modelling analogical change. A history of Swedish and Frisian verb inflection.*
138. He Sun (2015). *Predictors and stages of very young child EFL learners' English development in China.*
139 Marieke Haan (2015). *Mode Matters*. *Effects of survey modes on participation and answering behavior*.
140. Nienke Houtzager (2015). *Bilingual advantages in middle-aged and elderly populations.*
141. Noortje Joost Venhuizen (2015). *Projection in Discourse: A data-driven formal semantic analysis.*
142. Valerio Basile (2015). *From Logic to Language: Natural Language Generation from Logical Forms*.
143. Jinxing Yue (2016). *Tone-word Recognition in Mandarin Chinese: Influences of lexical-level representations.*
144. Seçkin Arslan (2016). *Neurolinguistic and Psycholinguistic Investigations on Evidentiality in Turkish.*
145. Rui Qin (2016). *Neurophysiological Studies of Reading Fluency. Towards Visual and Auditory Markers of Developmental Dyslexia.*
146. Kashmiri Stec (2016). *Visible Quotation: The Multimodal Expression of Viewpoint*.
147. Yinxing Jin (2016). *Foreign language classroom anxiety: A study of Chinese university students of Japanese and English over time.*
148. Joost Hurkmans (2016). *The Treatment of Apraxia of Speech. Speech and Music Therapy, an Innovative Joint Effort.*
149. Franziska Köder (2016). *Between direct and indirect speech: The acquisition of pronouns in reported speech.*
150. Femke Swarte (2016). *Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors.*
151. Sanne Kuijper (2016). *Communication abilities of children with ASD and ADHD. Production, comprehension, and cognitive mechanisms.*
152. Jelena Golubović (2016). *Mutual intelligibility in the Slavic language area*.
153. Nynke van der Schaaf (2016). *"Kijk eens wat ik kan!" Sociale praktijken in de interactie tussen kinderen van 4 tot 8 jaar in de buitenschoolse opvang.*
154. Simon Šuster (2016). *Empirical studies on word representations.*
155. Kilian Evang (2016). *Cross-lingual Semantic Parsing with Categorial Grammars.*
156. Miren Arantzeta *Pérez (2017). Sentence comprehension in monolingual and bilingual aphasia: Evidence from behavioral and eye-tracking methods*.
157. Sana-e-Zehra Haidry (2017). *Assessment of Dyslexia in the Urdu Language.*
158. Srđan Popov (2017). *Auditory and Visual ERP Correlates of Gender Agreement Processing in Dutch and Italian.*
159. Molood Sadat Safavi (2017). *The Competition of Memory and Expectation in Resolving Long-Distance Dependencies: Psycholinguistic Evidence from Persian Complex Predicates.*
160. Christopher Bergmann (2017). *Facets of native-likeness: First-language attrition among German emigrants to Anglophone North America.*
161. Stefanie Keulen (2017). *Foreign Accent Syndrome: A Neurolinguistic Analysis.*
162. Franz Manni (2017). *Linguistic Probes into Human History.*
163. Margreet Vogelzang (2017). *Reference and cognition: Experimental and computational cognitive modeling studies on reference processing in Dutch and Italian.*
164. Johannes Bjerva (2017). *One Model to Rule them all. Multitask and Multilingual Modelling for Lexical Analysis: Multitask and Multilingual Modelling for Lexical Analysis.*
165. Dieke Oele (2018). *Automated translation with interlingual word representations*.
166. Lucas Seuren (2018). *The interactional accomplishment of action.*
167. Elisabeth Borleffs (2018). *Cracking the code - Towards understanding, diagnosing and remediating dyslexia in Standard Indonesian.*
168. Mirjam Günther-van der Meij (2018). *The impact of degree of bilingualism on L3 development English language development in early and later bilinguals in the Frisian context*.
169. Ruth Koops van 't Jagt (2018). *Show, don't just tell: Photo stories to support people with limited health literacy.*
170. Bernat Bardagil-Mas (2018). *Case and agreement in Panará*.
171. Jessica Overweg (2018). *Taking an alternative perspective on language in autism.*
172. Lennie Donné (2018). *Convincing through conversation*: *Unraveling the role of interpersonal health communication in health campaign effectiveness.*

173. Toivo Glatz (2018). *Serious games as a level playing field for early literacy: A behavioural and neurophysiological evaluation.*
174. Ellie van Setten (2019). *Neurolinguistic Profiles of Advanced Readers with Developmental Dyslexia*.
175. Anna Pot (2019). *Aging in multilingual Netherlands: Effects on cognition, wellbeing and health*.
176. Audrey Rousse-Malpat (2019). *Effectiveness of explicit vs. implicit L2 instruction: a longitudinal classroom study on oral and written skills*.
177. Rob van der Goot (2019). *Normalization and Parsing Algorithms for Uncertain Input*.
178. Azadeh Elmianvari (2019). *Multilingualism, Facebook and the Iranian diaspora*.
179. Joëlle Ooms (2019). *"Don't make my mistake": Narrative fear appeals in health communication.*
180. Annerose Willemsen (2019). *The floor is yours: A conversation analytic study of teachers' conduct facilitating whole-class discussions around texts*.
181. Frans Hiddink (2019). *Early childhood problem-solving interaction: Young children's discourse during small-group work in primary school*.
182. Hessel Haagsma (2020). *A Bigger Fish to Fry: Scaling up the Automatic Understanding of Idiomatic Expressions*.
183. Juliana Andrade Feiden (2020). *The Influence of Conceptual Number in Coreference Establishing: An ERP Study on Brazilian and European Portuguese.*
184. Sirkku Lesonen (2020). *Valuing variability: Dynamic usage-based principles in the L2 development of four Finnish language learners*.
185. Nathaniel Lartey (2020). *A neurolinguistic approach to the processing of resumption in Akan focus constructions*.
186. Bernard Amadeus Jaya Jap (2020). *Syntactic Frequency and Sentence Processing in Standard Indonesian.*
187. Ting Huang (2020). *Learning an L2 and L3 at the same time: help or hinder?*.
188. Anke Herder (2020). *Peer talk in collaborative writing of primary school students: A conversation analytic study of student interaction in the context of inquiry learning*.
189. Ellen Schep (2020). *Attachment in interaction: A conversation analytic study on dinner conversations with adolescents in family-style group care*.
190. Yulia Akinina (2020*). Individual behavioural patterns and neural underpinnings of verb processing in aphasia*.
191. Camila Martinez Rebolledo (2020). *Comprehending the development of reading difficulties in children with SLI*.
192. Jakolien den Hollander (2021). *Distinguishing a phonological encoding disorder from Apraxia of Speech in individuals with aphasia by using EEG*.
193. Rik van Noord (2021). *Character-based Neural Semantic Parsing*.
194. Anna de Koster (2021). *Acting Individually or Together? An Investigation of Children's Development of Distributivity*.
195. Frank Tsiwah (2021). *Time, tone and the brain: Behavioral and neurophysiological studies on time reference and grammatical tone in Akan*.
196. Amélie la Roi (2021). *Idioms in the Aging Brain*.
197. Nienke Wolthuis (2021). *Language impairments and resting-state EEG in brain tumour patients: Revealing connections*.
198. Nienke Smit (2021). *Get it together: Exploring the dynamics of teacher-student interaction in English as a foreign language lessons*.
199. Svetlana Averina (2021). *Bilateral neural correlates of treatment-induced changes in chronic aphasia*.
200. Wilasinee Siriboonpipattana (2021). *Neurolinguistic studies on the linguistic expression of time reference in Thai*.
201. Irene Graafsma (2021). *Computer programming skills: a cognitive perspective*.
202. Pouran Seifi (2021). *Processing and comprehension of L2 English relative clauses by Farsi speakers*.
203. Hongying Peng (2021). *A Holistic Person-Centred Approach to Mobile-Assisted Language Learning*.
204. Nermina Cordalija (2021). *Neurolinguistic and psycholinguistic approaches to studying tense, aspect, and unaccusativity*.
205. Aida Salčić (2021). *Agreement processing in Dutch adults with dyslexia*.
206. Eabele Tjepkema (2021). *Exploring content-based language teaching practices to stimulate language use in grades 7 and 8 of Frisian trilingual primary education*.
207. Liefke Reitsma (2021). *Bilingualism and contact-induced language change: Exploring variation in the Frisian verbal complex*.
208. Steven Gilbers (2021). *Ambitionz az a Ridah: 2Pac's changing accent and flow in light of regional variation in African-American English speech and hip-hop music*.

209. Leanne Nagels (2021). *From voice to speech: The perception of voice characteristics and speech in children with cochlear implants*.
210. Vasilisa Verkhodanova (2021). *More than words: Recognizing speech of people with Parkinson's disease*.
211. Liset Rouweler (2021). *The impact of dyslexia in higher education*.
212. Maaike Pulles (2021). *Dialogic reading practices: A conversation analytic study of peer talk in collaborative reading activities in primary school inquiry learning*.
213. Agnes M. Engbersen (2022). *Assisting independent seniors with morning care: How care workers and seniors negotiate physical cooperation through multimodal interaction*.
214. Ryssa Moffat (2022). *Recognition and cortical haemodynamics of vocal emotions—an fNIRS perspective*.

GRODIL
Center for Language and Cognition Groningen (CLCG)
P.O. Box 716
9700 AS Groningen
The Netherlands

Propositions

1. Normal hearing listeners rely more heavily on variations over time in fundamental frequency than intensity or speech rate, to extract emotional meaning from speech.

2. Faced with attenuated variations in F0, listeners do not confuse emotions (e.g., *happy* and *sad*), but rather consider the speech void of emotions (*unemotional*).

3. Cortical activity evoked by specific emotions (e.g., happy relative to unemotional) cannot be observed with region-of-interest-based nor channel-based analyses of fNIRS data recorded from the superior temporal, inferior frontal, or middle frontal gyri.

4. Right lateralisation of cortical responses to vocal emotions can be observed in fNIRS data recorded from superior temporal gyri.

5. Cortical haemodynamic activity in right superior temporal gyrus may be viable as a biomarker for emotion recognition performance when F0 cues are reduced.

6. Subtracting deoxygenated (HbR) from oxygenated (HbO) haemoglobin estimates to obtain a single measure of haemodynamic response magnitude is a promising way to simplify fNIRS analyses, while accounting for the dynamics of both chromophores.