PYSAN3 Lecture 3

The social and emotional voice

Emotional prosody and voice quality

- "Tone of voice"
- Not the linguistic or semantic information
  - Can be isolated from linguistics
  - Can interact with verbal content (e.g. sarcasm)
- Conveyed with changes in pitch, loudness, timbre, speech rate, pauses
- Conveys emotion, attitude, identity, "personality"?

How important is it?

http://www.boingboing.net/2008/02/11/spongebob-voice-acto.html

- From the earliest stages of development, infants respond to affect-laden vocal expressions from their mothers (Fernald, 1989; Fernald and Morikawa, 1993).
- Brain regions specialised for human vocal sounds (Zatorre; Binder)

Voice processing in the brain

- Regions in Superior Temporal Sulcus and Gyrus (STS and STG) respond preferentially to human voices/speech (Zatorre, Binder)
- Binder (2000) and also Belin (2000): Compared fMRI BOLD responses to FM modulated noise, words, nonwords, reversed speech

http://cercor.oxfordjournals.org/cgi/content-nw/full/10/5/512/T1

Table 1 Stimulus characteristics

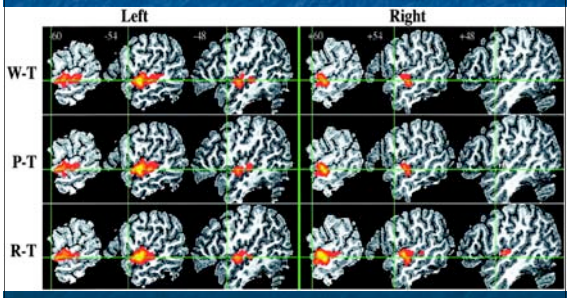
Stimulus	Physical attributes			Associated representations			
	Frequency bandwidth	AM and FM	Diversity	Pitch	Phonemic	Lexical	Semantic
Noise	20-11000	–	–	–	–	–	–
Tones	50-2400	+	+	++	–	–	–
Reversed	50-3000	++	++	+	+	–	–
Pseudowords	50-3000	++	++	+	++	+	–
Words	50-3000	++	++	+	++	++	++

The degree to which each stimulus is characterized by a particular physical attribute or set of associated representations is indicated by '–' (not characteristic), '+' (characteristic) or '++' (highly characteristic).

Done zotero

- Binder (2000)

Voice processing in the brain

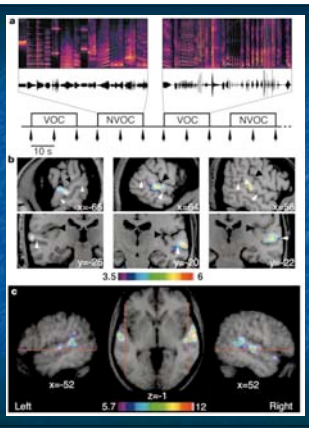


Binder (2000)

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P. and Pike, B. (2000) Voice-selective areas in human auditory cortex. Nature, 403, 309-312.

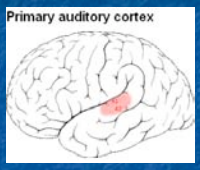
- http://www.zlab.mcgill.ca/supplements/supplements_intro.html#

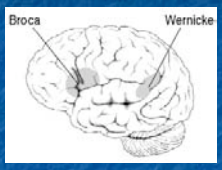
Belin et al (2000)



Belin et al (2000)

Auditory Areas





Emotional prosody and voice quality

- Perceived/decoded slightly worse than facial expressions
- Decoding varies with emotions:
  - Anger, sadness: high
  - Fear, happiness: medium
  - Disgust: poor

  (e.g. Banse and Scherer, 1996)
- Universal: same emotional cues used and understood across different cultures (Scherer and Wallbott, 1994)

Emotional prosody and voice quality

Table 1

Comparison of Accuracy Percentages for Individual Emotions in Two Empirical Studies

Study	Fear	Disgust	Joy	Sadness	Anger
van Bezooijen (1984)	58	49	72	67	74
Scherer et al. (1991)	52	28	59	72	68

Note. From "Vocal Expression and Communication of Emotion" by J. Pittam and K. R. Scherer, 1993, in M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions*, New York: Guilford Press. Copyright 1993 by Guilford Press. Reprinted by permission.

The brain and vocal emotion

- Is vocal emotion processed in the same brain regions as verbal elements of speech? (e.g. LH dominance, Broca and Wernicke)?
- If not, are their distinct neural regions involved in perception of vocal emotion?
- How is voice combined with other modalities (e.g. face)?

The brain and vocal emotion: hemispheric specialisation

- Early evidence: Observations from Hughlings-Jackson (1915) that patients with LH brain damage had severe verbal & linguistic deficits, but could express emotion in the voice
- RH brain damage associated with impaired ability to decode vocal emotion (Tucker et al., 1977; Bowers et al., 1987; Heilman et al., 1984; Peper and Irle, 1997; Ross, 1981)
- Depending on the location of the lesion, patients with RHD may also have difficulty expressing emotional prosody (Gorelick and Ross, 1987; Brådvik et al., 1991; Pell, 1999).

The brain and vocal emotion: hemispheric specialisation

- Using positron emission tomography (PET), George et al. (1996) reported greater right prefrontal activation during processing of the emotional prosody than during processing of the emotional propositional content of spoken sentences.
- Pihan et al. (1997) reported a right hemisphere lateralization in scalp electroencephalography (EEG) signal for the perception of both temporal (accented syllable duration) and frequency (F0 range) mediated emotional prosody.

The brain and vocal emotion: hemispheric specialisation

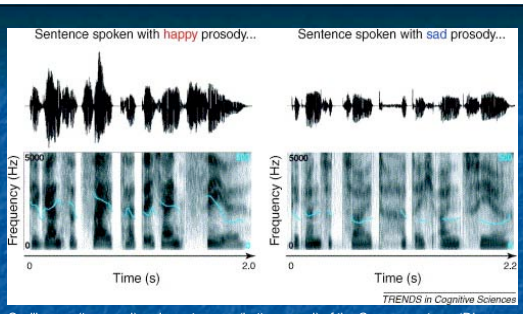
- Ross (1981) has proposed that there exists a RH specialisation for processing vocal emotion:
  - more posterior (temporal/parietal) regions are involved in decoding vocal emotion (during listening)
  - more anterior RH (prefrontal & anterior temporal) regions are involved in encoding vocal emotion (during speech)
- A little like an emotional prosody version of LH linguistic speech regions

But... it's more complicated than that...

- If there exists a RH bias for emotional prosody, what is it due to?
  - Specialisation for processing emotional content? *Functional Hypothesis*
  - Specialisation for processing pitch, timbre, loudness? *Acoustic Hypothesis*

RH/LH & prosodic cues

- Van Lancker & Sidtis (1992) examined RHD, LHD & NBD controls on affective prosody identification
- 4 emotions: happy, sad, angry, surprised
- RHD and LHD both did poorly compared to NBD controls
- Analysed errors made by groups as a function of acoustic properties of vocal tokens (fundamental frequency (FO) variability, mean FO, and syllable durations)
- Found that LHD and RHD were using acoustic cues differently
  - NOT emotion per se, but rather the acoustic cues used to portray emotion, which are often those processed in the RH



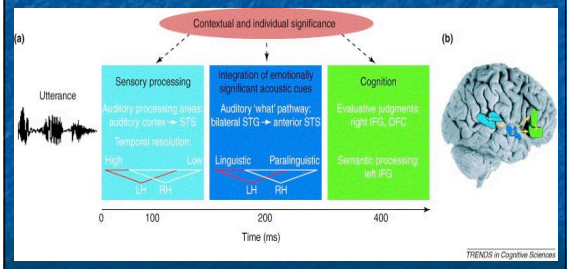
Oscillogram (top panel) and spectrogram (bottom panel) of the German sentence 'Die ganze Zeit hatte ich ein Ziel.' ('During all this time I had one goal.'). The sentence is shorter when produced with a happy prosody (2 s) than with a sad prosody (2.2 s). The speaker is louder as can be seen by comparing the sound envelope illustrated in the oscillogram. Spectral differences between happy and sad prosody are illustrated in the spectrogram. The dark shading indicates the energy of frequencies up to 5000 Hz. The superimposed blue lines represent the fundamental frequency contour, which is perceived as speech melody. This contour shows greater variability and a higher mean for happy than for sad

RH/LH & prosodic cues

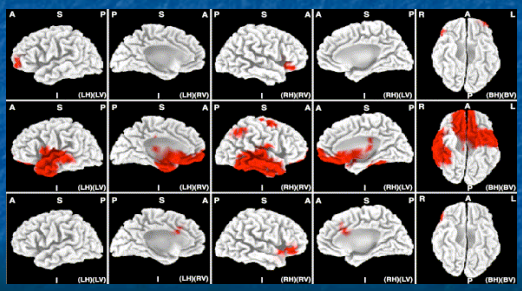
- Rapid aspects of speech processed more in LH
- Slowly changing aspects of speech processed more in RH
- Slowly changing aspects carry more emotional information

→ emotional prosody processed more in RH

Schirmer & Kotz (2006)



The brain and vocal emotion: sequence of processing



Essler et al. (

Outstanding questions

- If the acoustic hypothesis is correct, then are aspects of speech such as speaker identity also processed in RH (because they are conveyed with similar acoustic cues)
- Females are better than males at recognising vocal emotion (and faces too), so are there detectable differences in brain activation between males and females when listening to emotional speech? What might the implications be?
  - {stay tuned for next week's presentations}

Combining faces and voices

- At some level, facial and vocal emotional expressions must be integrated
- Are there regions of the brain that support such multimodal integration?
- What happens if the information from one modality doesn't match the other?