Interspeech 2015
Dresden, September 9, 2015

Voices of power, passion, and personality

Klaus R. Scherer

Swiss Center for Affective Sciences

AGENDA

PREMISES

- Power – real or perceived - is a major determinant of voice quality and speech prosody
- Speech has evolved from affect-related animal vocalizations
- Vocal cues are powerful markers of traits and emotional states, reflecting complex biological and social processes
- Emotions are emerging patterns not natural categories
- Vocal communication depends on both sender and receiver characteristics and intentions

CONSEQUENCES FOR RESEARCH

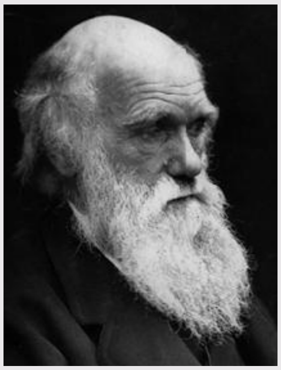
- Research design must reflect the underlying mechanisms (especially in the case of emotion recognition)
- Speech production and perception should be studied conjointly
- Choice of acoustic parameters should be based on physiological considerations, taking interdependencies into account
- More replication and accumulation of research results needed

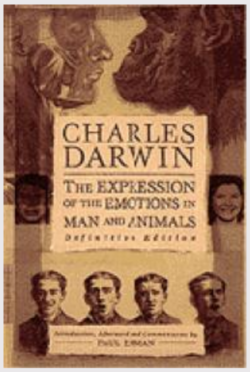
Functions of animal vocalizations for:

- Competition, rivalry -- determining the winner of a contest over a resource
- Mating rituals -- attracting the attention of possible mates
- Social bonding -- creating or solidifying bonds or affective contacts with offspring or other group members
- Ownership/territory -- Signaling a claim or defense of a territory, food, or a mate.
- Food resources -- advertising a food source to a mate, offspring, or other members of the social group
- Predator alarm -- communicating potential threats to the social group
- Meta-communication -- modifying the meaning of subsequent signals (e.g. 'play sounds')

Most, if not all, of these signals inform receivers about affective reactions to events and specific action tendencies of the sender.

Charles Darwin (1872)
The Expression of the Emotions in Man and Animals

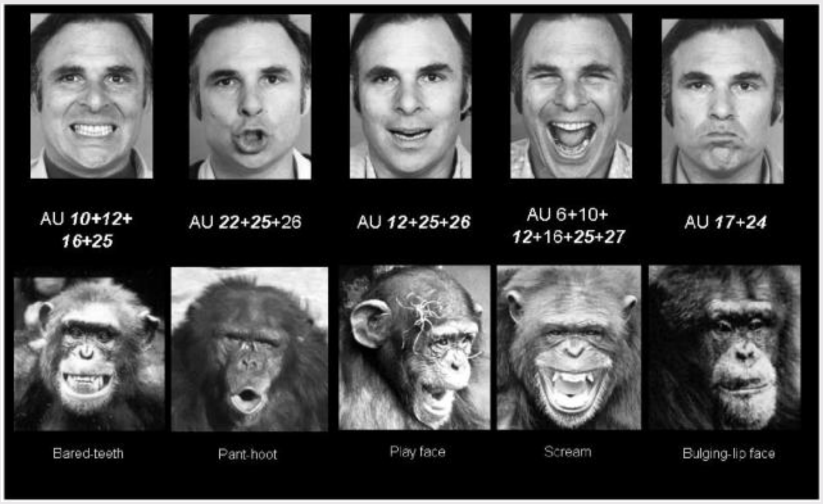




Phylogenetic continuity of affect vocalization

The expression of emotion via facial and vocal expression serves adaptation and is comparable across different species and cultures.

The phylogenesis of emotion expression



Donald Hebb: Species with > cognition have > emotion

The affect burst – mother of all emotional expression

Affect bursts:
Sudden, full-blown displays of emotion in response to highly affectively charged, and often unexpected, events. These eruptive outbursts of emotion are considered as concomitants of automatically produced emotional response patterns (push effects) that consist of brief, synchronized changes in facial, vocal, and possibly gestural and postural, expression.

Hypothesis: Affect bursts were at the origin of the co-evolution of speech and music.

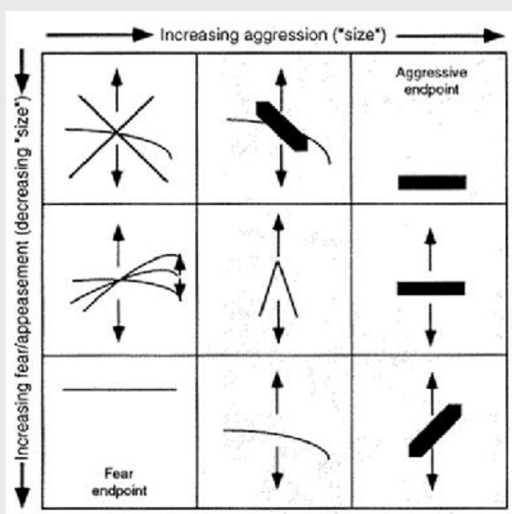
Scherer, K. R. (1991). Emotion expression in speech and music. In J. Sundberg, L. Nord, & R. Carlson (Eds.). Music, language, speech, and brain. (pp. 146-156). Wenner-Gren Center International Symposium Series. London: Macmillan.

Scherer, K. R. (2013). Affect bursts as evolutionary precursors of speech and music. In G. A. Danieli, A. Minelli, T. Pievani (Eds.). Stephen J. Gould: The Scientific Legacy (pp. 147-168). Springer-Verlag Italia.

Affect bursts are phylogenetically continuous and universal

Affective Sciences

Morton's motivational-structural rules: The evolutionary origin of vocal affect expression



Affective Sciences

Vocal affect universals

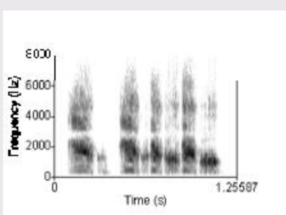
Affective-motivational state of sender	*Acoustic characteristics of animal vocalizations or human affect bursts*
Relaxation, contentment. comfort, play	Repeated short sounds with relatively low frequencies
Dominance, hostility, agonistic intention	Low frequency sounds, harshness, falling frequency
Defense, fear	Short tone-like calls with rising frequency, high amplitude onset, and broad frequency spectrum
Submission, resignation	High-frequency, tone-like sounds with repeated frequency shifts

Scherer, K.R. (1985). Vocal affect signalling: A comparative approach. In J. Rosenblatt, C. Beer, M.-C. Busnel, & P.J.B. Slater (Eds.), Advances in the study of behavior, Vol. 15. (pp. 189-244). New York: Academic Press.

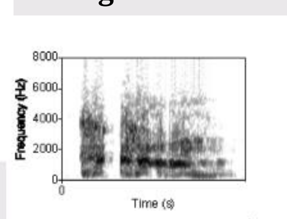
Affective Sciences

Cheney & Seyfarth
Vervet monkey alarm calls

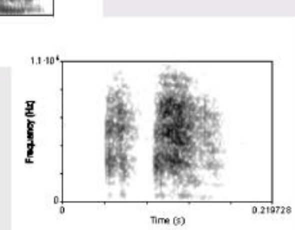
Leopard



Eagle



Snake



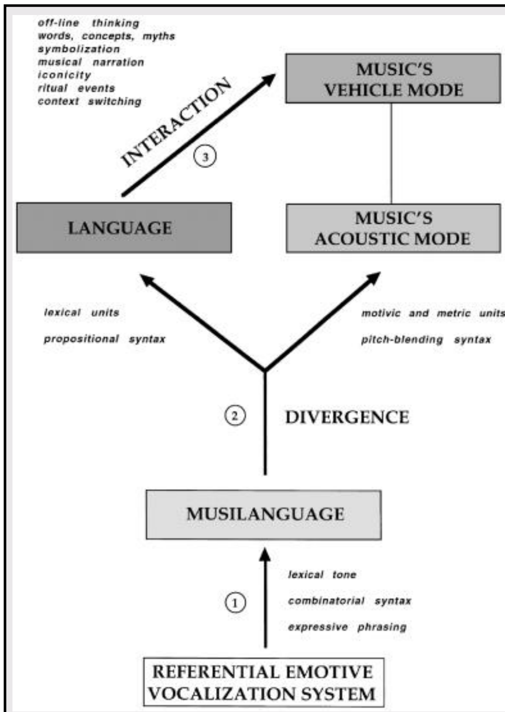
Mithen (2005) – Singing Neanderthals

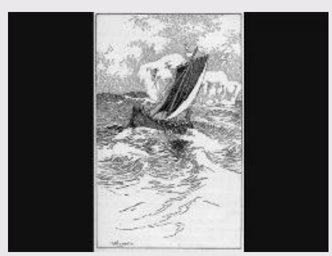
Neanderthal proto-musical language, the "Hmmmmm communication system"

- Holistic— whole phrases rather than words, rather like music
- Manipulative— focused on manipulating behavior of others rather than the transmission of information
- Multimodal — using body as well as the voice
- Musical—using variations in pitch, rhythm, and timbre for emotion expression, care of infants, sexual display, and group bonding
- Mimetic—involving high degree of mime and mimicry of the natural world.

Affective Sciences

Steven Brown (2000) - Musilanguage





Brief illustration of potential elements involved

- Expression and impression are closely linked, resulting in conventionalization and ritualization (Leyhausen, 1967)
- Expressive signals shaped by the constraints of transmission characteristics, limitations of sensory organs, and other factors
- Operation of pull effects facilitated by social pressure for regulation, control, strategic expression
- Resulting flexibility fostered the evolution of abstract, symbolic language and music systems
- Likely to have co-evolved with brain functions & complexity
- Speech as a digital system of information encoding and transmission makes use of primitive, analogue vocal affect signaling system as a carrier signal
- Multimodality – links to posture, gesture, dance

Functions of human vocalizations (including speech) for:

- Competition, rivalry -- determining the winner of a contest over a resource
- Mating rituals -- attracting the attention of possible mates
- Social bonding -- creating or solidifying bonds or affective contacts with offspring or other group members
- Ownership/territory -- Signaling a claim or defense of a territory, food, or a mate.
- Food resources -- advertising a food source to a mate, offspring, or other members of the social group
- Predator alarm -- communicating potential threats to the social group
- Meta-communication -- modifying the meaning of subsequent signals (e.g. 'play sounds')
- Emotions -- affective reactions to events preparing specific action tendencies

There is now a dual code – linguistic and paralinguistic

The central role of power in vocal communication

In most, if not all, of the major functional types of human and animal vocalizations some form of real or perceived power plays a central role in shaping the type of acoustic signal produced

Power: degree of action potential for influencing outcomes by acting on other person's behavior and/or situational determinants

Types of power	Source
- Size, force, attraction	biological, physical
- Mastery, available resources	individual (material, intellectual, referential)
- Authority, rank, position	social-structural

Power has a dispositional aspect, relatively stable over time and contexts, and a situational aspect, dependent on the degree of relevance of the respective type of dispositional power in a given situation and the degree of disponibility or manipulability of the respective resources.

Affective Sciences

Signaling sender characteristics and intentions

Traits
- Biometric (age, sex, weight, size, etc.)
- Social status (power, ascription, region of upbringing, native language)
- Dispositions (personality, attitudes, cognitive and affective disturbance)

States
- Health (hormonal cycles, cold, drug effects)
- Affect (emotion, mood)
- Physiological alertness (fatigue, intoxication)

Intentions
- Interpersonal stances (politeness, hostility, rhetorical-persuasive intent, etc.)
- Message intensions (irony, emphasis, etc.)

Patel, S. & Scherer, K. R. (2013). Vocal behaviour. In J. A. Hall and M. L. Knapp, Eds. Handbook of Nonverbal Communication (pp. 167-204). Berlin: Mouton-DeGruyter.

Affective Sciences

Signaling sender characteristics and intentions

Traits

- Biometric (age, sex, weight, size, etc.)
- Social status (power, ascription, region of upbringing, native language)
- Dispositions (personality, attitudes, cognitive and affective disturbance)

States

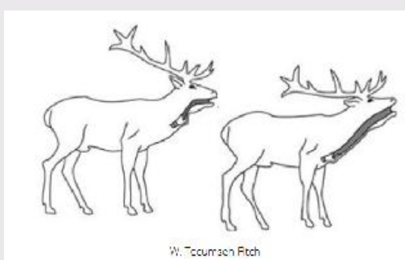
- Health (hormonal cycles, cold, drug effects)
- Affect (emotion, mood)
- Physiological alertness (fatigue, intoxication)

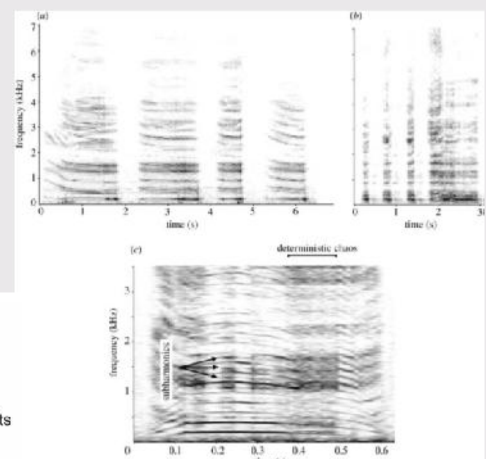
Intentions

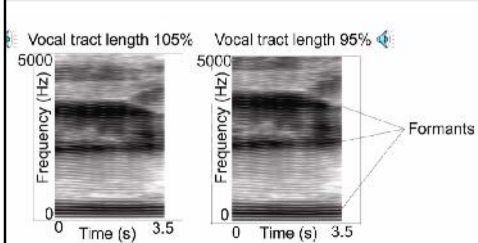
- Interpersonal stances (politeness, hostility, rhetorical-persuasive intent, etc.)
- Message intensions (irony, emphasis, etc.)

Patel, S. & Scherer, K. R. (2013). Vocal behaviour. In J. A. Hall and M. L. Knapp, Eds. Handbook of Nonverbal Communication (pp. 167-204). Berlin: Mouton-DeGruyter.

Physical power – Size (formant dispersion - Fitch, Reby, Feinberg)





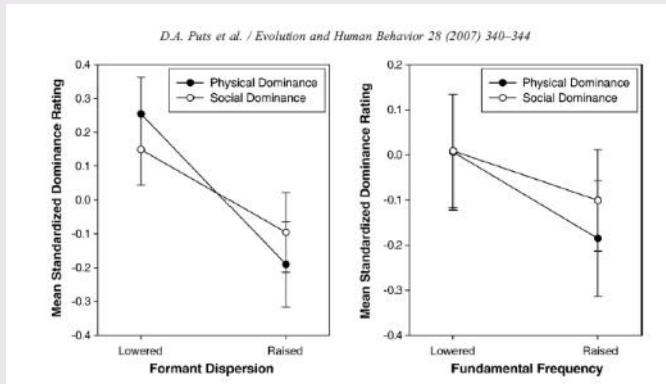


(a) a bout of common roars, clearly defined harmonic structure and modulated formants. The formants drop as the stag extends its vocal tract (by lowering its larynx) (b) A harsh roar, showing a noisy glottal wave (deterministic chaos) and well defined, non-modulated formant frequencies. (c) roar that includes nonlinear phenomena: subharmonics and deterministic chaos; both features confer a harsh quality to roars (Reby & McComb 2003b)

Dominance perception

Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men

David Andrew Puts[a,*], Carolyn R. Hodges[b], Rodrigo A. Cárdenas[c], Steven J.C. Gaulin[b]



Affective Sciences

Baboon rank

Julia Fischer · Dawn M. Kitchen · Robert M. Seyfarth · Dorothy L. Cheney

Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion

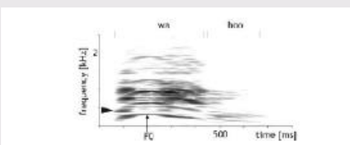
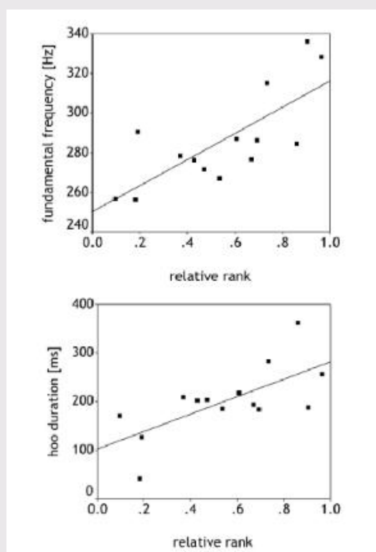


Fig. 1 Spectrogram of a wahoo call. 'Wa' and 'hoo' syllables, and fundamental frequency (*F0*) are shown. The *black triangle* indicates dfa1 start

Here we report that acoustic features of wahoos also reveal information about male competitive ability. High-ranking males give wahoos with higher fundamental frequencies (F0) and longer 'hoo' syllables. Within-subject analyses revealed that, as males fall in rank, the hoo syllables tend to shorten within a period of months. As males age and continue to fall in rank, F0 declines, hoo syllables shorten, and formant dispersion decreases. Independent of age and rank, within bouts of calling F0 declines and hoo syllables become shorter. Because wahoos are often given while males are running or leaping through trees, variation in these acoustic features may function as an indicator of a male's stamina. The acoustic features of contest wahoos thus potentially allow listeners to assess a male's competitive ability.



Affective Sciences

Voice of authority ? (Ko et al, Psych.Sci., 2014)

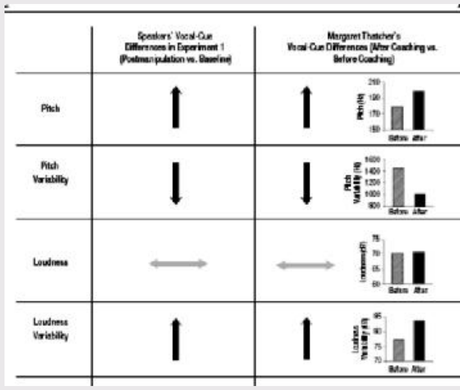


Table 4. Results From Experiment 1: Adjusted Means for the Hierarchy-Based Acoustic Cues and Effect Sizes for Between-Condition Differences in These Means

Acoustic cue	High-rank condition	Low-rank condition	Effect of condition: η^2
Pitch (f_0, in Hz)*	158.61	155.52	.03
Pitch variability (Hz)*	1425.02	1648.37	.03
Loudness (dB)	59.34	58.67	.01
Loudness variability (dB)*	196.73	183.48	.03
Resonance (D_f, in Hz)	1129.39	1128.81	.00
Resonance variability (Hz)	42170.78	43654.54	.00

Note: Each mean was adjusted by its corresponding baseline acoustic cue. Asterisks indicate significant differences between the conditions (*$p < .05$).

Affective Sciences

The voice of confidence

		Voice	Confident	Doubtful
Energy	Text	Confident	108.7	96.2
		Doubtful	115.2	94.6
F0 (Hz)	Text	Confident	113.0	112.1
		Doubtful	131.8	105.9
Rate of speech	Text	Confident	15.6	20.3
		Doubtful	17.7	19.6

Scherer, K. R., London, H., & Wolf, J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7*, 31-44.

Affective Sciences

Voice and personality: A Brunswikian model

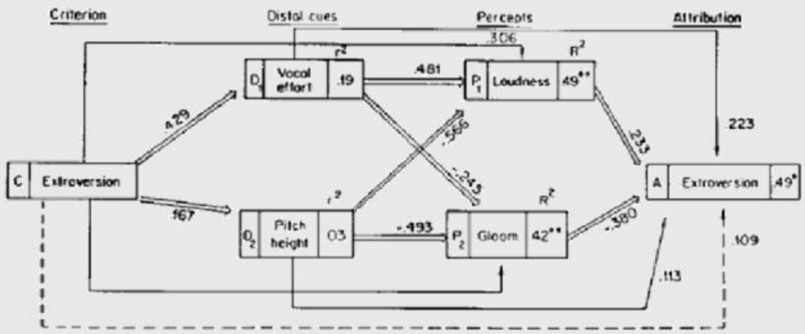


Figure 5. Two-dimensional path analysis model for the extroversion inference

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology, 8*, 467-487.

Speaker personality and perceived influence

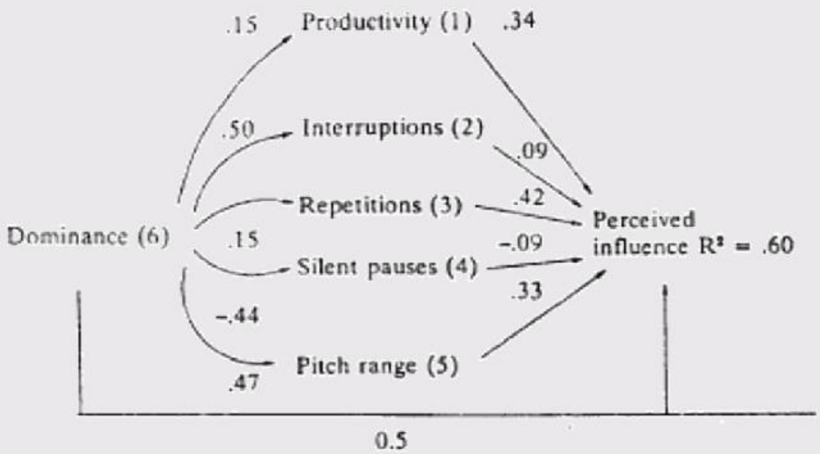
Table 5.2 Correlations between perceived influence and self and other jurors' personality ratings

Personality rating scale	Perceived influence: American jurors (N = 28) Self	American jurors Other jurors	German jurors (N = 31) Self	German jurors Other jurors
Dependability	0.03	0.52**	–0.42*	0.33
Task ability	0.12	0.73***	0.22	0.72***
Neuroticism	0.24	–0.26	0.43*	–0.45*
Stability	0.09	0.75***	–0.28	0.40*
Sociability	0.33	0.56**	–0.32	0.79***
Dominance	0.40*	0.72***	–0.31	0.63***
Likeability	–0.27	0.43*	–0.31	0.46*
Aggressiveness	0.10	0.16	–0.07	0.25

*p <.05, **p <.01, ***p <.001

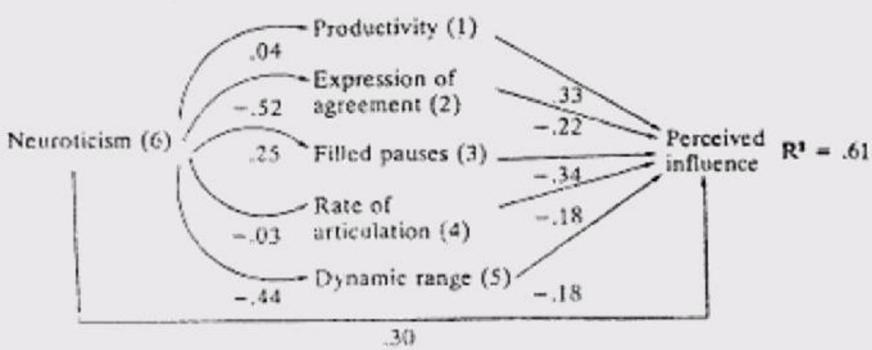
Scherer, K. R. (1979). Voice and speech correlates of perceived social influence in simulated juries. In H. Giles & R. St.Clair (Eds.), The social psychology of language (pp. 88-120). London: Blackwell.

Determinants of perceived influence
American jurors



Affective Sciences

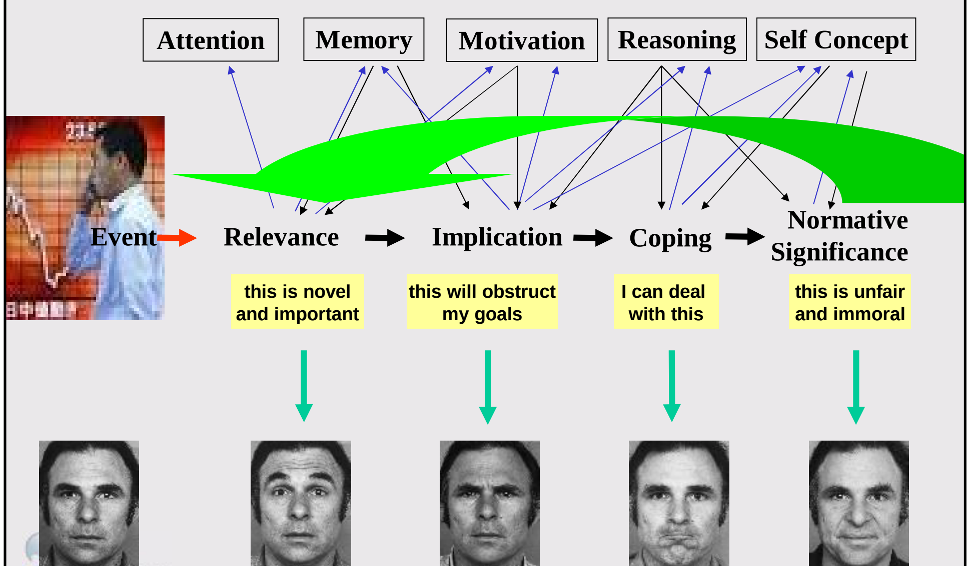
Determinants of perceived influence
German jurors



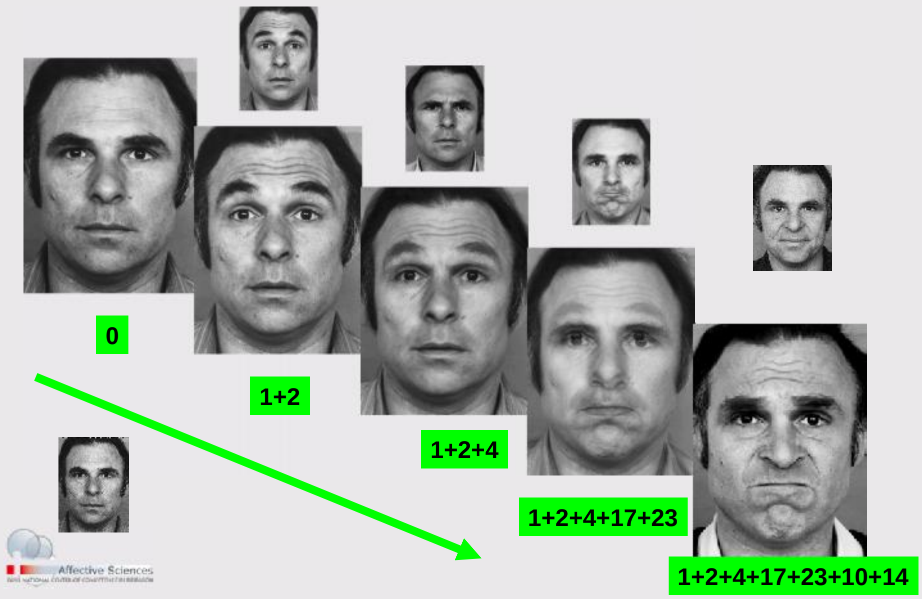
Affective Sciences

And now...for Emotion!

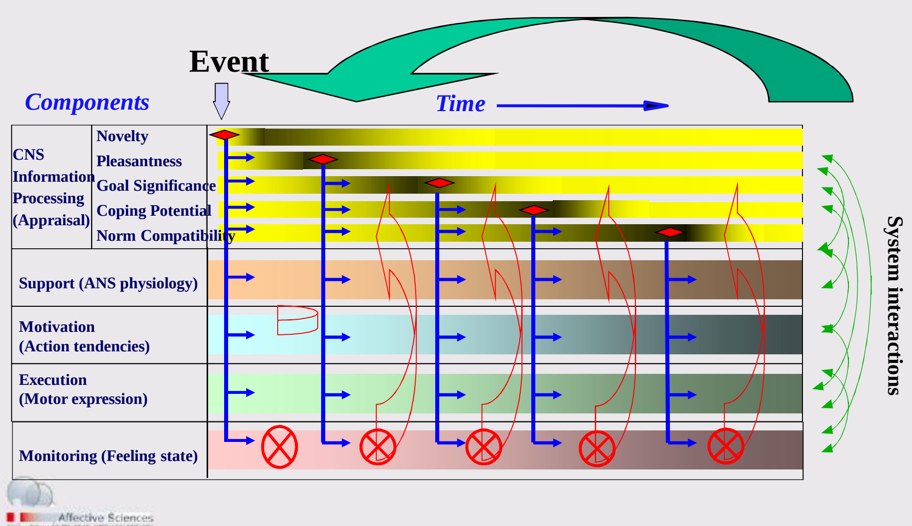
Component process model
Appraisal and efferent motor expression



Appraisal-generated sequential cumulative expression



Efferent effects of appraisal results on different components

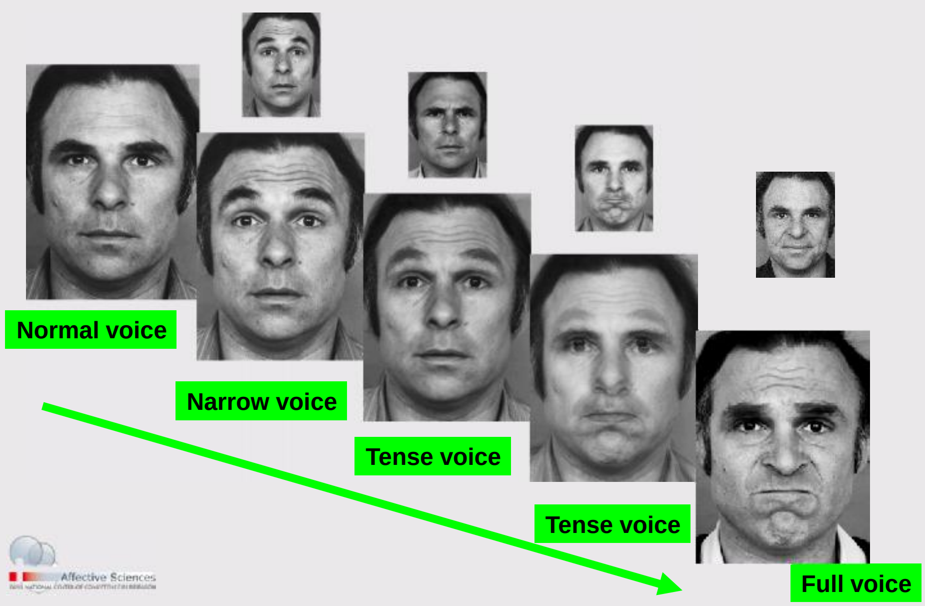


Predictions on appraisal check effects on voice production

Check	Production	Type of voice	Acoustic parameters
Pleasant:	faucal and pharyngeal expansion, relaxation of tract walls, vocal tract shortened due to AU 25 action	wide	increase in low frequency energy, F1 falling, slightly broader F1 bandwidth, velopharyngeal nasality, resonances raised
Unpleasant:	faucal and pharyngeal constriction, tensing of tract walls, vocal tract shortened due to AU 15 action	narrow	more high frequency energy, F1 rising, F2 and F3 falling, narrow F1 bandwidth, laryngopharyngeal nasality, resonances raised
Relevant and consistent:	overall relaxation of vocal apparatus	relaxed	F0 at lower end of range, low-to-moderate amplitude, balanced resonance with slight decrease in high-frequency energy
Relevant and discrepant:	overall tensing of vocal apparatus	tense	F0 and amplitude increase, jitter and shimmer, increase in high frequency energy, narrow F1 bandwidth, pronounced formant frequency differences
Low control:	hypotonus of vocal apparatus	lax	low F0 and restricted F0 range, low amplitude, weak pulses, very low high-frequency energy, spectral noise, format frequencies tending toward neutral setting, broad F1 bandwidth -
High control:	tensing	tense	See tense voice
High power:	Chest register phonation	full	low F0, high amplitude, strong energy in entire frequency range
Low power:	Head register phonation	thin	raised F0, widely spaced harmonics with relatively low energy

Affective Sciences

Appraisal-generated sequential cumulative expression



Affective Sciences

CPM hypotheses for push effects

	Stress	Anger/ Rage	Fear/ Panic	Sad- ness	Joy	Bore- dom
Intensity	↗	↗	↗	↘	↗	
F0 floor/mean	↗	↗	↗	↘	↗	
F0 variability		↗		↘	↗	↘
F0 range		↗	↗(↘)	↘	↗	↘
Sentence contours		↘		↘		
High frequency energy		↗	↗	↘	(↗)	
Speech and articulation rate		↗	↗	↘	(↗)	↘

Affective Sciences

Emotion enacting by professional actors: The Munich corpus

9 actors/speakers

2 sentences

hätt sandik prong nju ventsie

fi gött leich jean kill gos terr

cold anger **(irrit)**
hot anger **(rage)**

sadness **(trist)**
despair **(desp)**

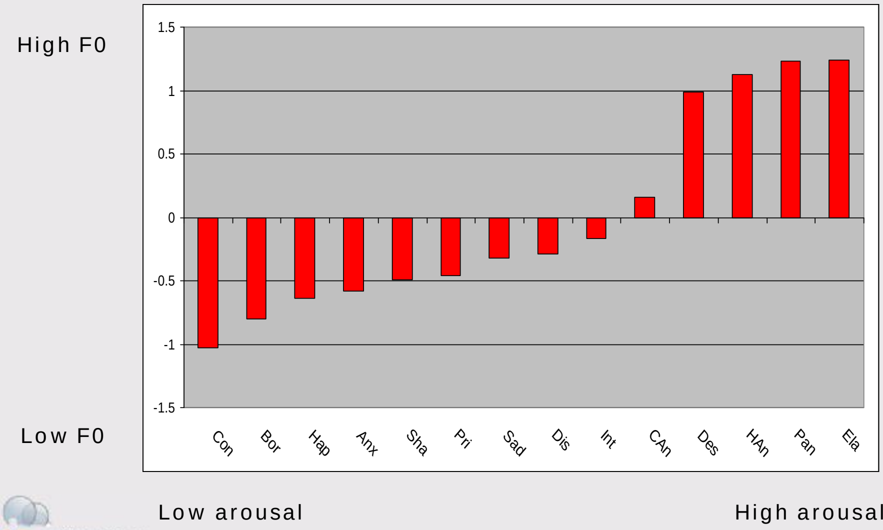
anxiéty **(anx)**
panic fear **(paniq)**

calm joy **(joie)**
elated joy **(exalt)**

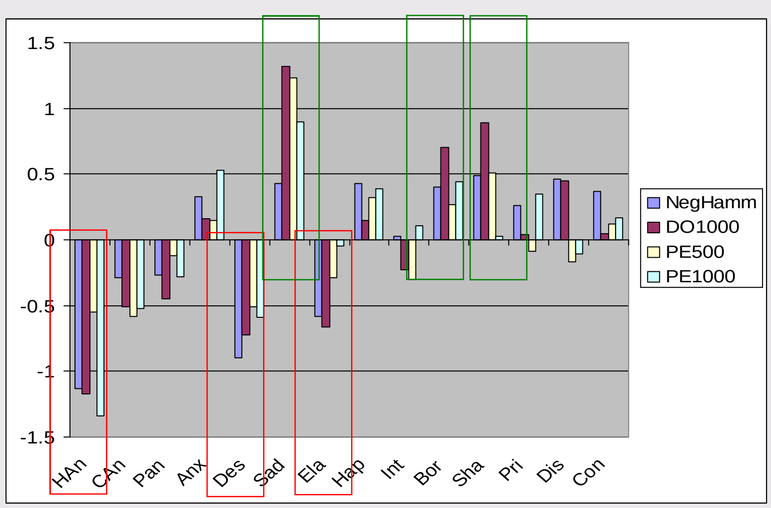
Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, 70(3), 614-636.

Affective Sciences

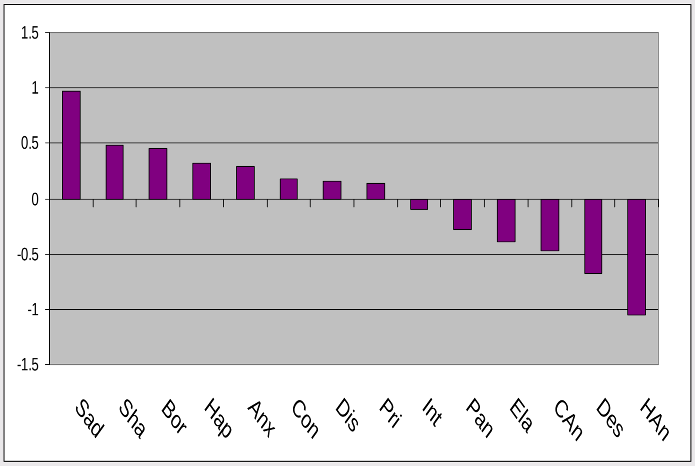
Arousal/activation dimension – Fundamental frequency



Energy distribution in the spectrum



Tense/Lax voice dimension reflected in spectral energy



Stronger Low Freq. Energy

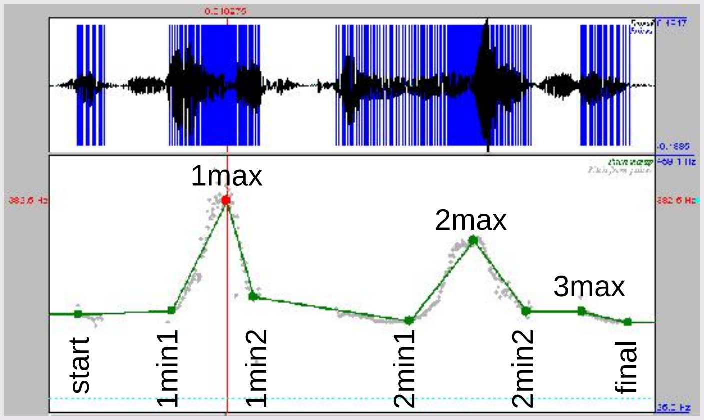
Stronger High Freq Energy

Lax

Tense

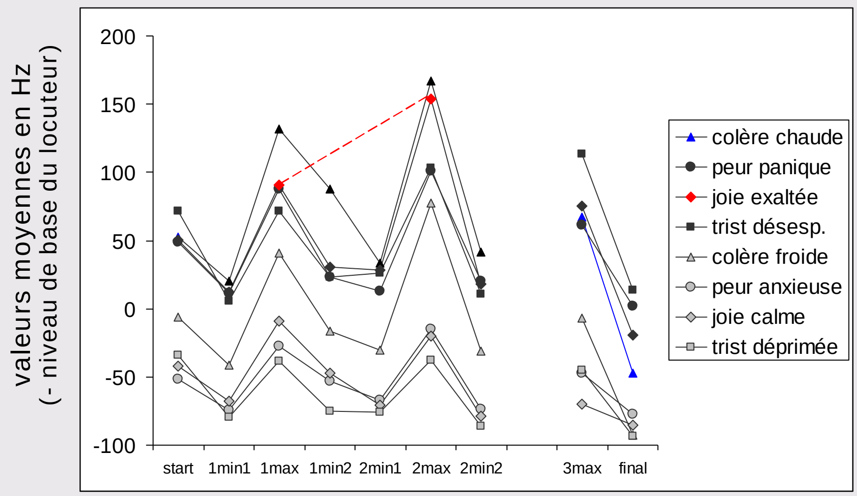
Stylization and quantitative measurement of intonation contours in standard, nonlinguistic pseudo-sentences

- Determination of F0 values for contour minima and maxima in 3 commonly realized accents

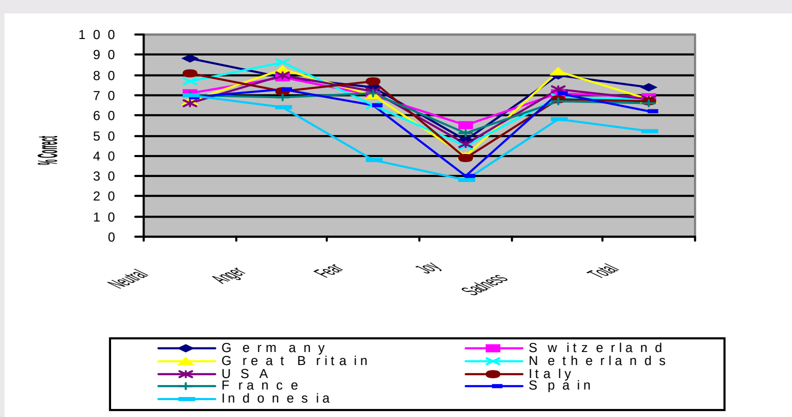


Bänziger, T. & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication, 46*, 252-267

One single contour shape effect found (for elated joy): Disproportional rise of second accent



Vocal communication of emotion: The effects of language and culture



Percentage of correct responses by type of emotion and country of judges

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.

Geneva Multimodal Expression Portrayals (GEMEP)

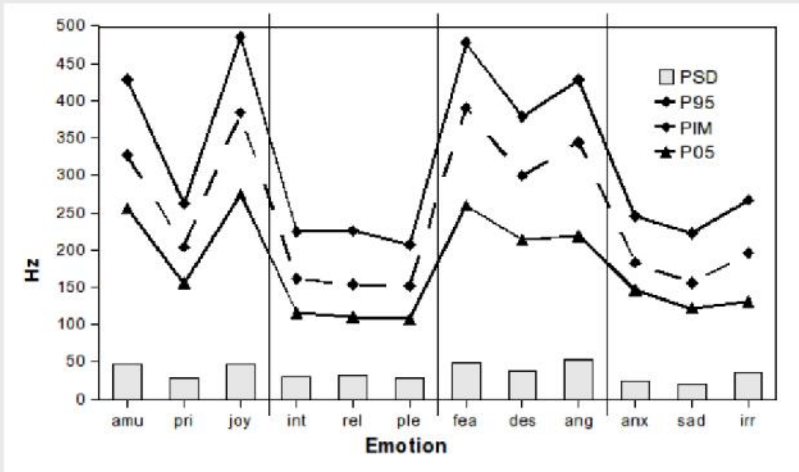
Bänziger, T. & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.). Blueprint for affective computing: A sourcebook (pp. 271-294). Oxford: Oxford University Press.

Emotions enacted by 10 professional actors (Stanislaski method with a professional director)

	Valence	
Arousal	**Positive**	**Negative**
High	Elation (joy)	Hot anger (rage)
	Amusement	Panic fear
	Pride	Despair
Low	Pleasure	Cold anger (irritation)
	Relief	Anxiety (worry)
	Interest	Sadness (depression)

Note. Additional states: shame, surprise, admiration, disgust, contempt, tenderness.

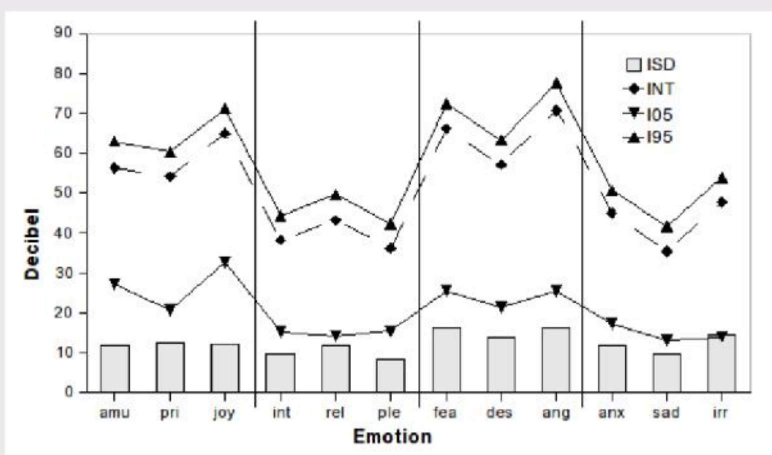
F0 measures



Goudbeek, M. & Scherer, K. R. (2010). Beyond arousal: Valence and potency/control in the vocal expression of emotion. Journal of the Acoustical Society of America, 128, 3, 1322-1336.

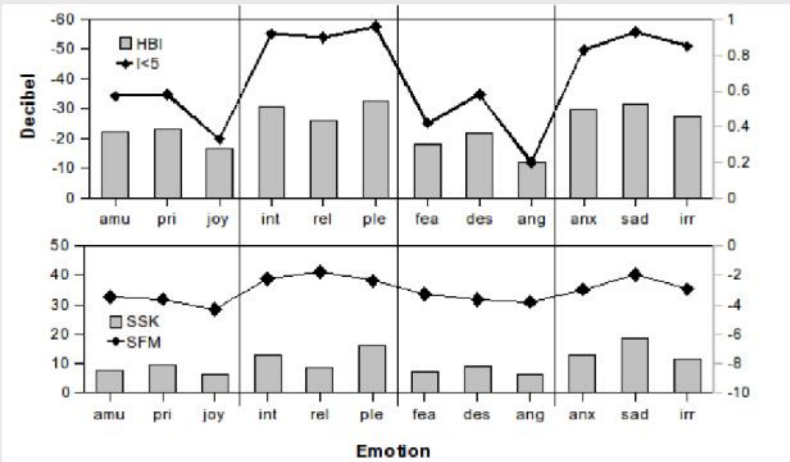
Affective Sciences

Intensity measures



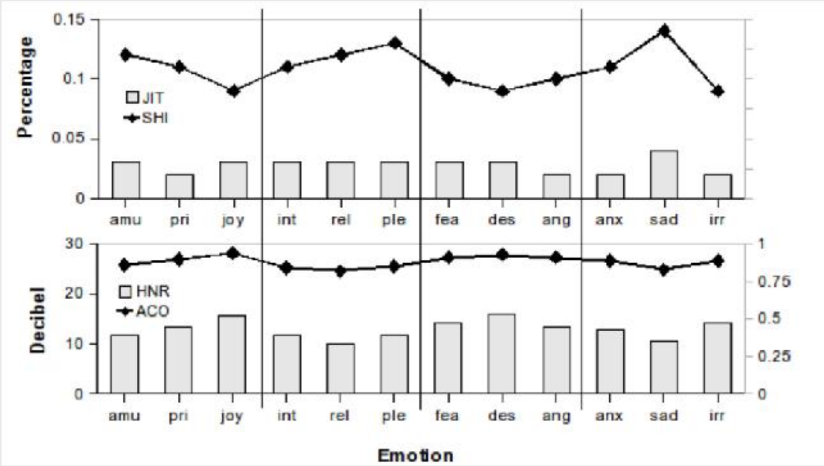
Affective Sciences

Spectral energy distribution



Affective Sciences

Perturbation measures



Affective Sciences

GEMEP actors producing /aa/s

TABLE 1
LIST OF ACOUSTIC PARAMETERS (AND ABBREVIATIONS) EXTRACTED FROM THE EMOTIONAL EXPRESSIONS

Feature	Description
Leq	Equivalent sound level
Alpha	Alpha ratio
MFDR	Maximum flow declination rate
Q_{Closed}	Relative duration of the closed phase
H1-H2	Level difference between the first and second harmonics
$H1\text{-}H2_{LTAS}$	Difference between the average long-term averaged spectrum level near mean F0 and the average level one octave higher
AC_{Amp}	Source flow pulse amplitude
NAQ	Normalized amplitude quotient
Mf0	Mean fundamental frequency
Jitter	Jitter (rap)
Shimmer	Shimmer (local)
HNR	Harmonics-to-noise ratio

	Function		
	1	2	3
Leq	.790*	0.002	-0.242
Alpha	-.586*	0.220	0.490
Mf0[a]	.524*	-0.104	0.024
$H1\text{-}H2_{LTAS}$	-.342*	0.206	0.263
AC_{Amp}[a]	.100*	0.057	-0.033
Shimmer	-0.090	.762*	-0.297
HNR[a]	0.203	-.307*	0.043
Jitter[a]	-0.018	.258*	0.024
NAQ	-0.003	-0.033	.702*
H1-H2[a]	-0.311	0.084	.437*
MFDR[a]	0.395	-0.087	-.430*
Q_{Closed}[a]	0.270	-0.039	-.356*

Multiple discriminant analysis yielding 87% accuracy with 12 predictors – and 78% with only the three marked in red

Affective Sciences

Sundberg, J., Patel, S., Björkner, E., Scherer, K. R. (2011). Interdependencies among voice sourc parameters in emotional speech. IEEE Transactions on affective computing, 99, 2423-2426.

TABLE 4
DIRECTION OF CHANGE IN THE INDICATED PARAMETERS FOR FEMALE AND MALE ACTORS' NORMALIZED SCORES

		Leq	Alpha	MFDR	QClosed	H1-H2	H1-H2LTAS	ACAmp	NAQ	Mf0	Jitter	Shimmer	HNR
Sadness	F / M	low	high				high			low	- / high	NC / high	high / low
Fear	F / M	high	low	NC / high	high / -	low	low	low / -	high	high	- / low	low	NC / high
Anger	F / M	high	low	high	high / -	high	low / -	high	low / -	NC / high	- / -	high	low
Rclicf	F / M	low	high	low	low	high	high	low	- / low	low	- / low	high / NC	low / NC
Joy	F / M	high	low	- / low	- / high	- / low	- / low	low	- / low	high	low	low	high

Affective Sciences

Singing emotions study

Following the colloquium "Emotional Power of Music", Geneva 2009, we recorded world class singers to record for us standard scales as well as a nonsense sentence ("Nec kalibam soud molen") in melodic phrases to express 9 emotions:

Admiration
Anger
Anxiety
Despair
Joy
Fear
Pride
Sadness
Tenderness

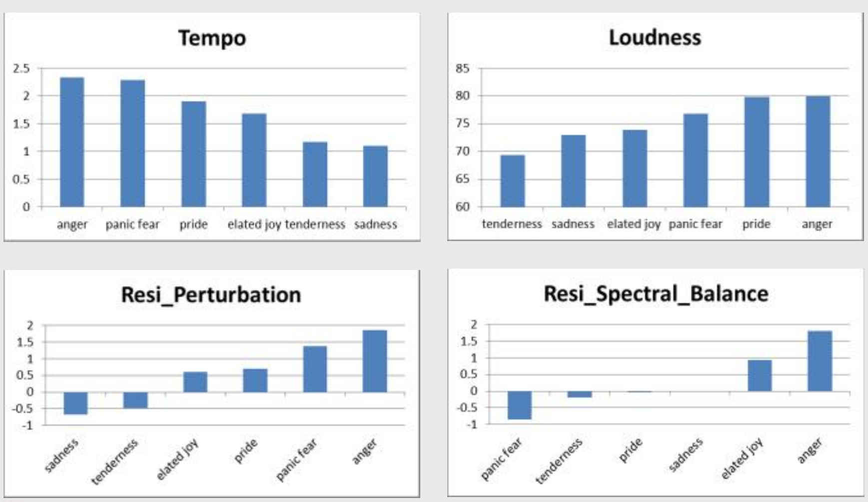
PCA of acoustic paremeters extracted from singer's phrases

	Component				
	1	2	3	4	5
ZHammarbergindex	**-0.85**	-0.31	0.08	-0.25	0.01
ZProportionenergybelow1000Hz	**-0.85**	-0.23	0.36	-0.05	0.21
ZSpectralcentroid	**0.84**	0.32	-0.34	0.16	-0.06
ZAlpha	**0.83**	0.26	-0.37	0.04	-0.25
ZSpectralflatness	0.62	0.46	-0.11	0.55	0.09
ZJitter	0.24	**0.91**	0.05	0.13	-0.03
ZShimmer	0.20	**0.82**	-0.23	0.05	-0.31
ZMeanautocorrelation	-0.41	**-0.80**	0.24	-0.24	0.02
ZMeanHNR	-0.42	**-0.72**	0.44	-0.19	0.19
ZProportionenergybelow500Hz	-0.31	0.01	**0.91**	0.04	0.10
ZH1H2LTAS	-0.13	-0.27	**0.88**	0.13	0.07
ZSIL	0.44	0.22	**-0.70**	0.36	-0.10
ZSlope	0.15	0.15	0.07	**0.94**	-0.04
ZTempo	-0.31	-0.55	0.28	-0.05	**0.68**

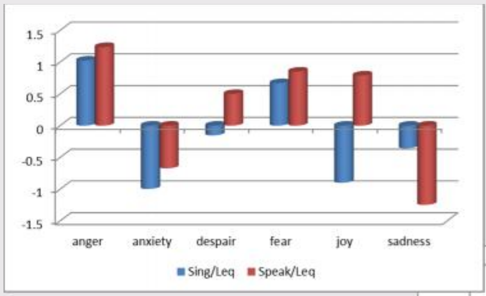
Affective Sciences

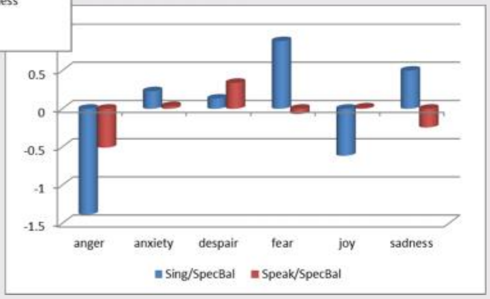
Five markers of voice quality

- Spectral balance = mean level difference between partials above and below 1 kHz; strongly influenced by vocal loudness
- Perturbation = composite measure of aperiodicity, regulated by glottal adjustment, subglottal pressure, vocal tract constriction.
- Low Partial dominance = mean level difference between the lowest and the higher spectrum partials; depending on glottal adduction, forceful adduction attenuating the lowest partials.
- Loudness, vocal effort
- Tempo

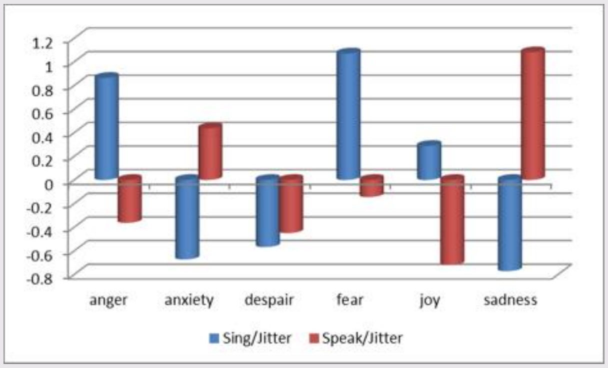


Loudness and spectral balance

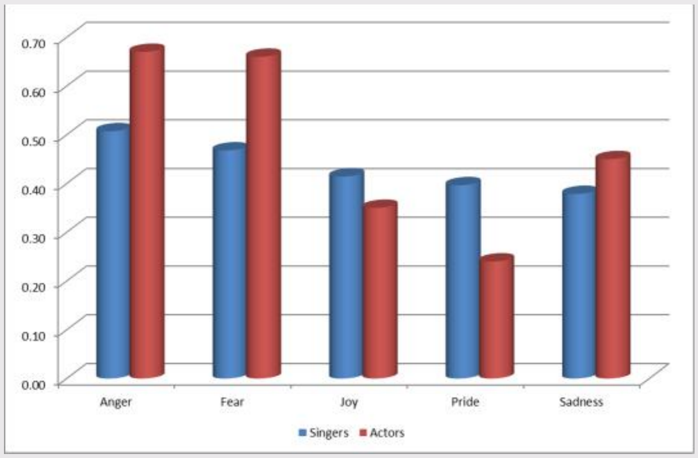




Jitter/Vibrato



Recognition of emotion enacting by singers and actors



Push effects

Examples for pure push effec

- animal expressions
- infant grunts
- affect bursts
- sudden, uncontrolled emotions

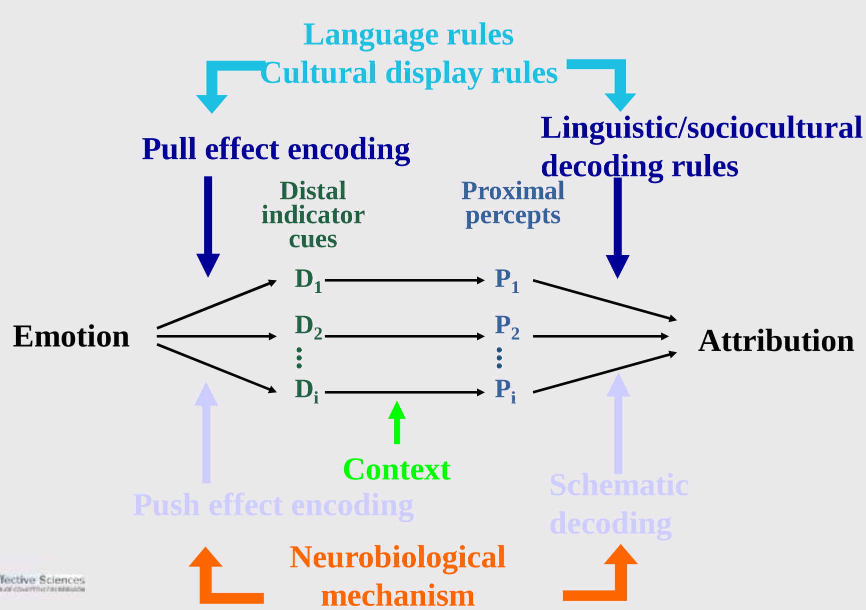
Pull effects

Examples for pure pull effects:

- sound symbolism
- symbolic coding systems, language
- conventionalized expression rules
- mimicking push effects

Dubious cases

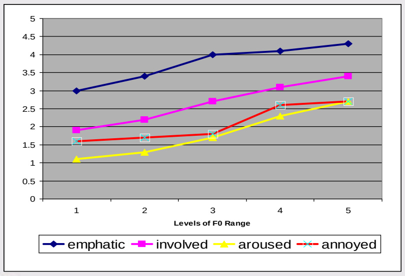
Brunswik's lens model: Combining the study of encoding and decoding



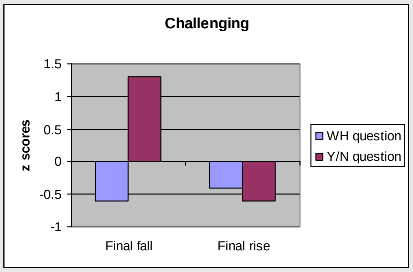
Push vs. Pull: Differential coding of parametric features

- Different features of prosody can be coded differentially
- Scherer, Ladd, & Silverman (JASA, 1984) have shown that F0 range is coded continuously whereas many intonation contours are coded configurationally

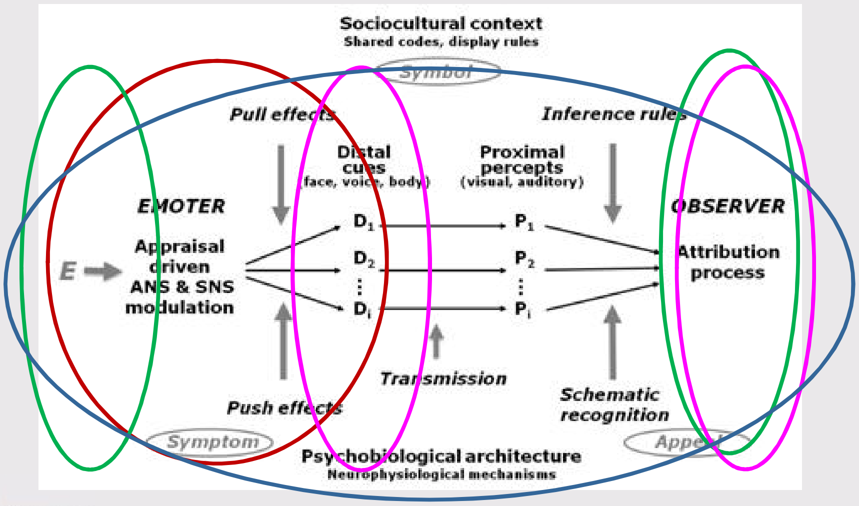
Covariation = mostly push



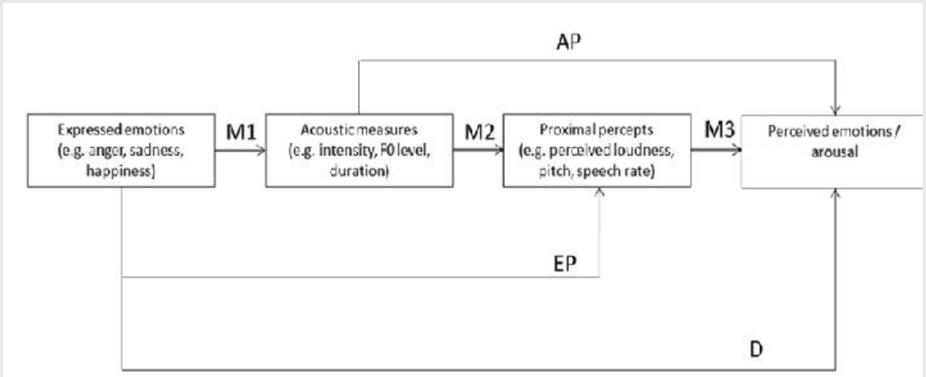
Configuration = mostly pull



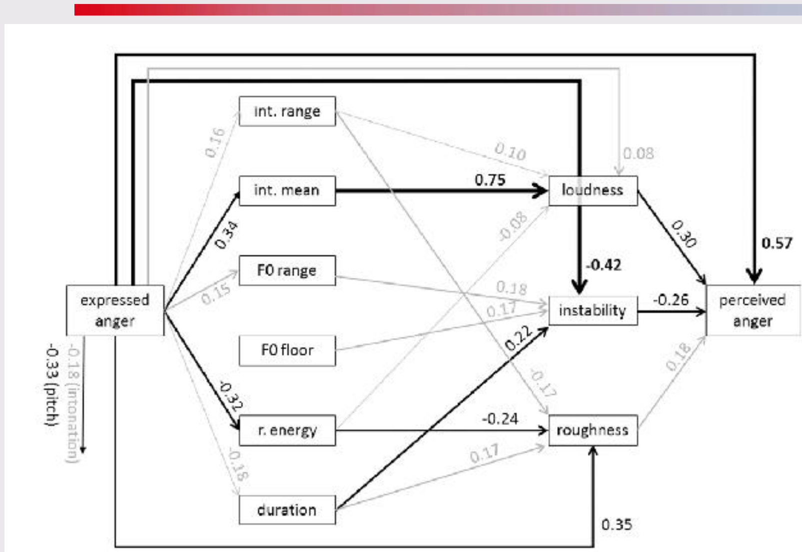
Tripartite Emotion Expression and Perception (TEEP) model



TEEP analysis GEMEP Core Set

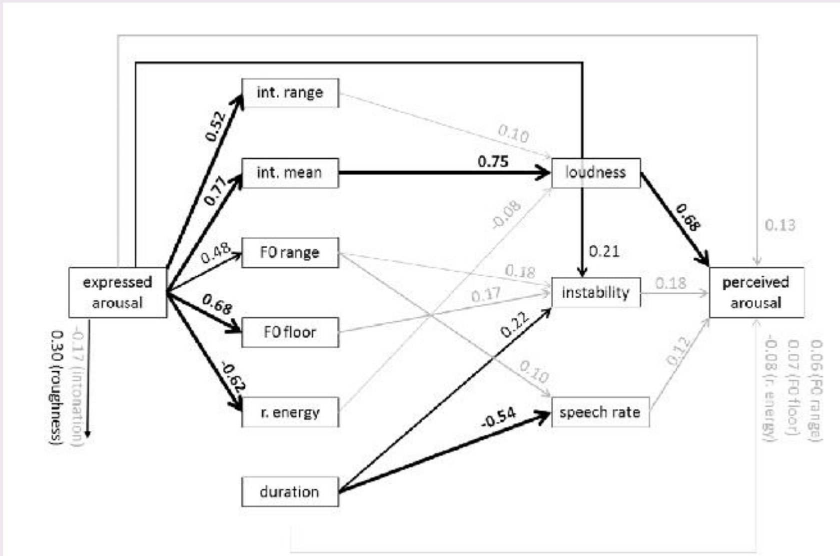


TEEP analysis GEMEP Core Set - ANGER



Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path models of vocal emotion communication. PlosOne, 10(9): e0136675. doi:10.1371/journal.pone.0136675.

TEEP analysis GEMEP Core Set -- AROUSAL



Affective Sciences

GeMAPS LLDs

IEEE TRANSACTIONS ON
AFFECTIVE COMPUTING

The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing

Florian Eyben , Klaus Scherer , Bjorn Schuller , Johan Sundberg , Elisabeth Andre , Carlos Busso , Laurence Devillers , Julien Epps , Petri Laukka , Shrikanth Narayanan , Khiet Truong

Frequency related parameters:
Fundamental frequency (F0, semitone)
Jitter
Formant 1, 2, and 3 frequency
Formant 1, 2, and bandwidth

Energy/Amplitude related parameters
Loudness measures
Shimmer
Harmonics-to-Noise Ratio (HNR

Spectral (balance) parameters:
Alpha Ratio
Hammarberg Index
Proportion < 500 and < 1000 of total energy
Spectral Slope 0–500Hz and 500–1500 Hz
Formant 1, 2, and 3 relative energy
Harmonic difference H1–H2
Harmonic difference H1–A3

Duration parameters

Affective Sciences

OpenSMILE package
Mean, SD, Coefficient of Variation, Rise/Fall
Extended Set: plus MFCCs, spectral flux, etc.

GeMAPS
comparison with other feature sets

Database	Best parameter set	Best UAR [%] with: best set	GeMAPS	eGeMAPS
FAU AIBO	ComParE	43.1[5]	40.4	41.5
TUM-AVIC	InterSp12	69.4	68.8	68.5
EMO-DB	ComParE	86.0	80.0	81.1
GEMEP	InterSp12	43.6	36.9	38.5
SING	ComParE	38.8	29.4	34.0
VAM	InterSp12	43.9	38.5	38.9
EMO-DB (A)	InterSp09	97.8	95.1	95.3
GEMEP (A)	**eGeMAPS**	84.6	84.5	84.6
SING (A)	ComParE	77.2	75.5	75.1
VAM (A)	InterSp11	77.4	74.7	75.3
FAU AIBO (V)	InterSp10	76.2[5]	73.1	73.4
TUM-AVIC (V)	InterSp11	75.9	73.1	73.4
EMO-DB (V)	ComParE	86.7	77.1	78.1
GEMEP (V)	InterSp10	71.4	64.3	65.6
SING (V)	**eGeMAPS**	67.8	66.5	67.8
VAM (V)	**eGeMAPS**	54.1	53.2	54.1

TABLE 3
Leave-one-speaker out classification of affective categories of each database (see each database for description) and binary arousal (A) and valence (V). UAR obtained with best SVM complexity C. Per speaker standardisation, instance up-sampling for balancing of training set.

Face Reading from Speech – Predicting Facial Action Units from Audio Cues

Fabien Ringeval[1,4], Erik Marchi[1], Marc Mehu[2], Klaus Scherer[2], Björn Schuller[3,4,5]

	COMPARE				GEMAPS			
AU	Onset	Apex	Offset	Occu.	Onset	Apex	Offset	Occu.
1	61.88	64.54	59.48	67.03	62.08	**67.96**	61.18	67.62
2	64.38	67.43	62.27	69.33	64.60	**70.99**	62.98	70.90
4	61.35	66.21	63.33	64.66	61.90	66.67	63.19	**67.05**
6	64.01	65.17	63.35	63.78	64.25	**67.71**	63.17	63.11
7	62.71	56.62	58.29	54.47	**62.85**	59.74	61.46	52.72
10	61.82	60.53	56.39	60.39	**61.92**	61.03	58.02	60.34
12	59.34	59.85	56.71	58.67	58.04	**60.92**	57.77	58.72
17	60.47	64.78	62.34	64.53	61.62	64.94	63.13	**65.88**
Avg.	62.00	63.14	60.27	62.86	62.16	**65.00**	61.36	63.29

Affective Sciences

ICMI 2013 – Sydney
Emotion Representations and Modelling for HCI Systems

On real-life inference and agent production of subtle, low intensity emotions

Klaus R. Scherer
Swiss Center for Affective Sciences
University of Geneva

An actuarial study of Swiss emotions

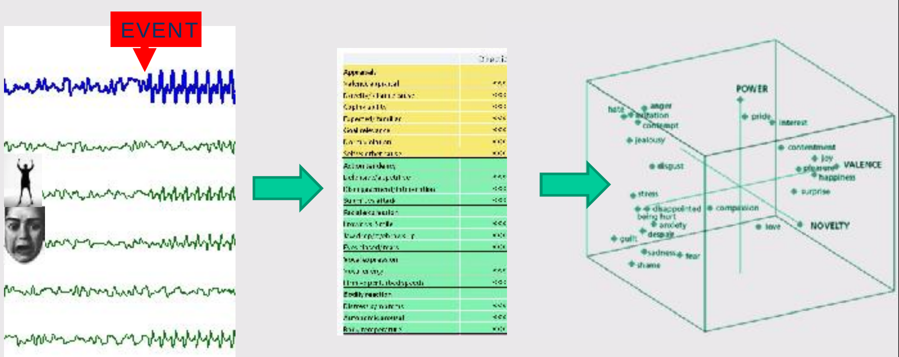
Representative sample of the Swiss population, N = approx 1500. Question: Describe the strongest emotion you had yesterday. How would you call it?

Happiness	14,8
Anger	14,1
Anxiety	9,6
Joy	8,8
Sadness	8,7
Disappointment	6,5
Tension / Stress	6,1
Desperation	6,0
Contentment	5,5
Uncodable	4,8
Irritation	4,1
Negative	3,7
Positive	2,7
Compassion	2,4
Pleasure / Enjoyment	2,1
Pride	2,0

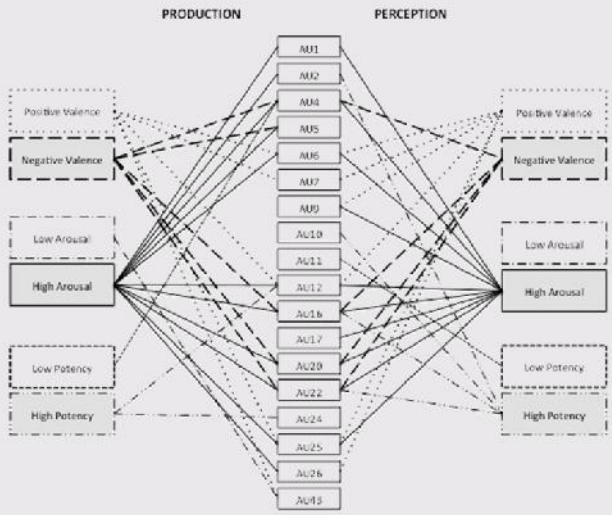
Fear	2,0
Stupefaction	2,0
Surprise	1,9
Guilt	1,7
Relaxation / Serenity	1,7
Relief	1,4
Love	1,1
Amusement	1,1
Gratitude	1,0
Hate	0,8
Interest	0,7
Disgust	0,6
Longing	0,6
Being touched	0,6
Admiration / Awe	0,5

Consequences for machine learning paradigms in expression research

- Ground truth should consist of estimated emergent emotion profiles and self-reported qualia
- This allows projection into four-dimensional affect space
- Potential labels can be determined from semantic GRID look-up tables



GEMEP facial expression – dimensional ratings and emotion categories

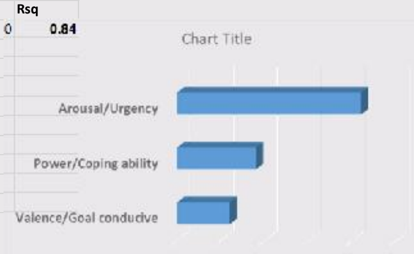


Affective Sciences

Mehu, M. & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: an integrated approach. Emotion (OnlineFirst).

GEMEP – Regression of acoustic parameters on dimension ratings

Dimension	Appraisal			F	p	Rsq
Valence	**Gc/Pleas**	beta	p <	7.49	0	**0.24**
loudness_pctlrange02		-0.251	0.005			
StddevUnvoicedSegmentLength		-0.326	0			
F3 bandwidth		-0.242	0.005			
F2 frequency		-0.261	0.004			
slope500-1500_CoV		-0.23	0.015			
Dimension	Appraisal			F	p	Rsq
Dominance	**Power**	beta	p <	10.9	0	**0.36**
F0 semitone_percentile20		-0.529	0			
loudness_pctlrange02		0.626	0			
F1 amplitudeLogRelF0		0.229	0.004			
logRelF0H1A3_amean		0.326	0.001			
slope0500_CoV		0.186	0.018			
F0semitoneFrom27.5Hz_meanRis		-0.161	0.036			
Dimension	Appraisal			F	p	Rsq
Activation	**Urgency**	beta	p <	21.8	0	**0.84**
mfcc2		-0.592	0			
spectralFlux_CoV		0.209	0			
F0 semitone_percentile50.0		0.643	0			
F1 frequency_CoV		0.102	0.013			
F0 semitone_FallingSlope		-0.153	0			
HNRdBACF		-0.364	0			
energyProportion0-500		0.389	0.004			
energyProportion0-1000_CoV		0.3	0			
mfcc1_CoV		-0.098	0.042			



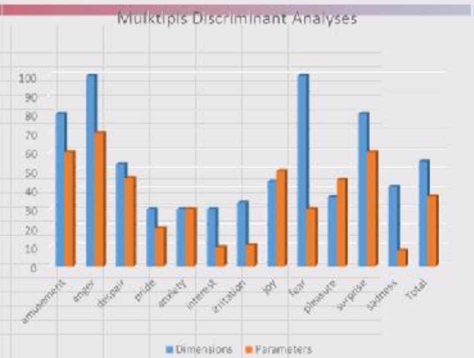
Affective Sciences

Discriminant analysis

Structure Matrix			
	Function		
	1	2	3
Mean_Activation	.755*	0.651	-0.08
Mean_Dominance	-0.249	0.277	.928*
Mean_Valence	0.211	-0.594	.776*

Functions at Group Centroids

emontion	Function		
	1	2	3
amusement	3.68	-1.35	1.50
anger	-0.25	4.38	0.68
despair	0.64	0.00	-1.48
pride	-1.38	0.76	1.20
anxiety	-0.70	-0.52	-0.41
interest	-2.07	-0.89	0.50
irritation	-1.45	0.78	0.62
joy	2.61	-0.32	0.22
fear	2.14	1.27	-1.72
pleasure	-1.30	-1.61	0.50
surprise	0.56	-1.14	0.24
sadness	-2.02	-0.95	-1.17



Classification Results

emontion	Predicted Group Membership												Total
	amusement	anger	despair	pride	anxiety	interest	irritation	joy	fear	pleasure	surprise	sadness	
amusement	80	0	0	0	0	0	0	10	10	0	0	0	100
anger	0	100	0	0	0	0	0	0	0	0	0	0	100
despair	0	0	53.8	0	0	0	0	7.7	15.4	0	23.1	0	100
pride	0	0	0	30	0	0	50	0	0	0	20	0	100
anxiety	0	0	30	0	30	10	10	0	0	0	10	10	100
interest	0	0	0	30	0	30	0	0	0	10	0	30	100
irritation	0	0	11.1	33.3	0	11.1	33.3	0	0	11.1	0	0	100
joy	33.3	0	11.1	0	0	0	0	44.4	11.1	0	0	0	100
fear	0	0	0	0	0	0	0	0	100	0	0	0	100
pleasure	9.1	0	0	0	0	27.3	0	0	0	36.4	9.1	18.2	100
surprise	0	0	10	0	10	0	0	0	0	0	80	0	100
sadness	0	0	16.7	0	8.3	25	8.3	0	0	0	0	41.7	100

a 60,5% of original grouped cases correctly classified.
c 54,8% of cross-validated grouped cases correctly classified.

Affective Sciences

If same variables used in MDA – 68.8/36.8

Personality disorders - Voice of depression

Figure 2. Speech rate of depressed patients. Individual values (syllables/minute) for female (○) and male (∇) patients during depression and remission. Medians = →.

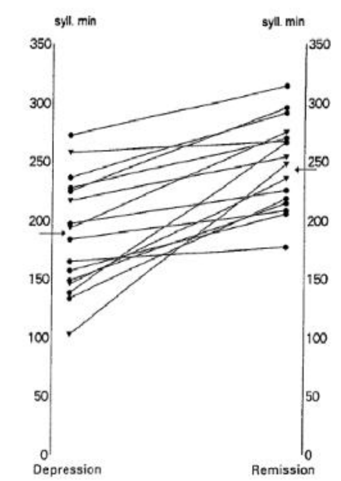
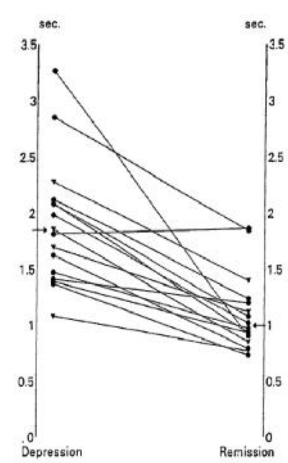


Figure 3. Mean pause duration (MPD) of depressed patients. Individual values (secs) for female (○) and male (∇) patients from speech samples during depression and remission. Medians = →.

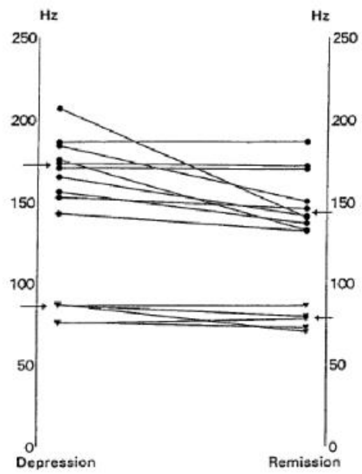


Affective Sciences

Ellgring, H. & Scherer, K. R. (1996). Vocal indicators of mood change in depression. Journal of Nonverbal Behavior, 20, 83-110.

Personality disorders - Voice of depression

Figure 4. Minimum fundamental frequency (MinF0) of depressed patients. Individual values (Hz) for female patients (○) and male patients (∇) from speech samples during depression and remission. Medians = →.



More recent evidence

J Neurolinguistics. 2007 January ; 20(1): 50–64. doi:10.1016/j.jneuroling.2006.04.001.

Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology

James C. Mundt[1], Peter J. Snyder[2,3], Michael S. Cannizzaro[2,4], Kara Chappie[2], and Dayna S. Geralts[1]

From the abstract:

“Patients responding to treatment had significantly greater pitch variability, paused less while speaking, and spoke faster than at baseline. Patients not responding to treatment did not show similar changes. “

Challenge

We need the speech community to counter the deleterious effects of DSM-5. We need to develop efficient diagnostic instruments to study risk factors, severity, and treatment effects for emotional disturbance and personality disorders. The voice, neglected so far, could play a major role and help early detection and prevention.

SPECIAL SECTION: NORMAL AND ABNORMAL EMOTIONS—THE QUANDARY OF DIAGNOSING AFFECTIVE DISORDER

emotion review

Emotion Review
Vol. 7, No. 3 (July 2015) 201–203
© The Author(s) 2015
ISSN 1754-0739
DOI: 10.1177/1754073915576689
er.sagepub.com

Normal and Abnormal Emotions—The Quandary of Diagnosing Affective Disorder: Introduction and Overview

Klaus R. Scherer
Swiss Center for Affective Sciences, University of Geneva, Switzerland

Marc Mehu
Department of Psychology, Webster Vienna Private University, Austria

Affective Sciences

Thank you for your attention!

I gratefully acknowledge contributions by many collaborators over the years, far too many to be listed on a slide

Recent research conducted by the Geneva Emotion Research Group

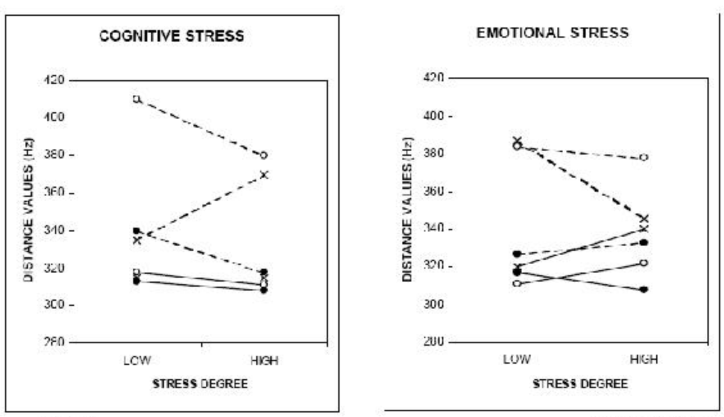
Verivox study

- Examining the role of stress and affect in automatic speaker recognition for 100 speakers from 3 language groups
- Comparing the effect of load, stress, and emotion
- Comparing acted to induced emotions

Collaborators: T. Bänziger, D. Grandjean, T. Johnstone, G. Klasmeyer

Affective Sciences

Figure 1. Distance values (difference between observed and normative formant frequencies) for cognitive and emotional stressors as a function of stress degree, with coping style and sex as parameters. (MAS = Manifest Anxiety Scale; SDS = Social Desirability Scale)



MALES
– – – FEMALES
x - anxiety deniers (MAS low; SDS high)
• low anxious (MAS low; SDS low)
° - high anxious (MAS high; SDS low)

reproduced from [5]

Affective Sciences

ACOUSTIC CORRELATES OF TASK LOAD AND STRESS

K. R. Scherer, D. Grandjean, T. Johnstone, G. Klasmeyer, and T. Bänziger

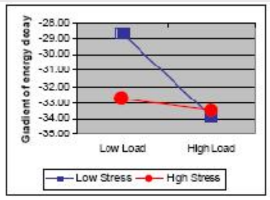


Figure 2: Differences in average gradient of energy decay between load and stress levels

homogeneous.

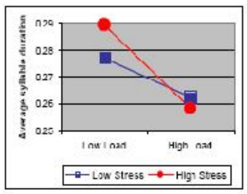


Figure 1: Differences in speech rate between load and stress levels

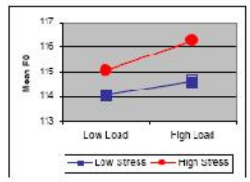


Figure 3: Differences in mean F0 between load and stress levels

Affective Sciences

Psychophysiology, 44 (2007), 827–837. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2007.00552.x

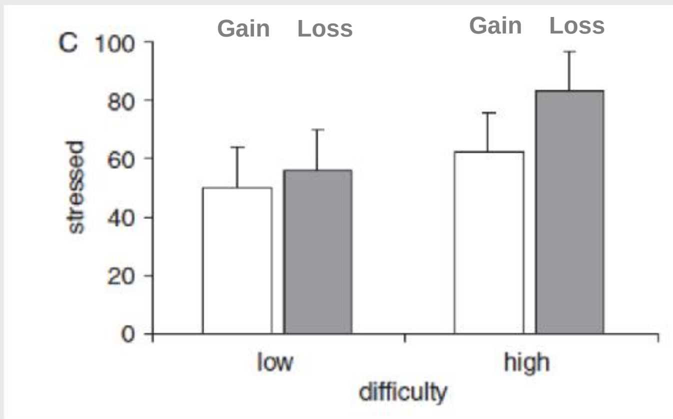
The effects of difficulty and gain versus loss on vocal physiology and acoustics

TOM JOHNSTONE,[a,b] CARIEN M. VAN REEKUM,[b] TANJA BÄNZIGER,[b] KATHRYN HIRD,[c] KIM KIRSNER,[a] AND KLAUS R. SCHERER[b]

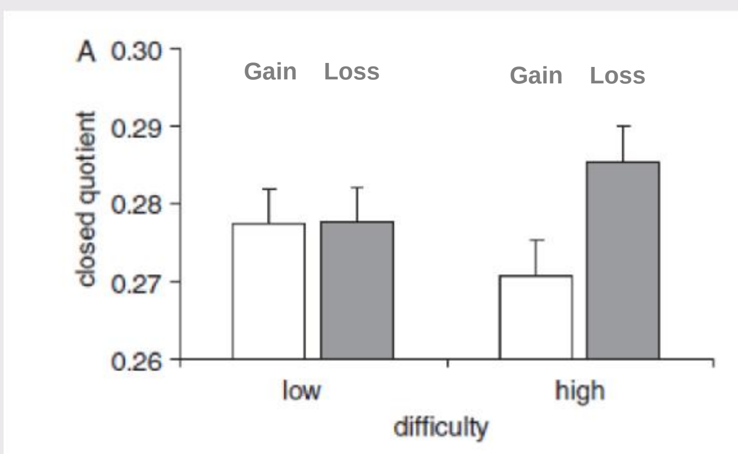
[a]School of Psychology, University of Western Australia, Perth, Australia
[b]Psychology Department, University of Geneva, Geneva, Switzerland
[c]School of Psychology, Curtin University of Technology, Perth, Australia

Affective Sciences

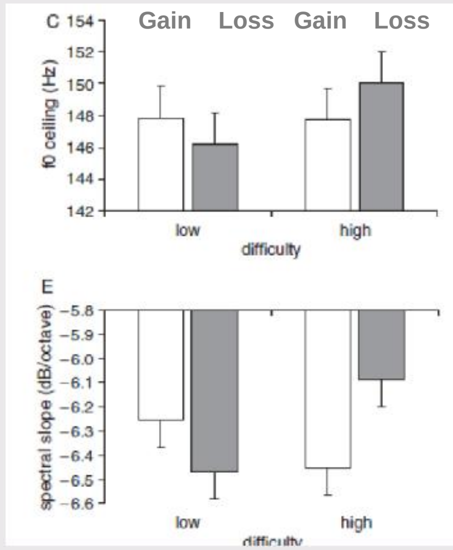
Affect ratings



EGG measures for exp. conditions



More vocal measures



Affective Sciences

Manipulating appraisal: The Xquest game

Affective Sciences

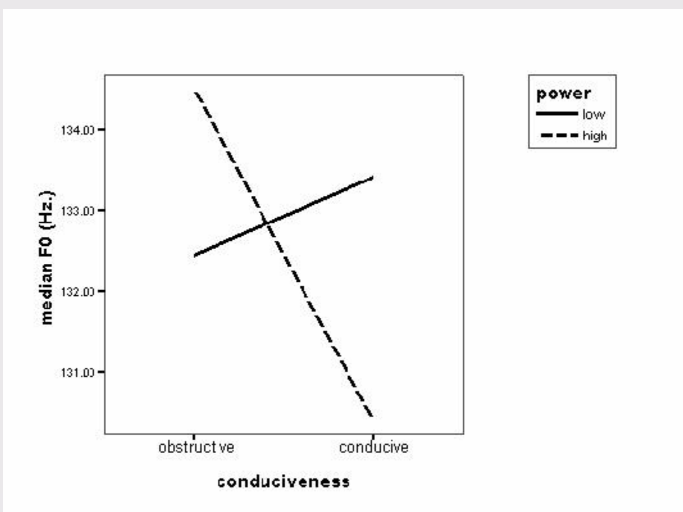
Manipulating appraisal: Design

	Low Power		High Power	
	Low Control	High Control	Low Control	High Control
Conducive	***	***	***	***
Obstructive	***	***	***	***

Power: Rate of bullets / **Control**: Wobbly cursor
Conducive: Reach next level / **Obstructive**: Loose ship

Affective Sciences

Tom Johnstone
Xquest study Vocal effects



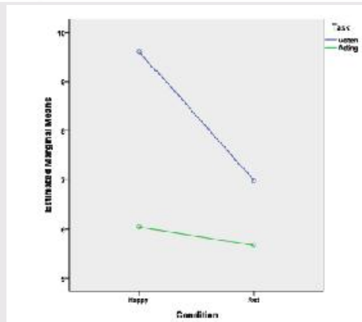
Affective Sciences

Vocal markers of emotion: Comparing induction and acting elicitation☆

Klaus R. Scherer*

Table 2
Partial eta square effect sizes for significant *Fs* resulting from a 2 × 2 repeated measures ANOVA of task and condition differences with a between test for language differences.[a]

	Within effects						Between effects
	Task	Task × language	Condition	Condition × language	Task × condition	T × C × L	Language
Max energy	0.063	0.092	0.205	0.072			0.406
FO 75			0.587				
Energy rises	0.675	0.076	0.560	0.071	0.479	0.069	
FO attack		0.079	0.242	0.120			
Low freq energy	0.076		0.406		0.078	0.228	0.535
Speech rate	0.197		0.569	0.198	0.217	0.092	0.185
Mean	0.253	0.082	0.428	0.115	0.258	0.130	0.375



Affective Sciences