The sound and the fury: emotional non-linguistic vocalizations in real life and in code

Andrey Anikin
Contents

- Non-linguistic vocalizations
- Study 1: sources of real stuff
- Study 2: real or fake?
- Study 3: emotion or vocalization?
- Study 4: synthetic stuff
- Future directions
Non-linguistic vocalizations: what?

- Non-articulated sounds (no clear phonemic structure)
- Some common to all human cultures or even all ape species (laughter)
- Some culture-specific (in-group advantage)
- Usually equated with emotion
Non-linguistic vocalizations: why?

- **Theoretical:**
  - universality of nonverbal communication
  - evolution of language

- **Practical (human-machine interaction):**
  - generate and recognize affective content without perfect natural speech processing
  - inform natural speech processing (clear parallels with prosody)
Contents

- Non-linguistic vocalizations
- **Study 1: sources of real stuff**
- Study 2: real or fake?
- Study 3: emotion or vocalization?
- Study 4: synthetic stuff
- Future directions
Study 1: sources

- YouTube videos:
  - sudden
  - unambiguous
  - powerful

- Multiple contexts, 9 emotions (ams, ang, disg, eff, fear, joy, pain, plsr, sdns)

- 260 validated sounds, ~600 total
Validation experiment: design

- Forced-choice classification (9 emotions) of 260 sounds
- 90 participants, 5 linguistic-cultural groups
- 59% overall accuracy

Which emotion? Pick one or more. Note that not all dimensions can be found in every sound!
Validation experiment: no in-group bias
Acoustic analysis

- Segmentation into syllables
- Pitch, harmonicity and other spectral measures
- Formants
Acoustic predictors of emotion
Study 1 is published as:

Contents

- Non-linguistic vocalizations
- Study 1: sources of real stuff
- **Study 2: real or fake?**
- Study 3: emotion or vocalization?
- Study 4: synthetic stuff
- Future directions
Study 2: real or fake?

- All we know about non-linguistic vocalizations is based on actor portrayals of emotion.
- Ecological validity..? Stereotypical, culture-specific..?
- Can people tell if a sound is genuine vs posed?
- Previously compared posed vs. spontaneous (actually, induced) only for laughs. How about all the rest?
Real-fake study: design

- 139 posed sounds from 6 published corpora
- 139 authentic sounds from my corpus
- 2 x 2 conditions:
  - noise vs. no noise
  - emotion shown vs. hidden
- 154 participants from 5 linguistic-cultural groups
- Accuracy 68/66% with/without noise

PROGRESS

DONE: 3/152 (2%)
CORRECT: 2/3 (67%)

Click to play!
Accuracy per corpus (with masking noise)
Accuracy per emotion
(with masking noise)

Highest for ach, ang, fear

Lowest for amu, disg, pain
Acoustic correlates of authenticity

- Similar for all emotions:
  - higher and more variable pitch
  - lower harmonicity (noisier)
  - widely and irregularly spaced syllables
- And...? Many questions remain!
“Real-fake” study is published as:

Contents

• Non-linguistic vocalizations
• Study 1: sources of real stuff
• Study 2: real or fake?
• Study 3: emotion or vocalization?
• Study 4: synthetic stuff
• Future directions
Study 3: emotion or call type?

- Classification errors suggest recognition of call type not emotion (scream, laugh, ...)
- More consistent with research on animal communication
- Questions:
  - What are human call types?
  - Are they language- / culture-specific?
  - How are they related to emotion?
Labeling study: design

- Cross-linguistic naming study
- 64 participants (English / Swedish / Russian)
- 132 sounds (name sound + emotion)
- + triad classification task (“odd one out”) to avoid linguistic labels
Semantic spaces of sound names: same 4-6 main types in EN / SV / RU
No one-to-one mapping to emotion

- skrik (scream)
- tjut (howl)
- vrål (roar)
- gnäll (whimper)
- rop (yell)
- stön (moan)
- flämtning (gasp)
- suck (sigh)
- skratt (laugh)
- fniss (giggle)
- Don't know
- gråt (cry)
- snyftning (sob)
Natural to classify call type first, then emotion

- Prefer to name sound before naming emotion:
  - start with sound name controlling for left-right,
  - faster response,
  - more consistent naming,
  - higher subjective certainty

- Non-linguistic triad task more consistent with acoustic not emotional classification
Thus...

- Only a few (4-6) clear call types recognized cross-linguistically: laughing, crying, screaming, moaning, sighing (?), roaring (?)
- Limited range of meanings, but not equivalent to a single affective state (context)
- Possibly two-stage processing, categorical perception (to be determined...)
Contents

- Non-linguistic vocalizations
- Study 1: sources of real stuff
- Study 2: real or fake?
- Study 3: emotion or vocalization?
- Study 4: synthetic stuff
- Future directions
Study 4: synthetic stuff

- No consensus on acoustic correlates of emotion (except general arousal) in human and animal research
- Untestable generalizations like Morton's rules in animal research
- Limitations of correlational studies
- **Thus**: want to synthesize sounds and vary parameters one by one
Sound synthesis: method

- Generate excitation source:
  - pulses (1/0)
  - sine waves (F0 + harmonics)
  - glottal pulses (KLSYN88)
- Filter (formants)
- Amplitude modulation, dynamics (HMM or deep learning)

~ speech synthesis, but with fewer and acoustically transparent parameters!
Work in progress: soundGen

SoundGen, ver. 3.0

Pitch at start, Hz
200
75 418 761 1,104,147 2,133 2,819 3,500

Pitch at anchor, Hz
400
75 418 761 1,104,147 2,133 2,819 3,500

Pitch at end, Hz
300
75 418 761 1,104,147 2,133 2,819 3,500

Anchor location, %
0 15
0 10 20 30 40 50 60 70 80 90 100

Pitch contour

Spectrogram

Generate

Speaker
M1
F1

Presets
Roar
Save wav
Save pars

Contents

• Non-linguistic vocalizations
• Study 1: sources of real stuff
• Study 2: real or fake?
• Study 3: emotion or vocalization?
• Study 4: synthetic stuff
• Future directions
Future directions

- Perfect soundGen, compare sine vs. pulse
- Use soundGen to test for CP of emotions and/or vocalizations, acoustic correlates
- Extract acoustic primitives (audio segmentation)
- Implement real-time recognition and production of non-linguistic vocalizations (social robotics)
- Phylogenetetic reconstruction of prosodic markers
Want to know more?

andrey.anikin@lucs.lu.se

http://cogsci.se