# Analysis and Re-synthesis of Natural Cricket Sounds Assessing the Perceptual Relevance of Idiosyncratic Parameters

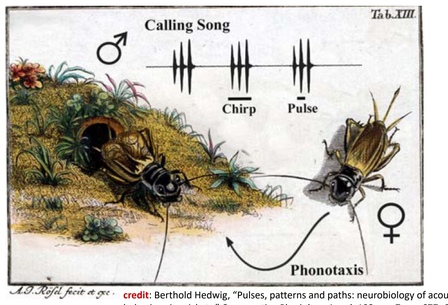

**Marco Oliveira**
**Vitor Almeida**
*FEUP-DEEC*
*University of Porto*

**João Silva**
**Aníbal Ferreira**
*INESC TEC*
*University of Porto*

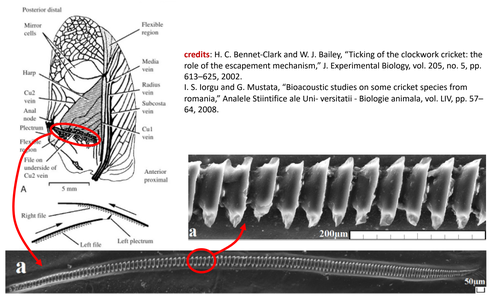

## context and motivation

- cricket sounds are very pleasant sounds
- due to their structural simplicity, cricket sounds can be easily reverse-engineered and re-synthesized with desired alterations in their defining temporal and spectral parameters
- cricket sounds can be used to design and conduct listening tests assessing the human auditory acuity to, and idiosyncratic value of, specific temporal and spectral features
- objective is to gain insight on the human auditory acuity to temporal and spectral features that are important in designing assistive technologies for dysphonic voice reconstruction

## structure of cricket sounds: stridulation

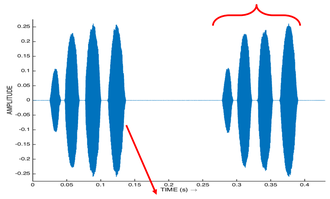



credit: Berthold Hedwig, "Pulses, patterns and paths: neurobiology of acoustic behaviour in crickets," Comparative Physiology A, vol. 192, no. 7, pp. 677–689, 2006.




credits: H. C. Bennet-Clark and W. J. Bailey, "Ticking of the clockwork cricket: the role of the escapement mechanism," J. Experimental Biology, vol. 205, no. 5, pp. 613–625, 2002.
I. S. Iorgu and G. Mustata, "Bioacoustic studies on some cricket species from romania," Analele Stiintifice ale Uni- versitatii - Biologie animala, vol. LIV, pp. 57–64, 2008.

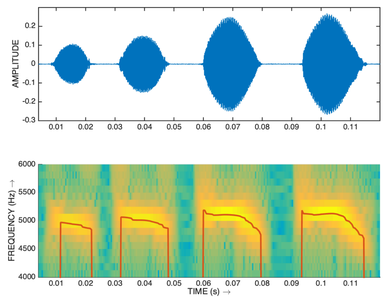
chirps containing 3-5 syllables




each syllable is a 5 kHz almost pure tone lasting around 20 ms

## structure of the *Grillus Campestris* stridulation

- each syllable exhibits a slight FM effect

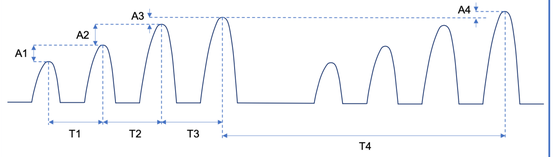



- main features of interest:
  - number of syllables per chirp
  - inter-syllable duration and amplitude variability
  - inter-chirp duration and amplitude variability
  - subtle frequency modulation within each syllable

- **to which features is the human auditory system most sensitive to in terms of detection and discrimination ?**

## cricket sound acquisition, parameters and analysis-synthesis

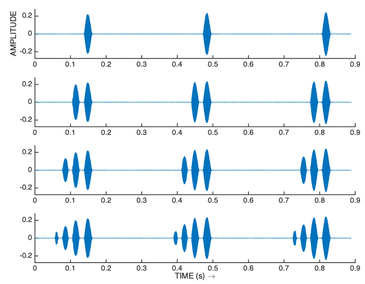
- 4 songs from 2 different crickets were recorded for analysis using the same lab conditions
- recording equipment and conditions were:
  - Sennheiser Ear Set microphone with phantom power (15 cm from cage)
  - Cakewalk UA-25EX A/D audio interface
  - Adobe Audition audio recorder/editor
  - 48 kHz sample rate and 16 bit/sample
- parameters of interest are:
  - inter-syllable time distances (T1, T2, T3) and inter-chirp time distance (T4)
  - inter-syllable amplitude differences (A1, A2, A3) and inter-chirp amplitude difference (A4)
  - frequency modulation effect within each syllable



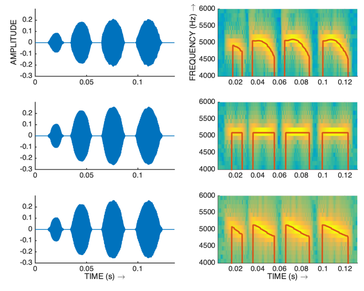

## cricket sound analysis and re-synthesis

- modified versions of the original recordings were created using a voice-oriented analysis/synthesis framework [19] such as to
  - normalize the inter-syllable/chirp time distances to their average value ( == to eliminate jitter)
  - normalize the inter-syllable/chirp amplitude differences to their average value ( == to eliminate shimmer)
  - suppress the leading 1, 2, or 3 syllables in chirps
  - modify the frequency sweeps in syllables either by flattening them to their average value (**F0fix**), or by approximating their variation to their first-order approximation (**F0var**)

- example of leading syllables suppression




- example of original and modified frequency sweeps




## objective experimental test results

- jitter and shimmer parameters (commonly used in voice analysis)

$$\text{jitter}[\ell] = \frac{1}{N-1}\sum_{k=2}^{N}\frac{|T_\ell[k]-T_\ell[k-1]|}{0.5(T_\ell[k]+T_\ell[k-1])} \qquad \text{shimmer}[\ell] = \frac{1}{N-1}\sum_{k=2}^{N}\frac{|A_\ell[k]-A_\ell[k-1]|}{0.5(A_\ell[k]+A_\ell[k-1])}$$

$\ell$ =1,2,3,4 N = number of chirps

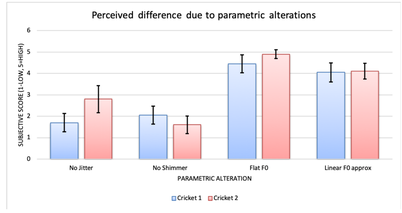
| Jitter (%) | A | B | C | D |
|---|---|---|---|---|
| T1 | 4.6 | 2.5 | 2.9 | 2.1 |
| T2 | 1.9 | 2.4 | 2.6 | 2.5 |
| T3 | 1.7 | 2.5 | 4.5 | 2.3 |
| T4 | 5.6 | 22.5 | 6.2 | 6.2 |

| Shimmer (%) | A | B | C | D |
|---|---|---|---|---|
| A1 | 14.6 | 9.2 | 10.8 | 17.2 |
| A2 | 7.6 | 5.3 | 4.8 | 3.1 |
| A3 | 3.0 | 4.0 | 3.0 | 2.8 |
| A4 | 2.1 | 1.9 | 5.1 | 3.3 |

- no consistent matches were found for the two recordings of the first cricket (A and C), or of the second cricket (B and D)

## subjective experimental test results

- listeners panel characterization
  - 20 volunteer listeners (7 FEM, 13 MAL)
  - age range is [19 - 58], average is 34.7 and std. dev. is 12.9
  - all are university students or professors/researchers
  - all have experience with speech/audio listening tests, 3 have formal musical education, none has hearing problems
  - all listeners were informed on the nature/purpose of the listening tests, and were provided with instructions and necessary material
- listening task: number of crickets
  - just by listening to the 4 recordings, half of the subjects correctly identified the number of crickets, which denotes that audible cues exist in the signals that are captured by the human ear
- listening task: number of preserved syllables in a chirp
  - comparing the original signals and the modified versions preserving a reduced number of syllables, listening test results suggest that important perceptual differences exist when the number of preserved syllables in a chirp is just 1 or 2, but not when it is 3
- listening task: plausible cricket sound pairs
  - just by listening to the 4 recordings, half of the subjects correctly identified which pair of recording corresponds to the same cricket, which confirms that audible idiosyncratic cues exist in the signals that are captured by the human ear
- listening tasks: audibility of alterations in crickets songs:
  - when jitter is reduced to zero
  - when shimmer is reduced to zero
  - when frequency sweeps are flattened or linearly approximated




## conclusions and future work

- main conclusions:
  - cricket sounds are simple signals that can be easily reverse-engineered and are suited for psychoacoustic experimentation
  - jitter and shimmer patterns in cricket songs have no idiosyncratic value and their flattening has no relevant perceptual impact
  - frequency sweep manipulations cause a high perceptual impact
- the human auditory system has a high discrimination capability regarding frequency variations even when the time scale is shorter than the average syllabic duration in speech (10-20 ms)
- the above conclusions, in addition to a recent conclusion [16] that pitch micro-variations in speech are not idiosyncratic, help the design of customized assistive technologies in voice rehabilitation
- future tests on subtle aspects of the human auditory acuity will involve altered short monosyllabic CVC speech sounds