# Do cochlear implants provide spectral envelope cues for voice gender identification?

**Kovačić, Damir; Balaban, Evan**

# DO COCHLEAR IMPLANTS PROVIDE SPECTRAL ENVELOPE CUES FOR VOICE GENDER IDENTIFICATION?

Damir Kovačić[1,#], Evan Balaban[2,3]

[1]Polyclinic SUVAG, Zagreb (Croatia), [#]Current address: Laboratory of Auditory Neurophysiology, KU Leuven (Belgium)

[2]McGill University, Montreal (Canada), [3]Hospital General Universitario Gregorio Marañón, Madrid (Spain)

## Abstract

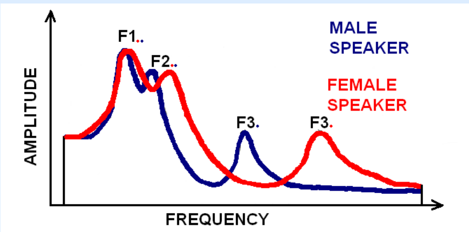
The spectral envelope of human speech contains information on speaker gender, reflecting differences in the vocal tract lengths of males and females. It is not clear whether cochlear implants (CI), which have reduced spectral resolution compared to unimpaired human hearing, are able to effectively transmit spectral envelope cues from speech sounds, and whether CI users can make use of them. A juvenile population of 41 CI users was tested using naturalistic short speech segments spoken by a variety of speakers. Stimulus output patterns of each CI device in response to vocalic /a/ segments of the speech items were also recorded and analyzed for the presence of spectral envelope cues. A majority of recorded CI devices preserved spectral envelope cues to voice gender, but subjects who were able to identify gender correctly did not appear to utilize spectral envelope information for gender identification. Future research will be required to understand whether it is possible to train CI users to utilize spectral envelope information, and to understand why do not use spectral envelope voice cues in spite of daily exposure.

## Background

Cues for voice gender recognition:

1) glottal pulse rate (GPR) -> fundamental frequency (F0)

2) spectral envelopes (SE) -> spacing between formants (F1,F2,F3...)




Current CI devices are designed to transmit F0 primarily via envelope temporal modulations of electrically stimulated pulse trains.

This study assessed the availability of spectral information for voice gender identification by CI users.

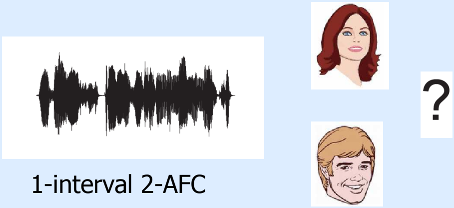
## Research strategy

1) Voice gender identification in 1-interval 2-AFC task
2) Extraction of vocalic segments (vowel /a/) from CI stimulus output patterns (CI SOPs)
3) Finding spectral envelope ratios (SERs) from CI SOPs and comparing with the SERs in acoustic signals
4) Correlation of electric SER values with the voice gender identification scores

## Methods: Subjects,stimuli & procedure

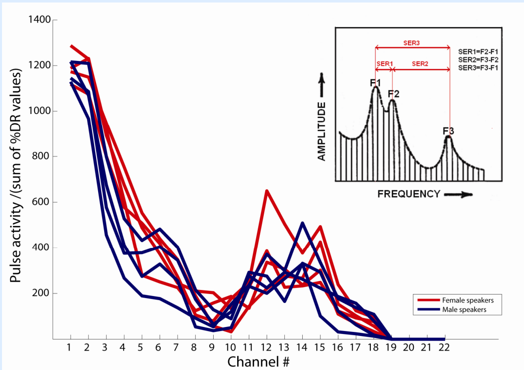
Forty-one CI subjects with Cochlear devices (20 males, 21 females; age range:5.3-18.8 years, mean age=12.3 years) together with 15 hearing children (8 males, 7 females, age range 6.7-10.6 years, mean age 9.3 years)

Short vocalic segments (vowel /a/) were extracted from 2-sec speech items spoken by 40 different speakers (20 male and 20 females)




1-interval 2-AFC

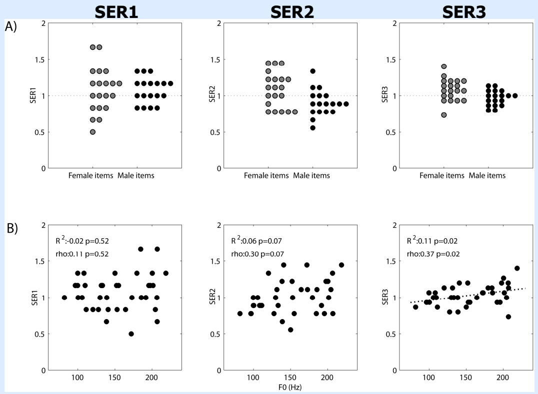
## Methods: Stimulus output patterns




## Methods: Extracting SER values

1) Wide-band long-term average spectrums (LTAS) with a 125-Hz bandwidth were applied to calculate the frequency positions of the first three peaks (formants), as well as the four valleys associated with them, yielding 7 data points (3 peaks and 4 valleys) for each speech item.
2) For each subject, the electrode positions (channels) for each of these 7 points were assigned according to the implant frequency allocation tables.
3) The SER values were obtained by normalizing the electrode distances between speech formants with the electrode distance of the speaker with the median value.

## Results: Acoustical availability of SER




*A) Gender distribution of the SER values in the original (acoustic) signals. B) Relationship between the F0 of the speakers and their corresponding SER values in the original (acoustic) signals. Column division: Left column shows F2-F1 related measures (**SER1**), the middle column F3-F2 related measures (**SER2**), and the right column F3-F1 related measures (**SER3**).*

## Results: Electrical availability of SER

Three measures were used to assess the availability of spectral envelope cues in output provided by the CI devices:

1) Regression between acoustic SERs and electric SERs : All subjects (except 1 for SER 2 and 2 for SER 3) had CI devices that preserved SER cues to a significant degree.
2) Gender difference in electric SER values: electric SER2 were significantly different in 26 out of 35 subjects, with an additional 2 subjects being marginally significant ($p=0.06$). Eleven subjects out of 35 had significant differences in the SER3 measure (spectral envelope ratios of F3-F1 differences), with an additional 3 being marginally significant. In contrast, the SER1 (F2-F1) measure did not yield any gender-dependent differences.
3) Pearson product-moment correlations between the F0 of the speech item and the SER values: 25 out of the 35 subjects demonstrated a significant correlation for SER2; (range of the correlations: 0.32 - 0.61), and 13 subjects showed significant correlations for SER3 (range: 0.34 - 0.68).

## Support

*Support provided by the Croatian Ministry of Science, Education and Sports Grant (207-0000000-2293) to DK, NSERC 298612 and CFI 9908 to EB*

## Results: Relationship between the availability of SER cues and gender identification performance

1) The mean difference for SER-related information between female and male speech items reflected in cochlear implant output was significant for all three SER measures (only SER2 and SER3 were significantly different for the acoustic signals). The mean SER1 difference was -0.03±0.01 (difference from zero: $Z=-2.21$, $T+=180$, $p=0.03$, $n=35$, Wilcoxon signed rank test); the mean SER2 difference was 0.23±0.01 (difference from zero: $Z=-5.16$, $T+=630$, $p<0.0001$, $n=35$, Wilcoxon signed rank test); and the mean SER3 difference was 0.13±0.01 (difference from zero: $Z=-5.16$, $T+=630$, $p<0.0001$, $n=35$, Wilcoxon signed rank test).

2) Comparison of the SER cue distances between performing and non-performing subjects: no systematic differences were found (SER1: mean for performing subjects= -0.015 ($n = 15$); mean for non-performing subjects= -0.045 ($n = 20$), $Z=1.28$, $p=0.2$, Mann-Whitney U-test; SER2: mean for performing subjects= 0.225 ($n=15$); mean for non-performing subjects=0.232 ($n=20$), $Z=0.12$, $p=0.9$,Mann-Whitney U-test; SER3: mean for performing subjects= 0.129 ($n = 15$); mean for non-performing subjects=0.121 ($n = 20$), $Z=-0.15$,$p=0.88$, Mann-Whitney U-test).

3) There were no significant differences in the amount of variance explained by linear regressions between F0 and SER values for performing and non-performing subjects (SER1: $R^2$(performing subjects) = 0.716 ($n = 15$), $R^2$ (non-performing subjects) = 0.726 ($n = 20$), $Z=-0.4$, $p=0.69$,Mann-Whitney U-test; SER2: $R^2$ (performing subjects) = 0.58 ($n = 15$), $R^2$ (non-performing subjects) = 0.58 ($n = 20$), $Z=-0.4$,$p=0.69$, Mann-Whitney U-test; SER3: $R^2$ (performing subjects) = 0.69 ($n = 15$), $R^2$ (non-performing subjects) = 0.72 ($n = 20$), $Z=-1.02$, $p=0.31$, Mann-Whitney U-test).

4) There was also no general difference in SER values between correctly and incorrectly identified items in performing subjects. Only 2 out of 15 showed larger SER1 values for correct items as compared to incorrect ones (Subject CI26: mean SER1 in correct responses = 1.2069 ($n = 21$), mean SER1 in incorrect responses = 0.9078 ($n = 6$), $Z=-2.73$ $p = 0.006$; Subject CI36: mean SER1 in correct responses = 1.16 ($n = 27$), mean SER1 in incorrect responses = 0.85 ($n = 7$), $Z=-2.31$ $p = 0.02$). Only a minority of CI subjects showed consistent differences between CI SER values for items that subjects called "male" and items that subjects called "female" (7 out of 15 for the SER1 comparison and 3 out of 15 for the SER2 comparison).

## Conclusion

Significant gender-related differences in all three spectral envelope measures were present in the electrical output signals of cochlear implant devices worn by members of the present study population. However, CI subjects who are able to discriminate male and female voices did not appear to use SER cues to identify gender.