Linguistic assessment tools for the Digisonic® Dual electric-acoustic speech processor[1]

Willemijn Heeren, Bart Vaerenberg, Martine Coene, Kristin Daemers, Paul J. Govaerts, Andrei A. Avram, Anna Cardinaletti, Geert De Ceulaer, Luca Del Bo, Steven Gillis, Alexandru Pascu, Johan Rooryck, Karen Schauwers, Vincent van Heuven, Francesca Volpato

The FP7-SME-222291-DUAL PRO Consortium (coordination: OTOCONSULT BVBA)

Introduction.

It is well known that cochlear implants enable profoundly deaf patients to reach high levels of speech intelligibility. They are, however, suboptimal for the perception of melody in music and speech. The reason for this is that implants are conceived to code for the mid and high frequencies of sound, where phonemic information is contained. Low frequencies contain information related to tonality, timbre, etc. ,but cochlear implants do not code low frequencies well.

In new hearing rehabilitation strategies, such as electric-acoustic stimulation (henceforth EAS), standard amplification via hearing-aid technology is combined with electric stimulation via cochlear implant technology. This strategy aims at exploiting the complementary benefits of acoustic and electric stimulation of the low and high frequency auditory cues, respectively. Potential candidates for EAS are hearing impaired individuals with residual hearing at low frequencies and a severe hearing loss at high frequencies, who do not benefit from classical hearing aids.

Low frequency auditory cues are essential in pitch perception, the primary acoustic cue to speech prosody. Prosody covers the properties of speech that cannot be explained by the intrinsic properties of the speech sounds that were uttered. It includes acoustic variation in e.g., fundamental frequency, duration, and intensity. Such variation can carry meaning, and therefore is crucial for spoken communication.

In this report, we describe the development and use of a new module of the Auditory Speech Sound Evaluation test (A§E ®, Govaerts e.a. 2009), as part of the FP7 European research project "DUAL PRO", which aims at assessing the perception of speech prosody across languages by means of discrimination and identification tasks. By means of this prosody test battery the use of current generation of cochlear implants and classical hearing aids can be assessed, as well as that of new, hybrid EAS strategies.

Method.

The use of prosodic cues varies between languages. In tonal languages, such as Cantonese or Mandarin, the fundamental frequency (F0) serves as the major acoustic cue for lexical tone contrasts (Shih 1988), though other acoustic properties such as intensity and duration may also contribute to their discrimination (Whalen and Xu 1992). In the typologically different languages that are part of the "DUAL PRO" project (Dutch, Italian, Romanian), prosody contributes to processing of semantics, syntax, discourse structure and also paralinguistic information.

In the creation of a prosodic assessment tool we aim to define the minimal set of acoustic test items that fully covers the F0 variation of linguistic functions. An analysis of the Dutch, Italian and Romanian prosodic systems has shown that both similarities and differences are found as to functions and forms. As for the similarities, all three languages for instance discriminate between lexical items through stress placement, and discriminate between statements and (certain types of) questions by falling versus rising boundary tones. As for the differences, inflection in Dutch is not determined by stress placement, whereas it is in Italian and Romanian, and Italian, as opposed to Dutch, make less use of prosodic means for conveying information structure. Romanian has syntactic structures that require specific use of nuclear stress placement, a phenomenon not found in Italian or Dutch.

Given these restrictions on the acoustic realization of prosodic functions across those three languages, a language-independent approach to stimulus design was opted for, motivated by the acoustic forms known to be relevant in a large number of languages. Hence, the test battery contains two linguistic functions of pitch movements: (i) sentence intonation, i.e. clause typing by marking phrases as statements or questions with a pitch movement on the sentences' final syllable (van Heuven & Haan 2000 a.o.); and (ii) lexical ("word") stress, i.e. the differentiation between word meanings of

---

[1] This research was funded by the EC within the 7th Framework (FP7-SME-22291 DUAL PRO, coordination: Otoconsult)

sound sequences with the same segmental make-up, but with the pitch accent on different syllables (Cutler 2007 a.o.).

Materials.

For both tests the segmental make-up of the stimuli consists of highly frequent phonemes in the three languages under analysis. Moreover, all speech sounds used are voiced in order to allow the overlay of an uninterrupted pitch contour. As a result, the following set of six CV syllables was selected (/ma/, /mi/, /mu/, /na/, /ni/, /nu/). To be able to make any combinations of these syllables a diphone and triphone grammar was constructed and recorded with a female, native speaker of Dutch.

For the Sentence Intonation test, a same-different (AX) discrimination task was chosen. The listener has to indicate whether two sentences are either exactly the same (AA) or different (AB), i.e. differing in the presence of a final rise on one of the sentences. Each sentence was modelled by a sequence of four to six syllables from the abovementioned set. Over each sentence a pitch contour was drawn with a fixed pitch accent on the second syllable and a variable-sized pitch rise on the final syllable, varying from a flat ending that remained at the 200 Hz baseline to a 408 Hz rise. The location of the maximum in the final rise was set at about 100 ms after syllable onset (Figure 1a).

As for the Word Stress test, a three-category identification task was chosen. Listeners are asked to indicate on which of the three syllables they hear a pitch accent. The possible sizes of the accent are taken from the same series as used for the clause typing test. The location of the maximum pitch in the syllable was set to 100 ms after syllable onset (Figure 1b).

An acoustic validation of the stimulus wave files was performed to ensure that the crucial information in the linguistic stimuli is contained in the low frequencies, i.e. that the stimuli only vary in pitch, but not in duration or loudness. Second, both tests were doubled by a low-pass filtered version. Low-pass filtering only maintains the frequencies in which the crucial information is contained, while suppressing the segmental information of the speech. All stimuli (filtered and non filtered) were presented in an adaptive algorithm until the Just Noticeable Difference was found.

Participants.

The test was validated on 90 hearing listeners (30 for each language background, e.g., Dutch, Italian and Romanian). All were between 18 and 53 years old and sex was evenly distributed in the sample. Participants were selected not to be students of linguistics or any language. Additionally, participants with a history of hearing problems or dyslexia were excluded. Normal hearing was screened through tonal audiometry: participants did not show a hearing loss of over 20 dB on the frequencies 125 through 8000 Hz. For the hearing impaired group, pilot testing on the unfiltered tests was done with 6 adult CI users and 4 subjects with classical hearing aids. Some of the tests were also administered to 4 cochlear implanted children aged 12 to 14 years.

Results.

For the hearing controls (Figure 2, Table 1), the median JND across all listeners was 12 Hz and 14 Hz for the non-filtered and the filtered Sentence Intonation tasks, respectively. For the Word Stress task, the median JND was 16 Hz for the unfiltered and 8 Hz for the filtered version. All 6 adult CI users obtained JNDs that were not within normal range. Four of them even obtained a JND of over 200 Hz on the Sentence Intonation task. On the other hand, 3 out 4 subjects with classical hearing aids obtained results that are within normal range. The observed differences between the two hearing impaired populations are less prominent for the Word Stress test. Additional pilot testing on cochlear implanted children showed that, despite intra-subject variation, the results are mainly in line with those obtained for the adult CI users, i.e. most JNDs are not within normal range. For 2 out of 4 children it was not even possible to obtain a JND on particular tests, meaning that they were unable to detect pitch rises that were as large as 350 Hz. Only one child obtained a JND on the filtered version of the Sentence Intonation test of 24 Hz, which is within normal range. Interestingly, this child's audiogram revealed residual low-frequency hearing in the left, non-implanted ear (Figure 3). This points in the direction of the particular role of low-frequency hearing in the perception of intonation.

Discussion.

Residual low-frequency hearing provides temporal fine structure that is not conveyed by current CI-processors. It is therefore anticipated that the additional acoustic stimulation within the Digisonic® Dual processor will substantially improve prosodic perception by providing pitch details of speech. Based on insights from developmental psycholinguistics in which early language development is taken to build on prosodic perception, EAS is expected to indirectly enhance spoken language development of young deaf children.

Conclusion.

The spectral resolution of current CI-processors has been optimized for speech sound discrimination, providing only restricted information about voice pitch. Yet, the perception of language prosody crucially depends on the perception of pitch. Therefore, CI users with electric stimulation alone have difficulties in discriminating and identifying sentence intonation and word stress patterns. It is expected that hybrid electric-acoustic stimulation will improve the access to low-frequency information in the speech signal and as such, enhance prosodic perception in deaf individuals.

REFERENCES

Cutler, A. 2007. Lexical Stress. *The Handbook of Speech Perception*. D.Pisoni and R.Remez, Blackwell Publishing: 264-289.

Govaerts PJ, Daemers K, Yperman M, De Beukelaer C, De Saegher G, De Ceulaer G. 2006. Auditory speech sounds evaluation (A§E®): a new test to assess detection, discrimination and identification in hearing impairment. *Cochlear Implants International* 2006; 7(2): 97-106.

Shih, C.-L. 1988. Tone and Intonation in Mandarin. *Working Papers of the Cornell Phonetic Laboratory* 3, 83-96.

van Heuven, V.J & J. Haan 2000. Phonetic correlates of statement vs question intonation in Dutch. *Intonation: analysis, modelling and technology*. A. Botinis. Dordrecht/Boston/London: Kluwer:119-144.

Whalen,D.H., Xu,Y. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25-47.

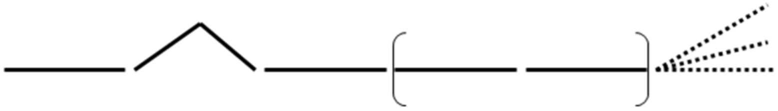

Figure 1a – Model of possible sentence pitch contours

Figure 1b – Model of possible pitch movements on the first, second and third syllable

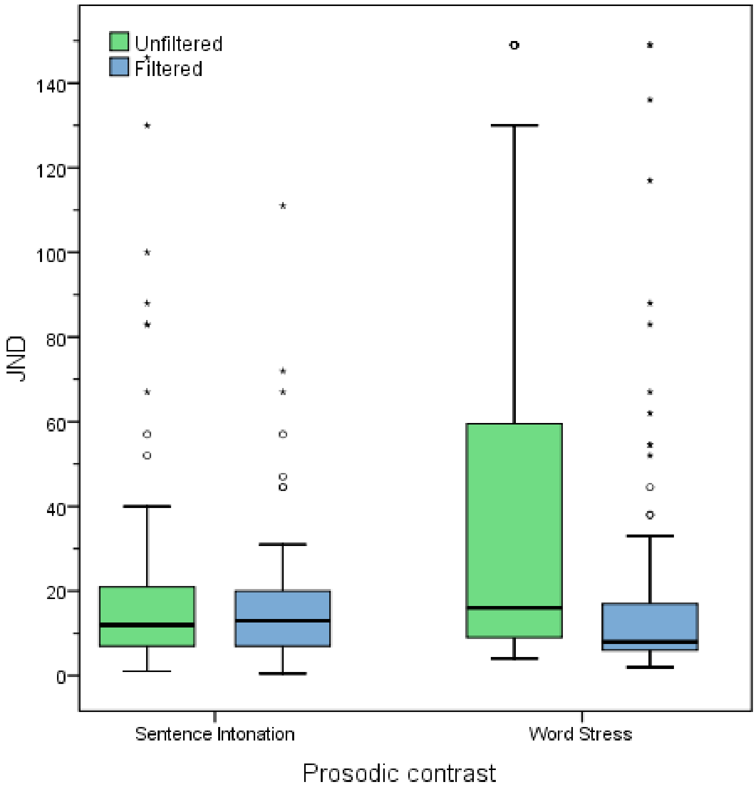

Figure 2 – JNDs for the unfiltered and filtered version of the two prosodic tests for 90 normal hearing listeners

Test	**Median JND**		
	Dutch	**Italian**	**Romanian**
Sentence Intonation	7.0	14.0	15.0
Sentence Intonation LPF	13.0	18.0	10.5
Word Stress Pattern	12.0	47.0	16.0
Word Stress Pattern LPF	7.0	10.0	10.5

Table 1 – Median JNDs for the four prosodic tests by language background

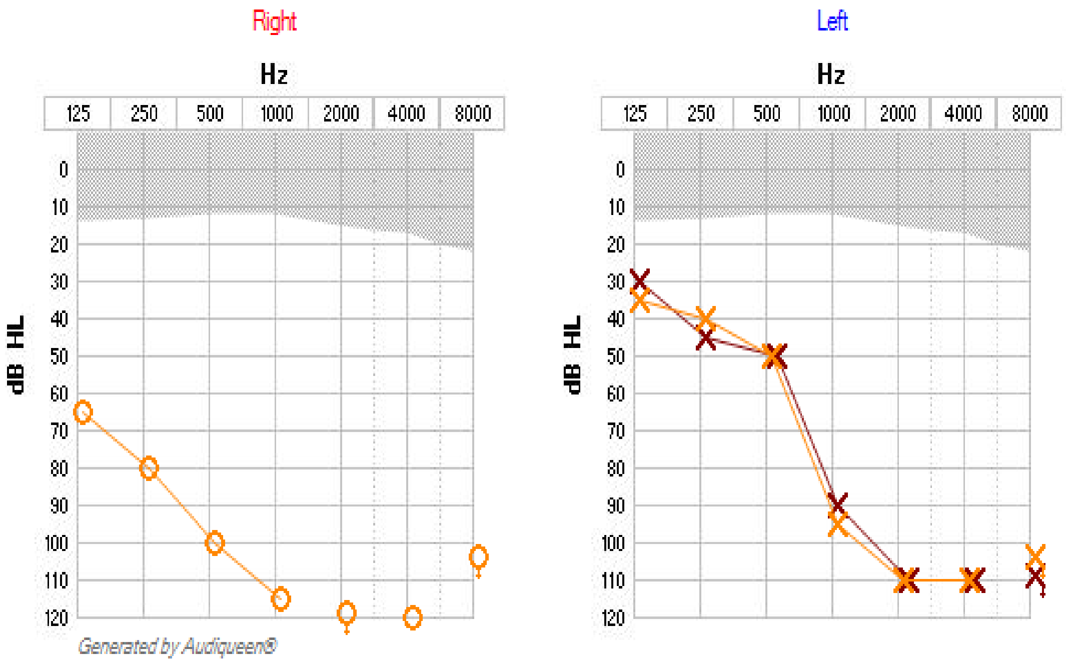

Figure 3 – Audiogram of the right and left ear of the only child obtaining a JND within normal range on the sentence intonation task. The left non-implanted ear clearly shows residual hearing in the lowest frequencies (125-500 Hz).