Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing" J. Acoust. Soc. Am., **88**, 1725-1736.

Grose, J. H., Mamo, S. K., and Hall, J. W., 3rd (**2009**). "Age effects in temporal envelope processing: speech unmasking and auditory steady state responses" Ear. Hear., **30**, 568-575.

Gustafsson, H. Å., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise" J. Acoust. Soc. Am., **95**, 518-529.

Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: the effects of temporal resolution" J. Acoust. Soc. Am., **119**, 3097-3108.

Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006a**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure" Proc. Natl. Acad. Sci. USA, **103**, 18866-18869.

Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006b**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception" Int. J. Aud., **45**, 487-495.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference" J. Acoust. Soc. Am., **125**, 457-468.

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people" J. Acoust. Soc. Am., **103**, 577-587.

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing" J. Acoust. Soc. Am., **125**, 3328-3345.

Stuart, A. (**2008**). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children" J. Am. Ac. Aud., **19**, 135-146.

Stuart, A., and Phillips, D. P. (**1996**). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing and presbyacusic listeners" Ear. Hear., **17**, 478-489.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly" J. Speech. Lang. Hear. Res., **35**, 1410-1421.

Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners" J. Acoust. Soc. Am., **111**, 401-408.

Prosody perception in simulated cochlear implant listening in modulated and stationary noise

DAVID MORRIS

Institute for Scandinavian Studies and Linguistics, University of Copenhagen, Njalsgade 120, DK-2300, Denmark

Cochlear Implant (CI) listeners can do well when attending to speech in quiet, yet challenging listening situations are more problematic. Previous studies have shown that fluctuations in the noise do not yield better speech recognition scores for CI listeners as they can do for normal hearing (NH) listeners. The aim of this experiment was to investigate the ability of simulated CI listeners in a prosodic task, where F0 Just Noticeable Differences (JND) were measured in modulated and stationary background noise.

A nonsense sentence was created from a recording with durations and overall contour derived from non-scripted Danish speech. The F0 temporal midpoint of the initial syllable was varied stepwise in semitones. Competing signals of modulated white noise and speech shaped noise at 0 dB and 12 dB SNR, were added to the tokens prior to 8-channel noise-excited vocoder processing. Stimuli were presented diotically to 8 NH listeners in a 2AFC task. A question/statement identification experiment was also performed. Results from the JND experiment indicate a significant noise effect for the modulated noise condition at the lower SNR.

INTRODUCTION

It is commonly accepted that CI processing schemes can provide the cues needed for speech recognition in quiet. Other listening situations are more problematic for CI listeners, for instance, speech in competing noise. In a study designed to evaluate the release from masking in fluctuating noise, Nelson et al., (2003) found that CI listeners and simulated CI listeners performed very poorly in a sentence identification task, even at favorable signal-to-noise ratios (SNR). They reported that there was no difference in the identification scores of CI listeners in steady noise and when noise modulation frequencies were between 2-4 Hz, or those approximately corresponding to word and syllable rates. Qin and Oxenham (2003) also found no significant improvement in Speech Reception Threshold (SRT) values between modulated and steady-state noise maskers in 24, 8 and 4 channel CI simulations. These studies highlight the deficit in the ability of CI listeners to 'listen in the gaps' of a temporally modulated masker, an ability that provides release from masking in NH listeners (Festen and Plomp, 1990; Howard-Jones and Rosen, 1993).

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

CI listening can be simulated by processing stimuli with an envelope-excited vocoder, after which it can be presented to NH listeners. The extracted envelope of the speech can be modulated with sine wave or noise carriers. The most appropriate simulation model for studying the perception of F0 is the noise vocoder as it eliminates fine-structure cues that can be preserved in the sine wave vocoder (Souza and Rosen, 2009). The number of vocoder processing channels has been shown to be a parameter that impacts on performance. Dorman et al., (1997) found that 8 channels were necessary to approach asymptotic performance in vowel recognition tasks.

In a favoured model of prosody in typical Danish sentences, F0 variation is mediated by the superposition of stress group patterns onto an intonation contour (Grønnum, 2007). The stress group pattern is a predictable and recurrent entity which has a bearing on the intonation contour of the sentence while it is also interrelated with linguistic stress (Thorsen, 1978). The intonation contour itself is communicatively relevant as it can impart attitudinal value to an utterance. For instance, it can alert a listener as to whether an utterance is a statement or a question.

In order to assess the ability of simulated CI listeners to perceive prosody in the presence of noise, a study was performed with a nonsense sentence and also with question and statement sentences. These studies assessed whether noise type or presentation SNR affects the performance of CI listeners on prosody perception.

METHOD

Stimulus preparation: JND Experiment

A male speaker recorded the isolated syllable [ba]. Recordings were made in a sound-treated studio with a quality microphone (Sennheiser MKH40 P48) and a 32-bit digital audio recorder (Sound Devices 722). The syllable was edited so that it was consecutively repeated three times. The duration of each segment was manipulated so that it was equivalent to Tøndering's (2008) average findings for a large corpus of non-scripted Danish speech (as per the durations in Figure 1).

The initial and final values of the overall intonation contour slopes for the three syllables were adjusted in Praat so as to resemble a standard declarative statement. The three syllables fell by six semitones (as per Grønnum, 2007). Also the F0 of the first syllable was pitch shifted in consecutive semitones up to 12 semitones (see Figure 1).

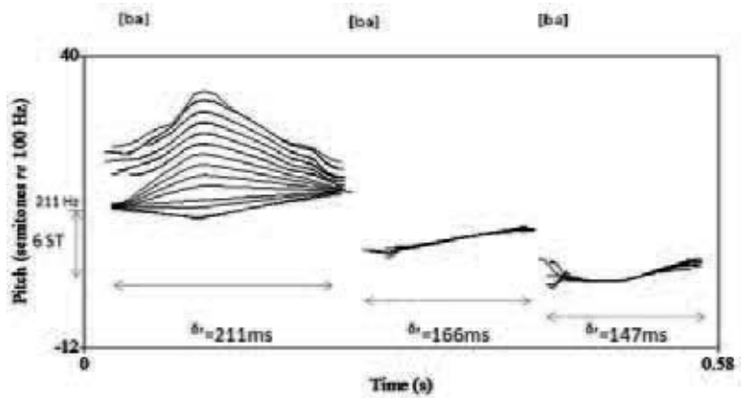

Fig. 1: Experimental Stimuli – showing fundamental frequency with pitch shifting on the initial syllable [ba] and duration modifications.

All pitch shifted stimuli were then added to the unmodified original with a 500 millisecond pause in between, so that the unmodified stimuli preceded the modified.

Stimulus Preparation: Question/Statement Identification Experiment

Stimuli for the question and statement tokens consisted of 8 sentences from Lieberman and Michaels (1962), 5 sentences from Møller and Post (1997) and 17 original sentences containing either three or four iambs. The stimuli were recorded 3 times as a statement and 3 times as a question, by both a male and a female speaker. Both speakers had Danish as their mother tongue. The best of the 3 recordings was then chosen for editing. This yielded 120 speech tokens, that is, 10 questions and 10 statements per noise condition spoken by both the male and female speaker.

CI simulation

The Tiger CIS software was used to simulate cochlear implant listening. In this software the input signals were bandpassed into 8- frequency channels. The high pass filter settings were at 200 Hz and the low pass at 7000 Hz with a filter slope of 24 dB per octave. White noise was selected as the carrier type and analysis and carrier filter settings were based on the Greenwood map. The temporal envelope in each channel was extracted via half wave rectification and low-pass filtering. The temporal envelope was then modulated with a white noise carrier. The modulated noise bands were then summed and the speech level adjusted. Noise was added at 0 dB SNR and 12 dB SNR to the JND stimuli and at 6 dB SNR to the question/statement stimuli.

Noise

Noise was added to the JND and question/statement stimuli prior to vocoder processing to best simulate listening in background noise. The modulated noise was generated with Adobe Audition v.1. An envelope filter with spline curves was used to modulate a white noise carrier at a rate of 36 Hz. The modulation depth was approximately 0.7. The speech shaped noise was stationary and was from the noise options available in the CI simulation software.

Participants

Participants were eight staff members from the University of Copenhagen. These were three men and five women with a mean age of 36 years (range 29-45 years). All reported normal hearing, and had Danish as their mother tongue.

Procedure

Stimuli were presented diotically to the participants who were instructed to judge whether the stimuli pairs were similar or different. Stimuli were presented via quality headphones (Sennheiser HD 590). Repetitions were permitted. Stimuli were presented in decreasing semitone steps in a 1-up, 1-down procedure starting at the 12th semitone modification. The criteria for a JND were set at 5 reversals per condition. The JND was then calculated as the mean of the total reversals. The question/statement identification task was also performed during the same testing session. The sentences were randomised prior to presentation. Together both tasks took approximately 55 minutes to complete.

RESULTS

Figure 2 presents the mean JNDs for the quiet and the noise conditions.

A paired samples test of the quiet and the modulated 0 dB SNR condition revealed a significant difference [t(7)= -2.512, p= 0.04]. There was no significant mean difference between the modified and the steady-state noise conditions, also when results were grouped together across SNRs.

To investigate whether SNR affects JND, the mean scores in both noise types with the 0 dB SNR and 12 dB SNR presentation levels were compared but significance was not reached. Despite the difference in mean JNDs between the modulated noise at the two presentation SNRs (0 dB SNR, M= 9.43; 12 dB SNR, M=8.7), a t-test revealed that this was not significant.

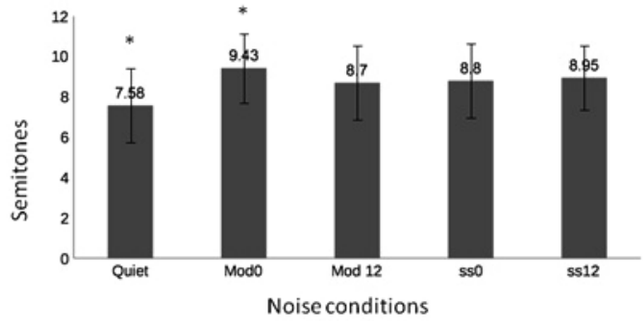

Fig. 2: Mean JNDs in semitones for the quiet and four noise conditions: modulated 0 dB SNR; modulated 12 dB SNR; speech shaped 0 dB SNR; and, speech shaped 12 dB SNR. Error bars represent the first standard deviation from the mean.

The question/statement identification task produced a data set with a wide range of correct identification scores. These extended from 50 to 81% (see Figure 3). The results from two listeners were omitted as they scored at chance levels at most conditions, possibly due to the relatively long test time and attention difficulties.

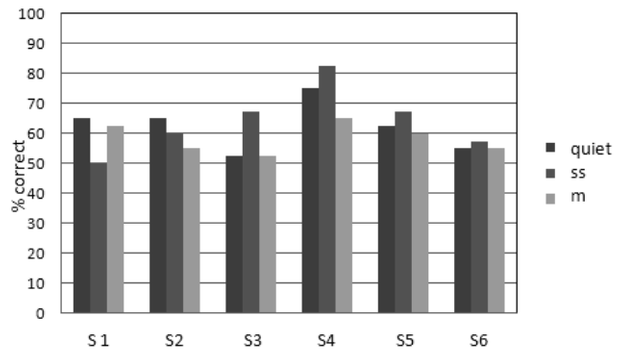

Fig. 3: Individual percentage correct scores for the question/statement identification task across the quiet, speech shaped and modulated noise conditions.

DISCUSSION

The results from this study show the inferior ability of simulated CI listeners to detect the changes in F0 necessary for the perception of prosody. Similar result patterns for F0 JND in quiet are reported in Meister et al. (2011) and for pure tones

in the 200 and 400 Hz JND results of Gfeller et al. (2002) and the UW-CAMP pitch test results of Uchanski (2011). While the introduction of noise did result in higher mean JNDs, it could have been expected that performance might have deteriorated more markedly.

Notable from the results in the JND study is the absence of a noise type effect and a presentation SNR effect, despite the relatively broad difference between the SNRs used. In the case of the modulated noise, this could be due to the relatively high amplitude modulation used. This combined with the short durations of the segments of the nonsense sentence and the spectral degrading entailed in 8-channel vocoding could have left the F0 variations imperceptible.

Manipulation of the syllable [ba] resulted in a rising-falling F0 event that extended over the first word which was also the first syllable. Identification of this 'crowned' pattern of excursion is thought to be crucial for identifying and discriminating between microprosodic features in related languages, for instance, word accents in Swedish (Bruce, 1998). The position of the manipulated segment in relation to the sentence could also be expected to play a role in the identification of F0 variation. This was reported by Meister, et al., (2011) when they observed that sentence stress was harder to identify when placed on the middle (verb) position of a subject-verb-object sentence than when placed on the final (object) position. This finding was attributed to either co-occurring intensity changes, the inclusion of the article preceding the subject and the object, or the grammatical function of the words. In light of these results it would be of interest to test F0 variations at other word positions in a nonsense sentence, to see if position is a relevant factor. Moreover, this could overcome a procedural problem encountered in this experiment. When manipulating the first syllable of the nonsense sentence in the present study, the F0 variations began at a relatively high initial F0. Further investigation within this nonsense sentence paradigm would do well to vary F0 from lower initial values.

The performance improvement in the question/statement identification in the speech shaped noise condition in subjects 3, 4, 5 and 6, is also noteworthy. This improvement could warrant investigation with an adaptive SNR procedure and an identification task where question sentences were less contrived.

CONCLUSION

A significant noise effect was found when modulated noise was introduced during a JND task where F0 was varied at the initial word of a nonsense Danish sentence. Neither presentation SNR nor noise type was a factor that significantly affected the JND results at the SNRs tested.

Acknowledgements and Caveat

Thanks to Preben Dømler, Hanne Trebbien Daugaard and Holger Juul for their assistance in recording the stimuli used in these experiments. This was an initial pilot study carried out as part of a Ph. D. project on CI listeners' perception of prosody in noise.

REFERENCES

Bruce, G. (**1998**). "Allmän och svensk prosodi" [General and Swedish prosody] in Praktisk Lingvistik 16. Reprocentralen, Lund.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs" J. Acoust. Soc. Am., **102**, 2403-2411.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing" J. Acoust. Soc. Am., **88**, 1725-1736.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., and Stordahl, J. (**2002**). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults" Cochlear Implants Int., **3**, 29-53.

Grønnum, N. (**2007**). "Rødgrød med Fløde. En lille bog om dansk fonetik" [Red pudding with cream. A little book on Danish phonetics] Akademisk Forlag.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard'noise" J. Acoust. Soc. Am., **93**, 2915-2922.

Lieberman, S., and Michaels, T. (**1962**). "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech" J. Acoust. Soc. Am., **34**, 922-927.

Meister, H., Landwehr. M, Pyschny, V., Wagner, P., and Walger, M. (**2010**). "The perception of sentence stress in cochlear implant recipients" Ear Hear, **32**, 1-9.

Møller, V., and Post, I. (**1997**). "Voksne med cochleaimplantat" [Adults with cochlear implants] Materiale til iagttagelse.

Nelson, P. B., Jin, S., Carney, A. E. and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners" J. Acoust. Soc. Am., **113**, 961-968.

Qin, M. K., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers" J. Acoust. Soc. Am., **114**, 446-453.

Souza, P. and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech" J. Acoust. Soc. Am., **126**, 792-805.

Thorsen, N. (**1978**). "An acoustical analysis of Danish intonation" J. Phonetics **6**, 151-175.

Tøndering, J. (**2008**). "Skitser af prosodi i spontant Dansk" [Sketches of prosody in spontaneous Danish] Ph.D. thesis.

Uchanski, R. M. (**2011**). "Listening to talkers and musical pitch" Presentation at Conference on Implantable Auditory Prostheses. July 2011, Asilomar, California.