Published in final edited form as:

Hear Res. 2007 June ; 228(1-2): 230–231.

The Effect of Frequency Estimation on Speech Recognition Using an Acoustic Model of a Cochlear Implant ^a

Chandra S. Throckmorton^a, M. Selin Kucukoglu^b, Jeremiah J. Remus^c, and Leslie M. Collins^d

Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

In a previous publication, the potential for improving speech recognition by increasing frequency resolution within each channel of a cochlear implant speech processing algorithm was investigated as an alternative to increasing the number of channels (Throckmorton et al., 2006). The study was conducted with normal-hearing subjects listening to speech tokens processed through an acoustic model of a speech processing algorithm with the intent to investigate whether the proposed method improved speech recognition. The results suggested that with as few as two discrete frequencies per channel, an increase in speech recognition performance was possible. Although several caveats must be considered and are discussed at length by Throckmorton et al. (2006), these results suggest the possibility that cochlear implant recipients might derive benefit from increased frequency resolution within channels when increasing the number of channels is not an option.

In the acoustic model, rather than presenting a constant frequency for each channel as is traditionally done (e.g. Dorman et al., 1997), the frequency content of each channel was estimated and a presentation frequency was chosen based on this estimate. For computational simplicity, a short-time Fourier transform (STFT) was used to analyze each 2 ms window of speech and select the closest 1 of N predefined presentation frequencies in each channel. However, given the relatively low frequency content in some of the channels, a 2-ms window duration does not sample enough of the signal to provide an accurate estimate of the frequency spectrum. The authors originally speculated that this lack of accuracy would be less of an issue for implementation in cochlear implants due to the inability to match stimulation rate to an exact frequency, but that the lack of accuracy may have influenced results in the normal-hearing study (Throckmorton et al., 2006).

Therefore, the authors have repeated a subset of the previous study utilizing two alternative frequency estimation techniques as well as the STFT. Only MCFA-2 (Multiple Carrier Frequency Algorithm with 2 predefined carrier frequencies per channel) was tested since the increase from one to two available presentation frequencies per channel had the greatest impact on speech recognition. The frequency was estimated for each channel using the STFT as originally proposed, an overlapping FFT with window durations that were channel dependent (see Table I), and a Flanagan phase vocoder (Flanagan and Golden, 1966). In addition to these

Corresponding Author: Leslie Collins, Phone: (919) 660-5260, Fax: (919) 660-5293, Email: lcollins@ee.duke.edu, Address: 130 Hudson Hall, Box 90291, Duke University, Durham, NC 27708

^aEmail address: cst@ee.duke.edu

^bEmail address: msk5@ee.duke.edu

^cEmail address: jjr6@ee.duke.edu

^dEmail address: lcollins@ee.duke.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

estimation techniques, a random selection strategy was used for comparison. Ten audiometrically-normal subjects were recruited from the Duke University staff and student population. Only vowel tokens ('h/V/d' format) were tested since enhancing frequency resolution within channel had the greatest effect on vowel recognition in the previous study. Tokens were presented in quiet as well as with speech-shaped noise added at three different signal-to-noise ratios (SNRs): 5, 0, and -5 dB.

The results of the vowel recognition task for MCFA-2 are shown in Fig. 1. The bar graphs are grouped for each noise condition, and from darker to lighter shades they present the results for the Flanagan phase vocoder, the overlapping FFT, the STFT methods, and the random strategy, respectively. The 95% confidence interval for each score was calculated and is indicated by the error bars on the bar plots. The performance with the STFT was significantly lower ($p < 0.05$) than the performance with the Flanagan phase vocoder and the overlapping FFT at 5 and -5 dB SNR respectively. However, neither the phase vocoder nor overlapping FFT techniques consistently provided improved performance; thus indicating an inconsistent benefit from increased frequency estimation accuracy. The significantly poorer performance ($p < 0.05$) with the random strategy at all noise levels indicates that the benefits of additional frequency resolution are only relevant if the selected frequencies represent the frequency content of the signal. Although the STFT was essentially a random strategy at the lowest channels, its ability to provide reasonable frequency estimates for the channels containing the formants resulted in significant performance gains over the random strategy ($p < 0.05$). Given that the more computationally complex algorithms investigated here provided limited, inconsistent benefit for normal hearing subjects, such approaches to frequency estimation may not be required for implementation in cochlear implant speech processing algorithms.

References

- Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America* 1997;102:2403–2411. [PubMed: 9348698]
- Flanagan JL, Golden RM. Phase vocoder. *Bell System Technical Journal* 1966;45:1493–1509.
- Throckmorton CS, Kucukoglu MS, Remus JJ, Collins LM. Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants. *Hearing Research* 2006;218:30–42. [PubMed: 16797896]

Fig 1.
Vowel recognition for the frequency estimation techniques at four different noise levels. Error bars indicate a 95% confidence interval.

TABLE I

Duration of windows for overlapping FFT for each channel.

Channel Number	Frequency Range	MCFA-2
1	[150 240]	30ms
2	[240 384]	30ms
3	[384 615]	16ms
4	[615 984]	16ms
5	[984 1574]	6ms
6	[1574 2519]	4ms
7	[2519 4031]	4ms
8	[4031 6450]	2ms (no-overlap)