Published in final edited form as:

Hear Res. 2008 May ; 239(1-2): 1–11.

Improving performance in noise for hearing aids and cochlear implants using coherent modulation filtering^a

Jong Ho Won^{1,3}, Steven M. Schimmel^{1,4}, Ward R. Drennan^{1,2}, Pamela E. Souza^{1,5}, Les Atlas^{1,4}, and Jay T. Rubinstein^{1,2,3}

¹ VM Bloedel Hearing Research Center, University of Washington Box 357923, Seattle, WA 98195-7923

² Department of Otolaryngology, Head and Neck Surgery, University of Washington

³ Department of Bioengineering, University of Washington

⁴ Department of Electrical Engineering, University of Washington

⁵ Department of Speech and Hearing Science, University of Washington

Abstract

This study evaluated the maximal attainable performance of speech enhancement strategies based on coherent modulation filtering. An optimal adaptive coherent modulation filtering algorithm was designed to enhance known signals from a target talker in two-talker babble noise. The algorithm was evaluated in a closed-set, speech-recognition-in-noise task. The speech reception threshold (SRT) was measured using a one-down, one-up adaptive procedure. Five hearing-impaired subjects and five cochlear implant users were tested in three processing conditions: (1) original sounds; (2) fixed coherent modulation filtered sounds; and (3) optimal coherent modulation filtered sounds. Six normal-hearing subjects were tested with a 6-channel cochlear implant simulation of sounds processed in the same three conditions. Significant improvements in SRTs were observed when the signal was processed with the optimal coherent modulation filtering algorithm. There was no benefit when the signal was processed with the fixed modulation filter. The current study suggested that coherent modulation filtering might be a promising method for front-end processing in hearing aids and cochlear implants. An approach such as hidden Markov models could be used to generalize the optimal coherent modulation filtering algorithm to unknown utterances and to extend it to open-set speech.

Keywords

Coherent modulation filter; speech enhancement; hearing aids; cochlear implants

1. Introduction

An envelope derived from an analytic signal (Dugundji, 1958) is known to contain speech information that varies at slow rates (Dudley, 1939). Many physical and acoustical signals such

^aPreliminary data of this work was presented at the 149th Meeting of the Acoustical Society of America, 16-20 May 2005.

Corresponding author address: Jong Ho Won, VM Bloedel Hearing Research Center, University of Washington, Box 357923, Seattle, WA 98195-7923, Phone: 206-616-2041, Fax: 206-616-1828, Email: jhwon@u.washington.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.

as speech and music can be modeled by low frequency modulators that modulate higher frequency carriers. Figure 1 shows this representation of a signal. The original signal (Figure 1(a)) has 800 Hz of acoustic frequency and 20 Hz of a modulation frequency as follows (Sampling frequency: 8 kHz, duration: 250 ms).

$$y = \cos(2\pi \times 800 \times t) \times (0.5 + 0.4 \times \cos(2\pi \times 20 \times t)) \quad \text{Eq. 1}$$

The signal can be also represented as a low frequency signal (20 Hz) that modulates a higher frequency signal (800 Hz). The left panels of Figure 1(b) and (c) show the envelope and fine-structure (i.e. carrier) of the original signal, respectively. The spectrum, which shows the modulation and the acoustic frequency spectrum, is shown in the right panels of Figure 1(b) and (c), respectively. Every sound has its unique acoustic and modulation frequency that allows us to analyze the sound on a two-dimensional transform domain, where one axis is for the modulation frequency and the other is for the acoustic frequency (the right panel of Figure 1 (d)).

Different speakers have different fundamental frequencies and different modulation of formants due to different vocal tract length, size, and mass. Humans can segregate sounds from different sources via fundamental frequency (Broadbent and Ladefoged, 1957) and formants (Darwin and Gardner, 1986). Likewise, the target speech and background noise differ in their fundamental frequencies, formants, modulation and in acoustic frequency. Figure 2 shows joint acoustic/modulation frequency representations of “had” (upper panel) and “hood” (lower panel) spoken by two male talkers and one female talker. The difference in formant location in acoustic frequency and formant modulation is not only seen between male and female talkers, but also between the two male talkers.

There is also ample evidence from animal neurophysiological and human psychoacoustical studies that suggests the importance of modulation for perception of sounds. Møller (1971) observed that the mammalian auditory system has a specialized sensitivity to amplitude modulation of narrowband acoustic signals. The neural representation of amplitude modulation was seen in all levels of auditory system including auditory cortex (Schreiner and Urbas, 1986). Magnetoencephalography (MEG) study suggests that frequency and modulation periodicity are represented via orthogonal maps in the human auditory cortex (Langner *et al.*, 1997). Schreiner and Langner (1988) found that cells in the inferior colliculus act as modulation frequency filters tuned to specific modulation frequencies. Liang *et al.* (2002) demonstrated that cortical neurons respond stronger to modulated stimuli than to unmodulated stimuli. Humans also have ability to detect modulated sounds (Patterson and Johnson-Davies, 1977; Rodenburg, 1977; Viemeister, 1977, 1979; Bacon and Viemeister, 1985; Formby *et al.*, 1992). Drullman *et al.* (1994) pointed out that the most important perceptual information lies at modulation frequency range of 4–16 Hz. Smith *et al.* (2002) demonstrated that envelope modulation sensitivity contributes significantly to the perception of speech.

Modulation analysis and filtering have been introduced to enhance or separate sounds (Arai *et al.*, 1996; Greenberg and Kingsbury, 1997; Kusumoto *et al.*, 2000). In previous methods of modulation filtering, an input signal is divided into sub-band signals via a filter-bank or a short-time transform. Then, each sub-band is decomposed into an slowly-varying envelope and a narrow-band carrier. Each sub-band envelope is passed through a linear time-invariant filter, and subsequently multiplied with the original, unmodified sub-band carrier to obtain modulation filtered sub-band signals. Finally, the filtered sub-bands are summed together to reconstruct a broad-band modulation filtered output signal.

The decomposition of a sub-band into its envelope and carrier is often done using the Hilbert transform (Drullman *et al.*, 1994) or by a direct magnitude estimate of the envelope (Vinton

et al., 2001). This type of envelope and carrier detection, however, is considered an *incoherent* way of estimation (Atlas *et al.*, 2004; Schimmel and Atlas, 2005), which has been shown to cause distortion that reduces the effectiveness of modulation filters (Ghitza, 2001; Schimmel and Atlas, 2005). This distortion can be avoided to a large degree by estimating the envelope and carrier in a *coherent* way (Atlas and Janssen, 2005; Schimmel and Atlas, 2005; Schimmel *et al.*, 2006). In the coherent approach, a sub-band is decomposed into an envelope and carrier by estimating a sub-band carrier signal using an instantaneous frequency (IF) estimator, and coherently demodulating the sub-band signal by the carrier to obtain the sub-band envelope. The general framework of coherent modulation filtering is illustrated in Figures 3(a) and 3(b) (see Methods).

Although coherent carrier estimation is a theoretical improvement over incoherent envelope detection for modulation filtering, not much is known about the practical potential of coherent modulation filtering techniques as a speech enhancement strategy. In this study, we assess the performance of an optimal adaptive coherent modulation filter for speech enhancement. A least-squares (LS) adaptive filter was used to optimize and achieve better performance of the proposed coherent modulation filtering than the conventional coherent modulation filtering, which uses a linear time-invariant filter.

We measured the speech reception threshold (SRT) for spondee words in two-talker babble noise with three different signal processing methods: (1) original, unprocessed signal; (2) fixed coherent modulation low-pass filtering; and (3) optimal adaptive coherent modulation filtering. We referred to these methods as unprocessed, fixed modulation filtering, and optimal modulation filtering, respectively. We measured the SRTs of all three signal processing methods for hearing-impaired subjects and cochlear implant (CI) users, as well as for normal-hearing subjects over a 6-channel vocoder simulation, to determine the relative benefit of each processing method for speech understanding in noise.

2. Methods

2.1. Optimal coherent modulation filter

The signal processing of the optimal coherent modulation filtering was based on the sub-band decomposition of coherent modulation filtering techniques shown in Figure 3(a). The conventional coherent modulation filtering uses a linear time-invariant filter shown in Figure 3(b). However, the current study proposed to use a LS adaptive filter shown in Figure 4(a) and 4(b) to achieve better enhancement. First, a broadband input signal was separated into 129 uniformly spaced sub-bands with a bandwidth of 207 Hz. In the case of the optimal coherent modulation filter, both the noisy input signal $x(t)$ and a desired output signal $d(t)$ were separated into frequency sub-bands, $x_k(t)$ and $d_k(t)$, using the same filter bank. Next, the sub-band carriers $x_{c,k}(t)$ of the noisy input signal were determined via an instantaneous frequency (IF) estimator, and they were used to coherently demodulate both $x_k(t)$ and $d_k(t)$ to obtain the envelope of the noisy input signal x and the desired envelope of the output signal $d_{m,k}(t)$, respectively. In all test cases for the coherent modulation filtering discussed below, the IF estimate was performed on the noisy sub-band input signal. Details on the IF estimator approach and its suitability for coherent modulation filtering can be found in Atlas and Janssen (2005) and Schimmel and Atlas (2006). Then, each sub-band signal was filtered using the LS filter $h_k(t)$ as shown in Figure 4(b). In each sub-band, the filter coefficients of the LS filter were determined by minimizing the squared error E_k between the filtered envelope $\hat{x}_{m,k}(t)$ and the desired envelopes $d_{m,k}(t)$:

$$E_k = \int_0^T |\hat{x}_{m,k}(t) - d_{m,k}(t)|^2 dt$$

$$= \int_0^T \left| \left(\int_0^{T_n} h_k(\tau) x_{m,k}(t - \tau) d\tau \right) - d_{m,k}(t) \right|^2 dt$$

Eq.2

A female speaker's spondee samples in quiet (see next section) were used as the desired output $d(t)$. The length T_h of the LS filters was restricted to 24 ms in each sub-band to have an acceptable amount of latency and sufficient modulation filtering effectiveness without overfitting the problem. In the optimal modulation filter case, filter coefficients were synchronously updated once every 250 ms in each sub-band.

2.2. Fixed coherent modulation filter

For the fixed coherent modulation filter, a low-pass filter with a cutoff frequency of 16 Hz was chosen. This choice was motivated by our hypothesis that the presentation of speech in babble noise might induce modulation frequency content above the conventional 4–16 Hz modulation frequency range of a single talker (Drullman *et al.*, 1994). In the fixed coherent modulation filter case, samples were processed with an identically-sized, fixed filter, which means that the same filter coefficients were used through all experimental conditions. Figure 5 shows the spectra of the fixed 16-Hz modulation filter (solid line), and the optimal modulation filter (dotted lines) of two different sub-bands at $T = 250$ ms, for one of the spondee words. Note that two of the optimal modulation filters of different sub-bands show a different frequency response, meaning that there are 129 different filters adapted to the input signal for each sub-band every 250 ms.

2.3. Vocoder simulation

Six-channel vocoder simulation was used to mimic the output signal of CIs. The input signal was divided into sub-band signals by using six band-pass finite impulse response (FIR) filters. The frequency ranges of the sub-bands were 80-308, 308-788, 788-1794, 1794-3906, 3906-8338, and 8338-17640 Hz. The envelope and the fine-structure were extracted by a Hilbert transform. Then, the Hilbert phases of the sub-band signals were randomized to make the randomized fine-structure. The randomized fine-structure of each sub-band was multiplied by the Hilbert envelope of each sub-band signal. Finally, the same band-pass filter was applied again to the sub-band output and all filtered sub-band outputs were summed together.

2.4. Spondee in babble test

The spondee in babble test was a closed-set task where the subjects were asked to identify one of 12 spondees that they heard in the presence of two-talker babble. The 12 spondees were of equal difficulty (Harris, 1991). The methods of this test follow those of Turner *et al.* (2004) and Won *et al.* (2007). Subjects were allowed to listen to each processed spondee word until they became familiar with the sounds. During testing, a spondee was presented in noise and listeners clicked on one of twelve boxes on a monitor which were labeled with the spondee words. Each subject completed three runs in each condition in random order. The design was a 3×3 using three different processing types and three repetitions. The spondees, spoken by a female talker, had fundamental frequencies that ranged from 212–250 Hz (Turner *et al.*, 2004). The background noise was two-talker babble with two competing sentences spoken by a male voice saying “Name two uses for ice” and by a female voice saying “Bill might discuss the foam” from the SPIN test (Bilger *et al.*, 1984). This female talker was different from the female talker who recorded the spondees. The same background noise was used on every trial and the onset of the spondees was 500 ms after the onset of the background noise. Forty-one different signal-to-noise ratio (SNR) conditions of the mixture of the target and two-talker babble noise were generated, from -50 dB SNR to $+30$ dB SNR with a 2 dB step size. The target spondee remained at 65 dBA and the babble level was tracked in 2-dB steps using a 1-down, 1-up procedure converging on 50% correct (Levitt, 1971). All wave files were computed offline and saved on a hard drive. Unprocessed, fixed, and optimal modulation filtered sounds were then processed using the 6-channel vocoder simulator described above. One test run was terminated when the subject completed 14 reversals. The threshold was calculated by taking

the mean of the last 10 reversals. The sound materials were presented directly to subjects through TDH-50P headphones. The procedures were implemented using MATLAB.

2.5. Subjects

Five bilateral hearing loss patients (ages 40–83) participated in this test. All had sensorineural loss with no significant air-bone gaps and immittance results within normal limits. Individual hearing-impaired subject information is listed in Table 1. Six normal-hearing subjects (ages 26–38) participated in the test. Five CI users (ages 34–77) participated in this study. Individual CI subject information is listed in Table 2. The use of human subjects in this study was reviewed and approved by the University of Washington Institutional Review Board.

3. Results

3.1. Acoustic analysis

Figure 6 shows the waveform (the first column), envelope (the second column), and spectra (the third column) of an original spondee word (drawbridge, the first row), mixed with the babble noise (−10 dB SNR, the second row), the same token of the mixed processed with the fixed modulation filter (the third row), and with the optimal modulation filter (the fourth row). The mixed sound shows that the envelope of the original spondee is contaminated by the presence of the babble noise. The original spondee shows clear spectral peaks, but the mixed sound does not show them. However, the optimal modulation filtered sound shows a similar envelope to that of the original spondee word. Some of the spectral peaks are also recovered in the optimal modulation filtered sound. Table 3 shows a quantitative analysis conducted on the envelope of each waveform with different SNR conditions. It shows that the correlation between the envelope of the original spondee word and the unprocessed mixed sound, fixed modulation, and optimal modulation filtered sound at 0 dB SNR are 0.64, 0.64, and 0.66, respectively. At −10 dB SNR condition, however, which is more challenging situation for the listeners, only the optimal modulation filtered sound provided a recovered envelope information to the listeners with correlation of 0.63. Correlations of less than 0.1 are found with the unprocessed and the fixed modulation filtered sound at −10 dB SNR condition, meaning that the envelope of the spondee word is contaminated by the presence of the background noise and the fixed modulation filtering can not recover the envelope information in such a strong background noise.

3.2. Speech perception evaluation

Individual data from the hearing-impaired listeners and improvements from each types of processing are shown in Figure 7. Error bars in all figures represent $\pm$ one standard deviation. The average SRT in two-talker babble for hearing-impaired listeners was -19.77 ± 7.38 dB SNR with unprocessed sounds, -12.41 ± 5.75 dB SNR with fixed modulation filtered sounds, and -24.30 ± 5.63 dB SNR with optimal modulation filtered sounds. Note that because the speech-in-noise task was a closed-set test, scores were as good as −20 dB SNR. A 3×3 repeated measures of analysis of variance (ANOVA, 3 different signal processing methods, 3 repetition times) demonstrated that there was a main effect of the signal processing types ($F_{2,6} = 13.726$, $p = 0.006$). No learning effect and no interaction between processing methods and repetition times were shown in this task. Subjects showed different responses in terms of the improvement provided from fixed and optimal modulation filtering. Average improvement in SRTs in the optimal modulation filtered sound compared to the unprocessed sound was 4.53 dB SNR, suggesting that the optimal modulation filtering helped the hearing-impaired subjects to perceive speech sound in higher levels of noise (by 4.53 dB) compared by the unprocessed sounds. However, in the fixed modulation filter case, SRTs for the hearing-impaired listeners got worse. Subjects' performance decreased by 7.36 dB when they listened to the fixed modulation filtered sounds. A detailed discussion of the results is provided in the next section.

Figure 8 shows individual data and improvements of fixed modulation filtering, and optimal modulation filtering followed by 6-channel vocoder simulation in normal-hearing listeners. The average SRT in two-talker babble was -10.23 ± 7.90 dB SNR with unprocessed sounds, -10.18 ± 6.77 dB SNR with fixed modulation filtered sounds, and -31.58 ± 9.46 dB SNR with optimal modulation filtered sounds. A 3×3 repeated-measures ANOVA (3 types of signal processing and 3 repetition times) demonstrated that, as expected, there was a significant effect of the signal processing method ($F_{2,10} = 181.571, p = 0.00001$). Subjects showed some learning effect ($F_{2,10} = 6.647, p = 0.015$), and no interaction between repetition and processing method ($F_{4,20} = 0.925, p = 0.47$). Average improvement in SRT in the optimal modulation filtered sound compared to the unprocessed sound was 21.35 dB SNR, which was larger than the improvement from the optimal modulation filtering for hearing-impaired listeners. The fixed modulation filtering didn't provide any improvement (-0.05 dB).

Figure 9 shows individual data and improvements of fixed modulation filtering, and optimal modulation filtering in CI listeners. The average SRT in two-talker babble was -5.93 ± 7.18 dB SNR with unprocessed sounds, -7.65 ± 7.83 dB SNR with fixed modulation filtered sounds, and -19.87 ± 14.90 dB SNR with optimal modulation filtered sounds. A 3×3 repeated-measures ANOVA (3 types of signal processing and 3 repetition times) demonstrated that, as with hearing-impaired and normal-hearing subjects, there was a significant effect of the signal processing method ($F_{2,8} = 10.17, p = 0.006$). No learning effect and no interaction between processing methods and repetition times were shown in CI subjects. Average improvement in SRTs in the optimal modulation filtered sound compared to the unprocessed sound was 13.93 dB SNR. The fixed modulation filtering provided improvement of 1.72 dB.

4. Discussion

The results suggested that all subjects improved with the optimal modulation filtering relative to unprocessed speech in noise. Group results for CI subjects and normal-hearing subjects with vocoder simulations showed that subjects had substantial benefit from the optimal modulation filtering, even larger than those provided to hearing-impaired listeners. Among 5 CI subjects, only one subject (C03) did not benefit from the optimal modulation filtered sound, suggesting that majority of CI users would be able to benefit from the optimal modulation filter.

The present study tested speech perception with competing talkers using normal-hearing listeners presented with CI vocoder simulations and in CI users. The CI vocoder simulations serve as a model of CI users' performance. The extent to which the CI users' results can be explained by the simulation model can be determined by comparing the CI vocoder simulations data with actual CI subjects data. Such CI vocoder simulation can also illustrate important aspects of CI processing without CI user's individual differences such as device, duration of deafness, age, nerve survival, and residual hearing. In the current study, both vocoder and CI subject data showed a similar pattern; 1) the greatest benefit provided by the optimal modulation filter, 2) little to no benefit provided by the fixed modulation filter, and 3) variation across the subjects. In Figure 8 and 9, the better two CI subjects (C1 and C4) data look exactly like the vocoder average data, suggesting that C1 and C4 might have spectral resolution and temporal coding as good as the vocoder. The data also suggest that the better CI users, in terms of their consonant-nucleus vowel-consonant (CNC) score, showed more improvement in the optimal modulation filter.

Three groups of listeners in the current study showed different trends of performance with the fixed modulation filtering. Normal-hearing listeners presented with vocoder simulations and CI subjects showed about the same performance with the fixed modulation filtering and the unprocessed sounds. However, hearing-impaired listeners performed worse with the fixed modulation filtering (Figure 7). We hypothesized that (1) the fixed modulation filter with a

cut-off frequency of 16 Hz might filter out modulation spectra of the babble background noise; and (2) subsequently, the subjects' speech discrimination performance might improve with the fixed modulation filtering. However, test results showed that removing high modulation frequencies didn't help hearing-impaired listeners to discriminate target speech from two-talker babble, suggesting that modulation frequencies above 16 Hz contain useful information for listeners to discriminate target speech from a two-talker babble background.

As noted in the results, the learning effect was minimal compared to the signal processing effect. Over the course of the three repeated runs, 16 subjects showed average learning of 2.69 dB, 3.2 dB, and 5.98 dB with the unprocessed, fixed modulation filtered, and optimal modulation filtered sound, respectively. The difference between the unprocessed and optimal condition was maintained throughout the three repetitions. There was a trend for the main effect of optimal processing to increase with learning, but it was not statistically significant. The average improvement due to optimal processing vs. unprocessed stimuli in 16 subjects was 13.77 dB.

The present study compared SRTs in two-talker babble noise between three different types of signal processing. Temporal envelope information is critical for high levels of speech perception (Shannon et al. 1995). Current cochlear implants use temporal envelope information of sounds to amplitude modulate the current pulses. In our tests, we had a female talker for target speech, and male and different female talkers for babble noise. Better representation of the temporal envelope of the target speech in the presence of the background noise is desirable if one wants to enhance the speech perception performance. The present study shows a large effect of the optimal modulation filtering on discriminating spondee in two-talker babble for hearing-impaired listeners, CI users, and vocoder simulations. This suggests that the optimal modulation filtering enhances the information of the target speech so that listeners can segregate different voices. Acoustical analysis (Table 3) suggests that part of the speech information recovered by the optimal modulation filtering is the temporal envelope of the target speech.

Several techniques have been suggested to improve hearing aids for speech understanding in noise such as adaptive filtering, directional microphones, and adaptive noise cancellation (Levitt, 2001). However, for hearing aid users, speech understanding in noise remains problematic. Kochkin (2000) reported that only 33% of hearing aid users are satisfied with their hearing aids in noisy environments. CI users have the same difficulty. Although the temporal envelope encoding strategy of the current cochlear implants has shown good performance in speech recognition in quiet, it is still difficult for CI users to understand speech in background noise (Nie *et al.*, 2005).

The present study demonstrated that benefit was provided for CI users and for hearing-impaired listeners when the sounds were processed with the optimal modulation filter. This outcome suggests that the coherent modulation filtering has a great potential as a front-end processing for hearing devices to improve speech intelligibility. However, the optimal modulation filter algorithm in this study is not a candidate for speech enhancement itself, as it requires knowledge of the exact spondee signals, which is not directly available in practice. In order to make this approach practical, offline trained models would be required to establish specific parameters for a target talker. These models, which would be trained to describe the probabilistic range of target talker variation, could make use of hidden Markov models, as are currently used in modeling for speech recognition. Or alternatively, the currently unconstrained wide range of possible modulation filtering functions could use other recent training-based approaches (e.g. Smaragdis, 2007) to capture and represent the variability of target talkers. Once trained, either of these techniques could be implemented in real-time systems.

Acknowledgements

We appreciate the dedicated efforts of our listeners. Two anonymous reviewers provided helpful comments on previous versions of this manuscript. This work was supported by the Washington Research Foundation (Souza, Atlas), the V.M. Bloedel Hearing Research Center (Atlas, Schimmel, Souza), the National Institutes of Health (Won, Rubinstein, Drennan, grant no. R01-DC007525 and P30-DC004661), and KOSEF (Won).

A list of the abbreviations

ANOVA	analysis of variance
CI	cochlear implant
FIR	finite impulse response
IF	instantaneous frequency
LS	least-squares
MEG	Magnetoencephalography
SNR	signal-to-noise ratio
SRT	speech reception threshold

References

- Arai T, Pavel M, Hermansky H, Avendano C. Intelligibility of Speech with Filtered Time Trajectories of Spectral Envelopes. *Proc ICSLP* 1996;4:2490–93.
- Atlas L, Janssen C. Coherent Modulation Spectral Filtering for Single-Channel Music Source Separation. *Proc IEEE ICASSP*. 2005
- Atlas L, Li Q, Thompson J. Homomorphic Modulation Spectra. *Proc IEEE ICASSP*. 2004
- Bacon SP, Viemeister NF. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology* 1985;24:117–134. [PubMed: 3994589]
- Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. *J Speech Hear Res* 1984;27:32–48. [PubMed: 6717005]
- Broadbent DE, Ladefoged P. On the fusion of sounds reaching different sense organs. *J Acoust Soc Am* 1957;29:708–710.
- Darwin CJ, Gardner RB. Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *J Acoust Soc Am* 1986;79:838–845. [PubMed: 3958326]
- Drullman R, Festen J, Plomp R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 1994;95:1053–1064. [PubMed: 8132899]
- Dudley H. Remaking Speech. *J Acoust Soc Am* 1939;11:169–177.
- Dugundji J. Envelopes and pre-envelopes of real waveforms. *IRE Trans Inform Theory* 1958;4:53–57.
- Formby C, Morgan LN, Forrest TG, Raney JJ. The role of frequency selectivity in measures of auditory and vibrotactile temporal resolution. *J Acoust Soc Am* 1992;91:293–305. [PubMed: 1737878]
- Ghitza O. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *J Acoust Soc Am* 2001;110:1628–1640. [PubMed: 11572372]

- Greenberg S, Kingsbury BED. The Modulation Spectrogram: in Pursuit of an Invariant Representation of Speech. ICASSP 1997:1647–50.
- Harris, RW. Signals, Sound, and Sensation. AIP Press; New York: 1991. Speech audiometry materials compact disk.
- Kochkin S. Customer satisfaction with single and multiple microphone digital hearing aids. Hearing Review 2000;7(11):24–34.
- Kusumoto A, Arai T, Kitamura T, Takahasi M, Murahara Y. Modulation enhancement of speech as preprocessing for reverberant chambers with the hearing-impaired. ICASSP 2000:853–856.
- Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am 1971;49:467–477. [PubMed: 5541744]
- Levitt H. Noise reduction in hearing aids: a review. J Rehab Res Develop 2001;38:111–121.
- Langner G, Sams M, Heil P, Schulze H. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. J Comp Physiol A 1997;181:665–676. [PubMed: 9449825]
- Liang L, Lu T, Wang X. Neural Representations of Sinusoidal Amplitude and Frequency Modulations in the Primary Auditory Cortex of Awake Primates. J Neurophysiol 2002;87:2237–2261. [PubMed: 11976364]
- Møller A. Unit responses of the rat cochlear nucleus to tones of rapidly varying frequency and amplitude. Acta Physiol Scan 1971;81:540–556.
- Nie K, Stickney G, Zeng FG. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans Biomed Engin 2005;52:64–73.
- Patterson, RD.; Johnson-Davies, D. Detection of a change in the pitch of AM noise. In: Evans, EF.; Wilson, JP., editors. Psychophysics and Physiology of Hearing. Academic; London: 1977. p. 363-371.
- Rodenburg, M. Investigation of temporal effects with amplitude modulated signals. In: Evans, EF.; Wilson, JP., editors. Psychophysics and Physiology of Hearing. Academic; London: 1977. p. 429-437.
- Smaragdis P. Convolutional Speech Bases and Their Application to Supervised Speech Separation. IEEE Trans Speech and Audio Processing 2007;15:1–12.
- Schimmel S, Atlas L. Coherent envelope detection for modulation filtering of speech. Proceedings of ICASSP 2005;2005:221–224.
- Schimmel S, Fitz KR, Atlas L. Frequency Reassignment for coherent modulation filtering. Proceedings of ICASSP 2006;2006:261–264.
- Schreiner C, Urbas J. Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). Hear Res 1986;21:227–241. [PubMed: 3013823]
- Schreiner CE, Langner G. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J Neurophysiol 1998;60:1823–1840. [PubMed: 3236053]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature 2002;416:87–90. [PubMed: 11882898]
- Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. J Acoust Soc Am 2004;115:1729–1735. [PubMed: 15101651]
- Viemeister, NF. Temporal factors in audition: A system analysis approach. In: Evans, EF.; Wilson, JP., editors. Psychophysics and Physiology of Hearing. Academic; London: 1977. p. 419-428.
- Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 1979;66:1364–1380. [PubMed: 500975]
- Vinton MS, Atlas LE. A scalable and progressive audio codec. Proc ICASSP 2001:3277–80.
- Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J Assoc Res Otolaryngol 2007;3:384–92. [PubMed: 17587137]

Figure 1.

(a) Left: original signal composed of 800-Hz carrier modulated at 20 Hz. Right: Spectrum of the original signal. (b) Left: envelope of the original signal. Right: Spectrum of the modulator of the original signal. (c) Left: Fine structure (i.e. carrier) of the original signal. Right: Spectrum of the carrier of the original signal. (d) Left: spectrogram of the original signal. Right: joint acoustic/modulation frequency representation of the original signal. Colors indicate energy over an 80 dB range, where red represents high energy and blue low energy.

Figure 2. Joint acoustic/modulation frequency representations of “had” (upper panel) and “hood” (lower panel) spoken by three different talkers. Colors indicate energy over an 80-dB range, where red represents high energy and blue low energy.

Figure 3. (a) General sub-band decomposition and reconstruction of modulation filtering techniques. (b) Sub-band processing of coherent modulation filtering techniques. A sub-band signal is decomposed into a carrier signal $x_{c,k}(t)$ using an instantaneous frequency estimator (IF) and a coherently demodulated envelope signal $x_{m,k}(t)$. The envelope signal is subsequently filtered by a linear time-invariant filter $g_k(t)$ and the modified envelope is recombined with the original, unmodified carrier into a modulation filtered sub-band signal.

Figure 4. Sub-band processing of the “optimal” adaptive coherent modulation filtering technique. A sub-band signal is decomposed into a carrier signal $x_{c,k}(t)$ using an instantaneous frequency estimator (IF), which is used to coherently demodulate the noisy sub-band signal $x_k(t)$ as well as the desired sub-band signal $d_k(t)$. The envelope signal is subsequently filtered by a linear time-invariant filter $g_k(t)$, and the modified envelope is recombined with the original, unmodified carrier into a modulation filtered sub-band signal. (b) Adaptive modulation filter.

Figure 5.
The spectra of the filters for one of the spondee words. Solid line: the fixed 16-Hz modulation filter. Dotted lines: the optimal modulation filter of two different sub-bands at $T = 250$ ms.

Figure 6.

The first column shows the waveform, the second column shows the envelope, and the third column shows the spectra of the original spondee word (first row), the mixed unprocessed sound (second row), the fixed modulation filtered sound (third row), and the optimal modulation filtered sound (fourth row).

Figure 7. The first panel shows the SRTs for five hearing-impaired subjects for three test conditions. The second panel shows the individual improvement in SRTs from the unprocessed to processed signal.

Figure 8. The first panel shows the SRTs for six normal-hearing subjects for three test conditions. The second panel shows the individual improvement in SRTs from the unprocessed to processed signal.

Figure 9. The first panel shows the SRTs for five CI subjects for three test conditions. The second panel shows the individual improvement in SRTs from the unprocessed to processed signal.

Table 1
Hearing-impaired subject characteristics: audiometric thresholds are shown in dB HL.

Subject	Ear	Frequency (Hz)					
		250	500	1000	2000	3000	8000
H1	Right	35	45	60	60	60	60
	Left	55	55	75	70	65	55
H2	Right	20	40	50	50	50	55
	Left	20	40	55	50	55	55
H3	Right	10	45	50	40	20	5
	Left	15	35	55	50	40	0
H4	Right	30	40	45	55	60	65
	Left	25	30	35	50	55	55
H5	Right	25	30	40	40	40	35
	Left	25	30	35	40	45	35

Table 2

CI subjects characteristics.

Subject	Age (yrs)	Duration of hearing loss (yrs) <i>a</i>	Duration of deafness (yrs)	Etiology	Implant type	Speech Processor Strategy	CNC word score (% correct)
C01	52	7	5	Hereditary	HiRes 90K	HiRes 120	86
C02	68	5	3	Hereditary	HiRes 90K	HiRes 120	44
C03	34	3	6	Unknown	Clarion CII	HiRes 120	24
C04	40	15	6	Genetic	Clarion CII	HiRes 120	98
C05	77		1/2	Noise exposure	HiRes 90K	HiRes 120	66

*a*The duration of their hearing loss before implantation

Table 3

Correlations between the envelope of the original spondee word and the processed sound.

Condition	0 dB SNR	-10 dB SNR
Unprocessed	0.64	0.09
Fixed modulation filtered	0.64	0.02
Optimal modulation filtered	0.66	0.63