

Effect of filter spacing on melody recognition: Acoustic and electric hearing

Kalyan Kasturi and Philipos C. Loizou

*Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688
loizou@utdallas.edu*

Abstract: This paper assesses the effect of filter spacing on melody recognition by normal-hearing (NH) and cochlear implant (CI) subjects. A new semitone filter spacing is proposed for music. The quality of melodies processed by the various filter spacings is also evaluated. Results from NH listeners showed nearly perfect melody recognition with only four channels of stimulation, and results from CI users indicated significantly higher scores with a 12-channel semitone spacing compared to the spacing used in their daily processor. The quality of melodies processed by the semitone filter spacing was preferred over melodies processed by the conventional logarithmic filter spacing.

© 2007 Acoustical Society of America

PACS numbers: 43.66.Ts, 43.66.Hg [QJF]

Date Received: March 12, 2007 **Date Accepted:** May 2, 2007

1. Introduction

Central to any speech coding strategy used in multi-channel cochlear implants is the decomposition of the acoustic signal into frequency bands. Given the large number (12–22) of electrodes available in commercial implant devices, it is becoming more important to find the best mapping (or, equivalently, spacing) of frequency bands to electrodes. The majority of implant processors use logarithmic (or semilog) filter spacing and that has worked well so far, at least for speech recognition.

A number of studies evaluated alternative filter spacings for vowel recognition and F0 discrimination (e.g., Fourakis *et al.*, 2004; Geurts and Wouters, 2004; Laneau *et al.*, 2004). Fourakis *et al.* (2004) advocated the placement of more filters in the F1/F2 region for better representation of the first two formants. Small but significant improvements were noted on vowel recognition with an experimental map which included one additional electrode in the F2 region. Similar outcome was reported in Skinner *et al.* (1995) and Loizou (2006). Other studies also considered the possibility of allocating more filters in the low frequencies for better place coding of individual harmonics and consequently better pitch perception. A new filter bank was proposed by Geurts and Wouters (2004) based on a simple loudness model used in acoustic hearing. The new filter bank, which allocated more filters in the low frequencies, was tested on an F0 detection task in the absence of temporal cues and yielded lower detection thresholds to F0 for synthetic vowel stimuli compared to a conventional filter bank based on log spacing.

The above studies demonstrated that the filter spacing can have a positive effect on vowel recognition and can in some cases reduce F0 difference limens, at least for steady-state vowels with a steady F0 contour. Little is known, however, about the effect of filter spacing on music signals which have a dynamic F0 contour. This paper investigates the hypothesis that a filter-bank spaced according to a musical scale would provide better place coding of individual harmonics and consequently improve melody recognition. The present experiments investigate the effect of semitone frequency spacing on melody recognition by normal-hearing and cochlear implant users.

Table 1. The 3 dB frequency boundaries of the semitone filterbank. Lower (L), upper (UP) and center (C) frequencies are given for each band in Hz.

	2 channels			4 channels			6 channels			12 channels		
	L	U	C	L	U	C	L	U	C	L	U	C
1	300	424	362	300	357	328	300	337	318	300	318	309
2	424	600	512	357	424	391	337	378	357	318	337	327
3				424	505	464	378	424	401	337	357	347
4				505	600	552	424	476	450	357	378	367
5							476	535	505	378	400	389
6							535	600	567	400	424	412
7										424	449	437
8										449	476	463
9										476	505	490
10										505	535	520
11										535	566	550
12										566	600	583

2. Experiment design

2.1 Subjects and material

Six Clarion CII (Advanced Bionics Corporation) cochlear implant users participated in this experiment. All subjects were postlingually deafened adults wearing the cochlear implant (CI) for a minimum of 2–3 years (no consideration was given to their musical training experience). For comparative purposes, we also tested ten normal-hearing (NH) subjects listening to stimuli processed via acoustic simulations of cochlear implants. A set of 34 simple melodies (e.g., “Twinkle Twinkle,” “Old McDonald”) with all rhythm information removed was used (Hartmann and Johnson, 1991) as test material. These same melodies were used in the study by Smith *et al.* (2002). Melodies consisted of 16 equal-duration notes synthesized using samples of a grand piano. The mean of all 16 note frequencies of each tune was concert A (440 Hz) plus or minus a semitone. The largest difference between the highest and lowest notes was 12 semitones.

2.2 Signal processing

For the NH listeners, the test material was first bandpass filtered (sixth order Butterworth) into 2–12 channels according to a semitone filter spacing that spanned an octave (300–600 Hz). This frequency range was chosen as it encompasses the mean note frequency (440 Hz) of the test stimuli. For the 12-channel condition, each filter had a bandwidth of 1 semitone (see Table 1). For the 6-channel condition, each filter had a bandwidth of 2 semitones, and for the 4- and 2-channel conditions the filters had a bandwidth of 3 and 6 semitones, respectively. In addition to the semitone spacing, a 16-channel logarithmic spacing (225 Hz–4.5 kHz) was used as control. Following the bandpass filtering, the channel envelopes are computed using a half-wave rectifier followed by a second order Butterworth low-pass filter with a cutoff frequency of 120 Hz. The resulting envelopes of each channel were modulated with white noise and re-filtered with the same analysis filters. The melodies were finally synthesized by summing up the outputs of all the channels.

For the CI users, the test material was processed through the continuous interleaved sampling strategy used in the subject’s daily processor, and implemented with different frequency spacings. Two different filter spacings were considered. The first one was based on the semitone scale mentioned above. Four semitone (4SM), six semitone (6SM) and 12 semitone (12SM) based filter banks were considered. In the 4SM condition, only the four most apical

electrodes were used for stimulation. Similarly, in the 6SM and 12SM conditions, only the 6 and 12 most apical electrodes were used for stimulation. The remaining electrodes were not stimulated.

The second filter spacing considered involved a combination of semitone and log spacings. This was done to account for a more realistic scenario in which the melodies might contain sung lyrics. In the 4SM condition, we considered utilizing a log spacing for the remaining 12 channels in the high frequencies. Similarly, a log spacing was used for the remaining ten channels in the 6SM condition and the remaining four channels in the 12SM condition. We refer to these hybrid frequency spacings as 4SM+LOG, 6SM+LOG and 12SM+LOG, respectively. All hybrid frequency spacings used 16 channels of stimulation. For comparative purposes, we tested subjects with the 16-channel logarithmic spacing (16LOG) used in their daily processor.

2.3 Procedure

The experiments with NH listeners were performed using a PC equipped with a Creative Labs SoundBlaster 16 soundcard. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones. The names of the melodies were displayed on a computer monitor, and a graphical user interface enabled the subjects to indicate their response. Prior to the test, each subject was asked to select ten familiar melodies from the list of 34 melodies (with a few exceptions, most subjects selected the same melodies). A training session (lasting about 10–15 min) with the ten selected melodies was performed using the original unprocessed melodies. Subjects were required to score above 90% with unprocessed melodies before participating in the experiment. After the training session, the subjects were tested with the melodies processed through the various number of channels. The order of test conditions was randomized across subjects.

The cochlear implant subjects were tested using the Clarion research interface-II (Advanced Bionics Corporation). Prior to the test, the subjects were asked to select ten known melodies from a list of 34 melodies. The subjects were given a practice session that lasted for about 10–15 min. Following the practice session, the subjects were tested on the ten selected melodies using the logarithmic spacing, semitone spacing, and hybrid spacings. The names of the melodies were displayed on a computer monitor, and a graphical user interface enabled the subjects to indicate their response. The subjects were tested for a total of seven different filter spacings. Each spacing was tested in two blocks of three repetitions each. The order of the various spacings tested was randomized across subjects.

Following the melody recognition test, the CI users participated in an AB paired preference test. In one condition, the task was to evaluate and compare the quality of melodies processed by the 16LOG and 6SM spacings. In another condition, the task was to compare the 16LOG and 6SM+LOG spacings. In each trial, the subjects listened to two stimuli each processed using a different filter spacing. The preference test included ten melody pairs composed of five different melodies. Five of the ten melody pairs were presented as filter spacing A followed by spacing B, while the other five were presented as spacing B followed by spacing A. The subjects were instructed to make a preference as to which stimulus sounded more “musical” (i.e., sounding like a melody with “natural” melodic contour) and more pleasant. In addition, they were asked to make a confidence rating on each comparison at six distinct scales: slightly better (or slightly worse), better (or worse), and much better (or much worse). A numeric score was assigned to each rating ranging in values from +3 (much better) to –3 (much worse). A total of six (signed) confidence ratings were assigned and a distance measure was computed. The percentage preference was computed as the percentage of the number of times stimulus B was preferred over stimulus A. The distance measure was computed to assess quantitatively how much stimulus B sounded better than stimulus A. Since the distance measure is computed over ten test pairs, it ranged in values from –30 to 30, with a positive value indicating that the strategy B is preferred, and a negative value indicating otherwise.

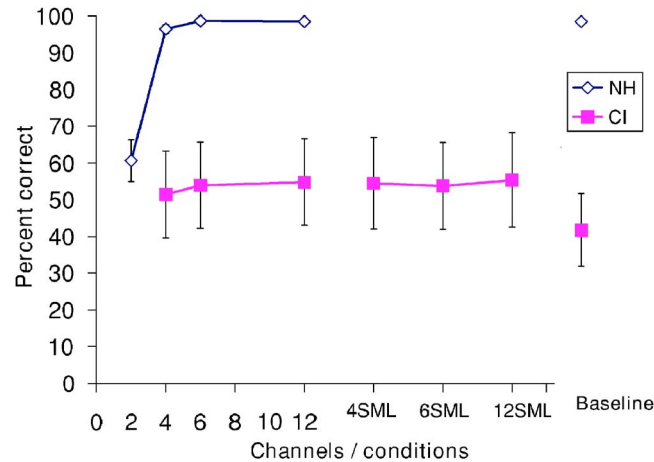


Fig. 1. (Color online) Mean percent correct scores for melody recognition by CI users (filled symbols) and normal hearing listeners (open symbols) as a function of number of semitone-spaced channels. The scores obtained by CI users in the hybrid spacing conditions (4SML, 6SML and 12SML) are also included. The baseline condition corresponds to the spacing (16 channels) used in the subjects' daily processor. Error bars indicate standard errors of the mean.

3. Results and discussion

The mean percent correct scores for melody recognition by normal hearing listeners are depicted in Fig. 1 (open symbols). Analysis of variance (ANOVA) with repeated measures showed a significant effect ($F[4, 16]=59.4, p<0.0005$) of number of channels on melody recognition. Nearly perfect melody recognition was achieved with only four channels of stimulation spaced according to a semitone scale. The three-semitone frequency resolution does not allow individual harmonics (spaced a semitone apart) to be resolved, yet it was found sufficient for accurate melody recognition, at least by normal-hearing listeners who receive acoustic envelope information at the correct place in the cochlea and have good frequency selectivity.

The mean percent correct scores for melody recognition by CI users are depicted in Fig. 1 (filled symbols). ANOVA (with repeated measures) indicated a significant effect ($F[6, 30]=2.8, p=0.026$) of frequency spacing on melody recognition. Post-hoc tests indicated that the scores obtained with the 12SM spacing were significantly higher ($p=0.021$) than the scores obtained with the conventional filter spacing (16LOG) used by CI users in their daily processor. Scores obtained with the other semitone spacings were not significantly higher, but approached ($p\approx 0.06$) the significance level.

The preference judgments for each subject are given in Table 2. Results indicated that the quality of melodies processed by the semitone filter spacing was preferred over melodies

Table 2. Subject preference scores (ranging from 0 to 100) indicating the number of times the semitone spacing (6SM) (or hybrid spacing, 6SM+LOG) was preferred over the conventional logarithmic filter spacing (16LOG). The distance scores in parentheses (ranging from -30 to 30) are positive if the semitone (or hybrid) spacings were preferred and negative if the log spacing was preferred. Large positive distance scores indicate stronger preference of the semitone-based filter spacing over the log spacing.

Filter spacing comparisons	Subjects						Mean
	S1	S2	S3	S4	S5	S6	
16LOG vs. 6SM	100 (25)	100 (20)	100 (20)	90 (14)	90 (14)	100 (20)	96.7 (18)
16LOG vs. 6SM+LOG	100 (12)	80 (8)	0 (-19)	20 (-10)	100 (25)	50 (0)	58.3 (3)

processed by the conventional filter spacing (16LOG). The quality of melodies processed by the 6SM strategy was preferred 97% of the time over the CI user's daily strategy (16LOG). The preference of the hybrid spacing (6SM+LOG) was not as strong (58%). This could be attributed to the fact that subjects were perhaps perceiving conflicting or noncoherent pitch cues in the low- (semitone spaced) and high-frequency (log spaced) channels. We cannot exclude the possibility that the hybrid spacing might yield higher preference scores when tested with music containing sung lyrics.

The above analyses indicate that the filter spacing can have a significant effect on music perception both in terms of melody recognition and subjective quality. These results suggest that the semitone filter spacing enhanced access to place (spectral) cues resulting in better F0 discrimination. The magnitude of the improvement, however, was not as large as that observed by normal-hearing listeners receiving the same number of channels of frequency information. Two factors could have contributed to that. First, in cochlear implants the acoustic information is rarely presented in the correct place in the cochlea due to the shallow insertion depth. As a result, the frequency-to-place mapping is somewhat compressed or expanded. There is evidence (Oxenham *et al.*, 2004) to suggest that a correct (i.e., matched) frequency-to-place mapping is necessary for complex pitch perception and consequently melody recognition. We cannot exclude, however, the possibility that if the subjects were given more time to adapt to the new frequency spacing, their scores might improve even further and this warrants further investigation. Second, the place-coding resolution in cochlear implants is limited and constrained by several factors including the electrode spacing, location of electrodes in terms of their proximity to excitable neuron elements and electrode configuration (monopolar vs. bipolar). All these factors limit the frequency specificity needed for complex pitch perception. If we assume that the mismatch in frequency-to-place mapping can be compensated over time with learning, then based on the outcome by NH listeners (Fig. 1), a place-coding resolution of 3 semitones (or better) would be required for accurate melody recognition by CI users.

The data from the present experiment demonstrate that the channel density in the low-frequency range plays a critical role in melody recognition. In cochlear implants, this channel density is influenced by the signal bandwidth and number of electrodes available. In a follow up experiment we investigated the effect of signal bandwidth on melody recognition using acoustic simulations and NH listeners. Test material was bandpass filtered into N ($N=2, 4, 6, 12, 40$) frequency bands using sixth-order Butterworth filters. The N bands were uniformly spaced on a logarithmic scale and spanned either a 5 kHz (225 Hz–4.5 kHz range) or an 11 kHz (225 Hz–10.5 kHz range) signal bandwidth. Following the envelope detection (120 Hz) and white noise modulation, the signals were re-filtered through the same analysis filters and summed up for reconstruction. A new group of ten listeners participated in this experiment using the same procedure and test material. The mean results are shown in Fig. 2. Two-way ANOVA (with repeated measures) indicated a significant effect ($F[1, 4]=10.5, p=0.031$) of signal bandwidth, a significant effect ($F[4, 16]=81.6, p<0.005$) of spectral resolution (number of channels) and a significant interaction ($F[4, 16]=5.8, p=0.004$). For the small-bandwidth condition, post-hoc tests (Fisher's LSD) showed that the performance asymptoted with 6 channels, while for the large-bandwidth condition performance asymptoted with 12 channels. Near perfect melody identification was achieved with 12 (or more) channels in both conditions.

The results shown in Fig. 2 clearly demonstrate that the signal bandwidth, which in turn affects the filter spacing (for a fixed number of channels), is extremely important for melody recognition. Higher performance was achieved with the small signal bandwidth, as more filters were allocated in the low-frequency range. In the 6-channel condition (based on large-bandwidth allocation), only one filter was allocated in the 300–600 Hz range, while in the corresponding 6-channel condition, based on small-bandwidth allocation, two filters were allocated within the same range. This small difference in the number of filters in the low-frequency range produced a difference of 34 percentage points in melody recognition (Fig. 2).

It is important to note that the proposed filter spacings were only tested with melodies and not with speech. Further tests are needed to assess the effects of the proposed filter-bank

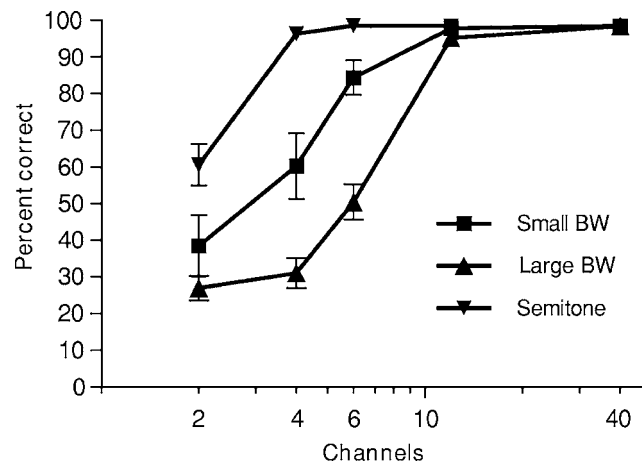


Fig. 2. Mean scores (percent correct) on melody recognition as a function of number of channels and signal bandwidth (small BW=5 kHz, large BW=11 kHz). Scores from the semitone spacing (300–600 Hz) are also plotted for comparison. Error bars indicate standard errors of the mean.

manipulations on speech recognition. The hybrid filter spacings (4SM+LOG, 6SM+LOG, 12SM+LOG) would clearly be more appropriate for speech recognition as they span the speech bandwidth. These spacings produced comparable performance on melody recognition as the semitone spacings (see Fig. 1). Alternatively, the semitone filter spacing could be programmed as a separate “music map” which CI users can switch to when wanting to listen to (instrumental) music.

Acknowledgment

This research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and Other Communication Disorders, NIH.

References and links

- Fourakis, M., Hawks, J., Holden, L., Skinner, M., and Holden, T. (2004). “Effect of frequency boundary assignment on vowel recognition with the Nucleus 24 ACE speech coding strategy,” *J. Am. Acad. Audiol.* **15**, 281–289.
- Geurts, L., and Wouters, J. (2004). “Better place coding of the fundamental frequency in cochlear implants,” *J. Acoust. Soc. Am.* **115**(2), 844–852.
- Hartmann, W. M., and Johnson, D. (1991). “Stream segregation and peripheral channeling,” *Music Percept.* **9**(2), 155–184.
- Laneau, L., Moonen, M., and Wouters, J. (2004). “Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees,” *J. Acoust. Soc. Am.* **106**(6), 3606–3619.
- Loizou, P. (2006). “Speech processing in vocoder-centric cochlear implants,” *Cochlear and Brainstem Implants*, edited by A. Moller, (Karger, Basel) Vol. **64**, pp. 109–143.
- Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (2004). “Correct tonotopic representation is necessary for complex pitch perception,” *Proc. Natl. Acad. Sci. U.S.A.* **101**(5), 1421–1425.
- Skinner, M., Holden, L., and Holden, T. (1995). “Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy,” *Ann. Otol. Rhinol. Laryngol.* **104**(Suppl. 166), 307–311.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). “Chimaeric sounds reveal dichotomies in auditory perception,” *Nature (London)* **416**, 87–90.