Effect of filter spacing and correct tonotopic representation on melody recognition: Implications for cochlear implants

Kalyan S. Kasturi and Philipos C. Loizou

Dept. of Electrical Engineering

The University of Texas at Dallas

Research supported by NIDCD/NIH (R01 DC 3421)

Introduction

- Several studies reported that cochlear implant listeners perform poorly (near chance) on melody identification tasks.
- This is partly due to the fact that current implant processors convey primarily envelope information and no fine-structure cues.
- Most devices use a logarithmic filter spacing, which is appropriate for speech, but not for music. Unlike speech, music is based on a highly-structured semitone scale.
- We therefore hypothesize that a filter spacing scheme that corresponds to a musical semitone structure might better capture pitch information for music perception (Exp 1).

Introduction (cont'ed)

- A corollary to the above hypothesis is that the signal bandwidth might be critical for melody recognition as it affects the number of filters that fall within the low frequency region (Exp 2).

Experiment 1

- Two different filter spacings were investigated: logarithmic and semitone-spaced.

- **Semitone-spacing**

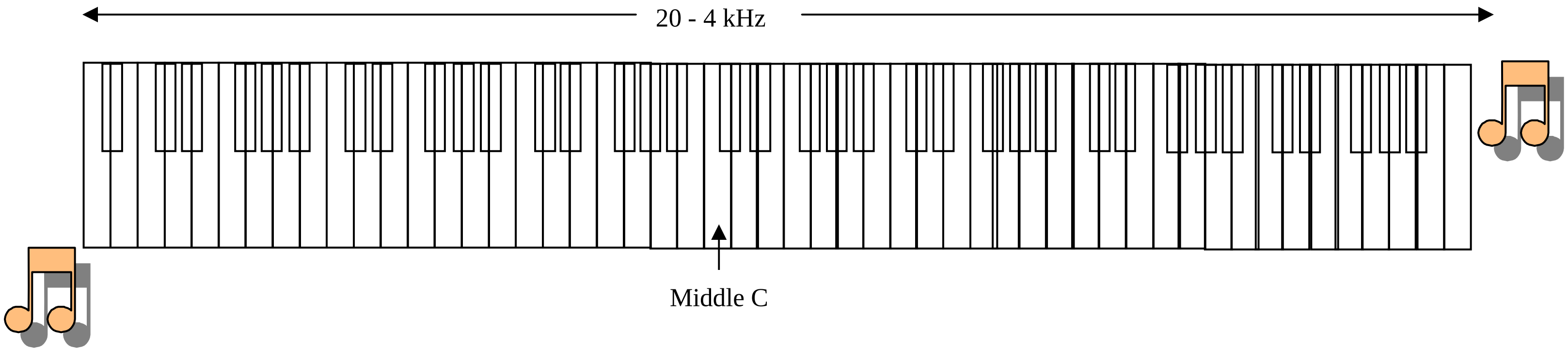
  We varied the number of channels from 2 to 12 with the following filter bandwidths:

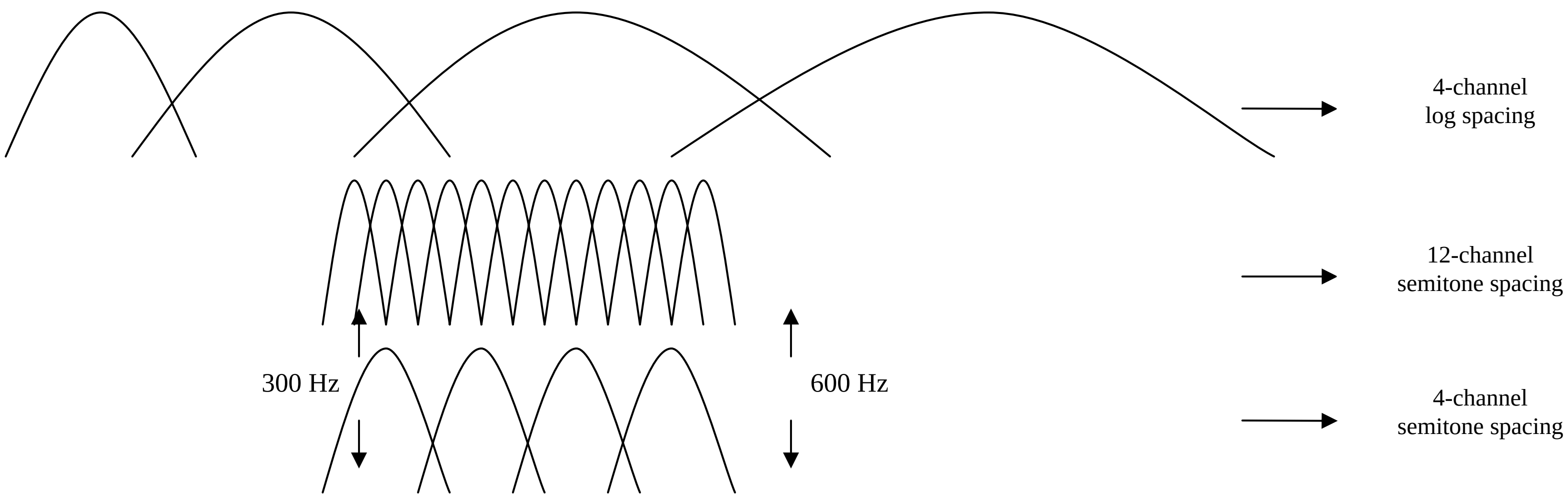
  - 12 channels - each filter had a bandwidth of 1 semitone
  - 6 channels - each filter had a bandwidth of 2 semitones
  - 4 channels – each filter had a bandwidth of 3 semitones
  - 2 channels – each filter had a bandwidth of 6 semitones

- **Logarithmic spacing (currently used by commercial devices)**

  Filters were logarithmically spaced. We varied the number of channels from 2 to 40.

Filter Spacing





Signal Processing

- Melodies were bandpass filtered into N channels using 6-th order Butterworth filters. The output of each channel was passed through a rectifier followed by a second-order Butterworth low-pass filter with cut-off frequency of 120 Hz to obtain the envelope of each channel.

- The envelope of each band-pass filter was modulated with white noise. Noise modulated envelopes were passed through synthesis filters that were essentially the same as the analysis filters.

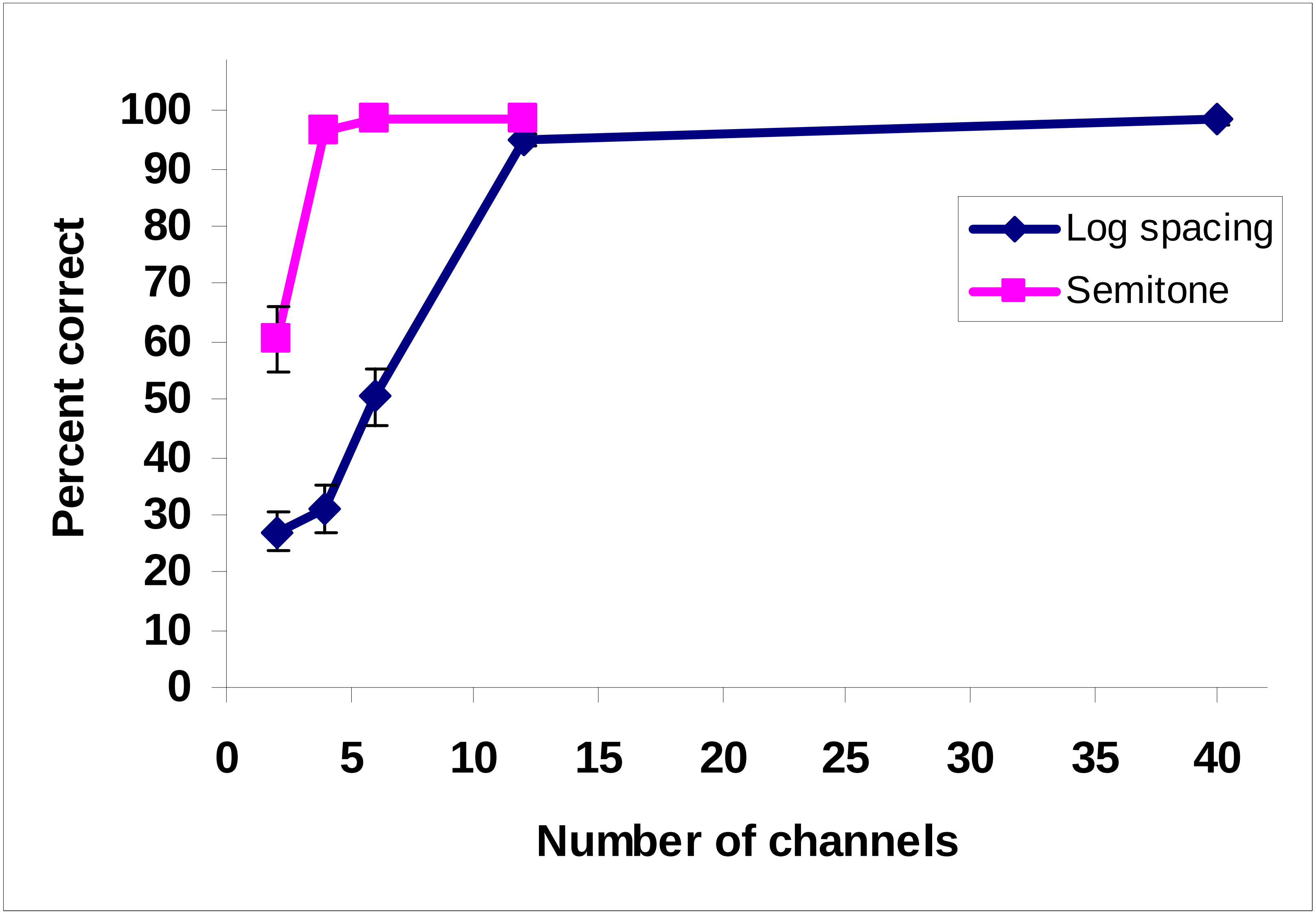
- The outputs of all channels were summed up to obtain the synthesized melodies.

- Synthesized melodies were presented to 10 normal-hearing subjects for identification in a closed-set format.

Melodies

- The melody test used thirty-four common melodies each consisting of sixteen isochronous notes as used by Hartmann [7].
- Isochronous notes were used to remove the rhythm cues from the melodies.
- The notes were synthesized using samples of acoustic grand piano.

Results:
Effect of filter spacing



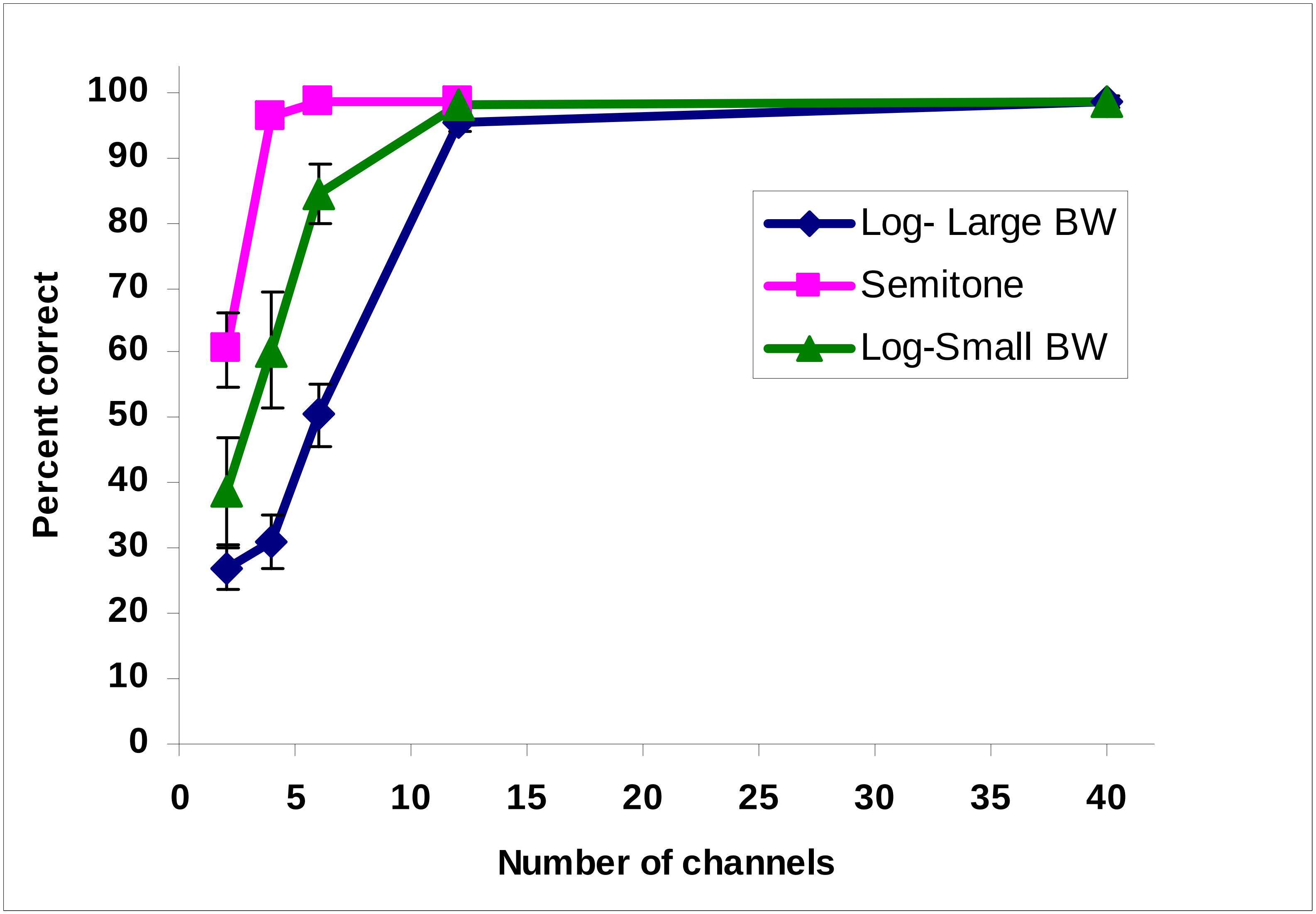
Analysis and Discussion

- Two-way ANOVA (repeated measures) indicated a significant effect of spectral resolution (number of channels), a significant effect of frequency spacing and a significant interaction ($p<0.005$).
- Semitone-spacing:
  - Post-hoc tests (Fisher's LSD) showed that performance asymptoted ($p>0.5$) with 4 channels.
- Performance with 4 channels based on semitone filter spacing as good as performance with 12 channels based on logarithmic filter spacing.
- Conclusion: Filter spacing is extremely important in melody recognition.

Experiment 2

- Investigated the effect of signal bandwidth on identification of melodies.
- **Hypothesis**: If a smaller signal bandwidth is used, then more filters would fall in the low-frequency region and melody recognition should improve.
- Added one more condition in which the filters were logarithmically spaced within a smaller bandwidth spanning the range of 225-4500 Hz.
- Five normal-hearing listeners participated in this experiment.

Results: Effect of Bandwidth



Analysis and Discussion

- Two-way ANOVA (repeated measures) indicated a significant effect of spectral resolution (number of channels), a significant effect of bandwidth and a significant interaction ($p<0.005$).
- Post-hoc tests (Fisher's LSD) indicated that:
  - 4 chan: performance with small bandwidth > large bandwidth ($p=0.013$)
  - 6 chan: semitone spacing > small bandwidth ($p=0.029$)
small bandwidth > large bandwidth ($p<0.005$)
- For small number of channels, using a small bandwidth brings significant benefits on melody recognition. Semitone spacing remains superior.

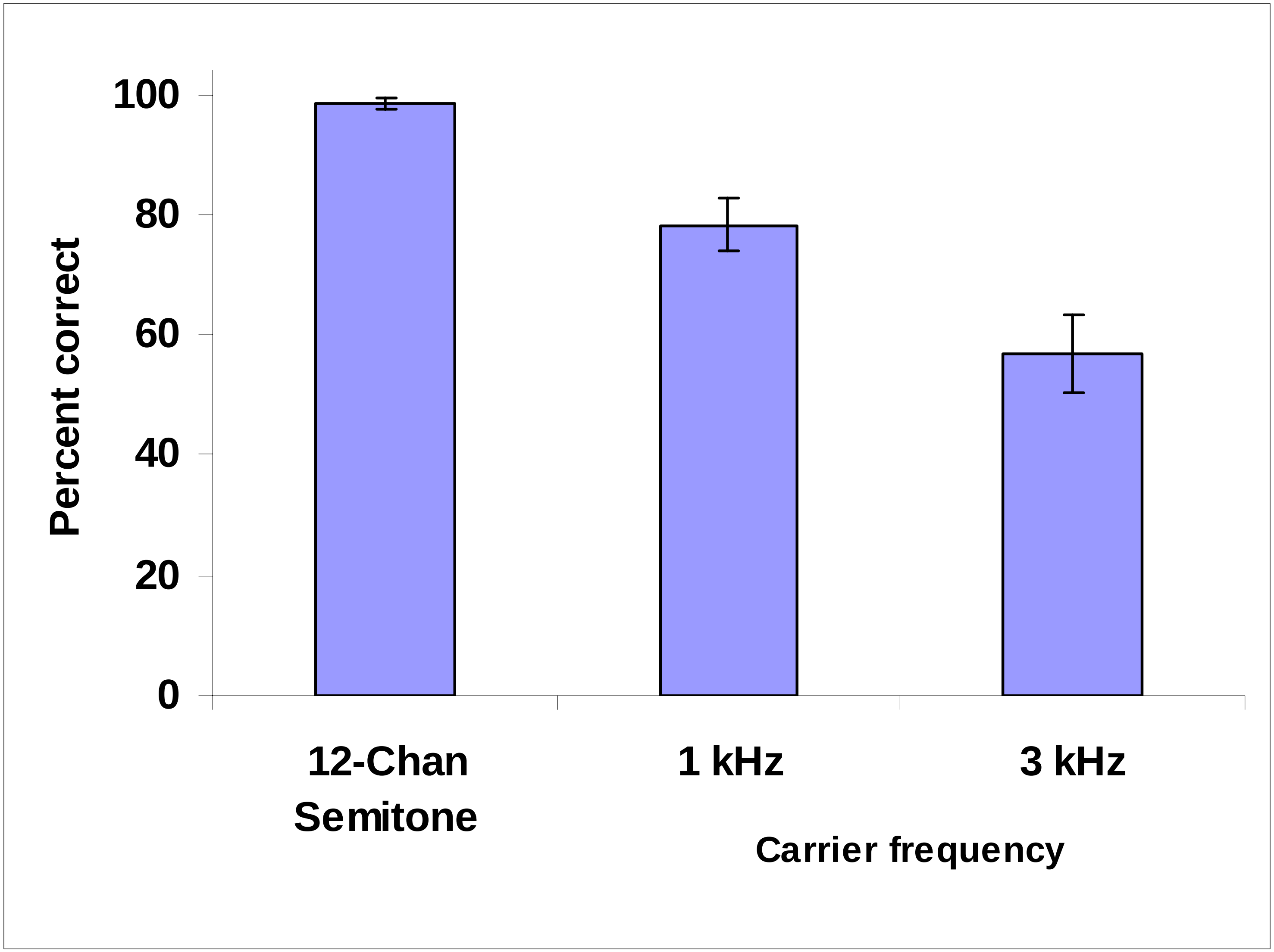
Experiment 3

- In cochlear implants, acoustic information is rarely presented in the correct place in the cochlea due to shallow insertion depths.
- CI patients typically receive frequency up-shifted stimuli. With speech, it is known that patients can tolerate large amounts of shift.
- The effect of frequency up-shifting on melody identification has not been thoroughly investigated.
- In the present experiment, we investigate the up-shifting effect by using frequency transposed melodies – i.e., melodies that are transposed to higher frequencies (1 and 3 kHz).

Experiment 3: Transposed Stimuli

- The transposed stimuli preserve the temporal structure of the signal and can thus be used to assess the importance of presenting the music stimuli at the correct tonotopic place in the cochlea (Oxenham *et al*., Proc. Nat. Proc. Sc., 2004).
- More specifically, the present experiment will examine whether pitch perception can be accounted for by a purely temporal code or whether a tonotopic representation of frequency (place code) is necessary.
- The transposed stimuli were generated by multiplying the original 12-channel stimuli (semitone spacing) by a high-frequency sinusoidal carriers at 1 and 3 kHz.

Results:
Frequency transposed melodies



Analysis and Discussion

- ANOVA (repeated measures) indicated a significant effect [$F(2,18)=21.2$, $p<0.005$] of correct tonotopic representation on melody recognition.
- *Post hoc* tests (Fisher's LSD) indicated that performance with 1 kHz carrier was significantly ($p=0.005$) lower than baseline, and performance with 3 kHz carrier was significantly ($p=0.003$) lower than performance with 1 kHz carrier.
- Correct tonotopic representation is critically important for complex pitch perception.

Conclusions

- The semitone-based filter spacing yielded the best performance among all the filter spacings investigated.
- Nearly perfect melody recognition (~98%) was achieved using only four channels.
- The distribution of filters in the low-frequency region is very important for melody recognition. Filters based on a smaller signal bandwidth yielded significantly higher scores.
- Correct tonotopic representation is necessary for complex pitch perception – melody recognition.

Discussion

- This shows that a finer filter spacing around the melody spectrum would better capture the fine structure cues and hence result in better melody recognition.

- As modulation frequency was increased melody recognition dropped. This indicates that preserving the place of stimulation is important.

- Upshifting the synthesized melodies with semitone spacing using four channels resulted in nearly perfect recognition and thus upshifting with a factor of 6.5mm did not degrade the performance.

Bibliography

1. Gfeller, K. and Lansing, C. R. (1991). “Melodic, rhythmic, and timbral perception of adult cochlear implant users,” *Journal of Speech and Hearing Research*., 34, 916 - 920.

2. Schulz, E. and Kerber, M. (1994). “Music perception with the MED-EL implants,” *Advances in cochlear implants*, 326 - 332.

3. Loizou, P. (1998). “Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants,” IEEE Signal Process. Mag., 15(5), 101 - 130.

Bibliography

4. Lobo, A., Toledo, F., Loizou, P. and Dorman, M. (2002). “Effect of envelope low-pass filtering on melody recognition,” *33rd Neural Prosthesis Workshop*, Bethesda, MD.

5. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). “Speech recognition with primarily temporal cues,” Science, 270, 303 - 304.

6. Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (2004). “Music perception with temporal cues in acoustic and electric hearing,” Ear and Hearing, 25(2), 173 - 185.

Bibliography

7. Hartmann, W. M. and Johnson, D., (1991). “Stream segregation and peripheral channeling,” Music perception, 9(2), 155 - 184.