Investigating the use of a Gammatone filterbank for a cochlear implant coding strategy

Journal Article

Author(s):
Tabibi, Sonia; Kegel, Andrea; La, Wai K.; Dillier, Norbert

Research paper

Investigating the use of a Gammatone filterbank for a cochlear implant coding strategy

Sonia Tabibi [a,b,*], Andrea Kegel [b], Wai Kong Lai [b], Norbert Dillier [b]

[a] Department of Information Technology and Electrical Engineering, ETH Zürich (ETHZ), Zürich, Switzerland
[b] Laboratory of Experimental Audiology, ENT Department, University Hospital Zürich, Zürich, Switzerland

HIGHLIGHTS

- A Gammatone filterbank has the potential to better resolve the harmonics of complex tones.
- A total delay of a Gammatone filterbank can be made smaller compared to an FFT filterbank with the same frequency resolution at low frequencies.
- Melody contour identification improved with longer frame size or higher frequency resolution.

ARTICLE INFO

ABSTRACT

Background: Contemporary speech processing strategies in cochlear implants (CIs) such as the Advanced Combination Encoder (ACE) use a standard Fast Fourier Transform (FFT) filterbank to extract envelopes. The assignment of the FFT bins to approximate the frequency resolution of the basilar membrane is only partly based on physiology, especially since the bins are distributed linearly below 1000 Hz and logarithmically above 1000 Hz.
New method: A Gammatone filterbank which provides a closer approximation to the bandwidths of filters in the human auditory system could replace the standard FFT filterbank in the ACE strategy. An infinite impulse response (IIR) all-pole design of the Gammatone filterbank was compared to the FFT filterbank with 128, 256 and 512 points resolutions and the effect of the frequency boundaries of the filters was also investigated.
Results: Melodic contour identification (MCI) and just noticeable difference (JND) experiments, both involving synthetic clarinet notes in octaves 3 and 4, were conducted with 6 normal hearing (NH) participants using noise vocoded stimuli; and 10 CI recipients just performed the MCI experiment. The MCI results for both NH and CI subjects, showed a significant effect of the filterbank on the percentage correct responses of the participants.
Comparison with existing methods: The Gammatone filterbank can better resolve the harmonics of tested synthetic clarinet notes which led to better performances in the MCI experiment.
Conclusions: The total delay of the Gammatone filterbank can be made smaller than the delay of the FFT filterbank with the same frequency resolution at low frequencies.

1. Introduction

Research efforts to improve coding strategies for cochlear implants (CIs) were recently more frequently directed towards physiologically-based approaches (El Boghdady et al., 2016; Nogueira et al., 2005; Sit et al., 2007). A Gammatone filterbank which was introduced by Johannsma (1972) for the peripheral filtering in the cochlea and estimated by the reverse correlation function of neural firing times can be used for the frequency decomposition of the acoustic signal in hearing devices (Holdsworth et al., 1988). Since the Gammatone filterbank is a more physiologically-based filterbank compared to the standard Fast Fourier Transform (FFT) filterbank which is being used in the Advanced Combination Encoder (ACE) strategy, an application of the Gammatone filterbank in the CI coding strategy could be advantageous.

The Gammatone filterbank is defined in the time domain by its impulse response. The direct implementation of the convolution

* Corresponding author at: Frauenklinikstrasse 24, 8091 Zürich, Switzerland.
E-mail address: sonia.tabibi@usz.ch (S. Tabibi).

sum in the time domain is computationally quite expensive. Thus, the Gammatone filterbank is not an optimum option for the filterbank in the CI processors due to its impractical implementation (Cosentino et al., 2014). However, more efficient solutions in the frequency domain have been proposed. It turns out that all zeros of a pole-zero filter implementation are located on the real axis which means that the zeros have a small effect near the center frequency of each filter and can therefore be omitted (Slaney, 1993). It was shown that the shape and the temporal fine structure of the Gammatone filter response are well preserved by an all-pole approximation (Hohmann, 2002). The all-pole model offers fewer parameters and makes it easier to model the bandwidth and center frequency shift of the filters (Lyon, 1997). This reduces the computational effort of implementation by approximately 50% (Slaney, 1993) and would be beneficial for CI processors.

The ACE strategy in the Nucleus devices is based on 8 msec time frames for 128 samples of the acoustic signal at a sampling rate of 16 KHz which is typically used in the CI processors. In each frame, only the "n" channels with the highest energy content are selected and the per channel stimulation rate is defined by the frame rate (Zeng et al., 2008). The Gammatone filterbank implementation was done in this frame based format in order to compare it with the standard FFT filterbank performance and explore the feasibility of the implementation within the ACE coding strategy. This implementation imposed the use of longer frames for the Gammatone filterbank which last for 256 samples. This increased frame size can also be used with the FFT filterbank and therefore a comparison condition of an FFT filterbank with 256 points was included in the experimental protocol.

Cutoff frequencies of the Gammatone filterbank are specified on the Equivalent Rectangular Bandwidth (ERB) scale, which is a psychoacoustic measure of the width of the auditory filters (Slaney, 1993). This modified assignment of the cutoff frequencies leads to more resolution in the lower frequency channels compared to the standard FFT in the ACE strategy (Laneau et al., 2004). In order to have the same frequency resolution as the Gammatone filterbank in the low frequency channels, the frequency mapping of the FFT filters needs to be modified. This can be done by increasing the number of FFT points to 512. To separate the effects of changing cutoff frequencies and increased frequency resolution, two different conditions were considered: the FFT filterbank with 512 points having either the same cutoff frequencies as the standard FFT in the ACE strategy or the Gammatone filterbank. Thus, different conditions were considered for the comparison of the filterbanks; the standard FFT with 128 points, the Gammatone, the FFT with 256 points (FFT256), the FFT with 512 points (FFT512) and the FFT with 512 points and matched Gammatone cutoff frequencies (FFTGT).

A study from Kasturi and Loizou (Kasturi and Loizou, 2007) showed a significant effect of frequency spacing on melody recognition. Since the Gammatone has more resolution in the lower frequency channels, the lower harmonics of complex tones could be better resolved which can help for better melody identification. This can also help in the identification of the tone that is different in pitch in the JND (just noticeable difference) test. Thus, two experiments were conducted with normal hearing (NH) volunteers; melodic contour identification (MCI) and JND. Finally, the possible effects of changing the frequency resolution of the filterbank for CI subjects were investigated in the MCI experiment.

The aim of this study was to compare the infinite impulse response (IIR) all-pole design of the Gammatone filterbank which was implemented in the frequency domain for a CI coding strategy with the standard FFT filterbank in the ACE strategy. This design of Gammatone filterbank has not been used for the CI coding strategies until now. In addition to that, the distribution of cutoff frequencies for the FFT filterbank was matched to the Gammatone

cutoff frequencies distribution which was not the case in the Laneau et al. study (Laneau et al., 2004). If this matched cutoff frequencies distribution was not taken into account, the participant's performance not only had the effect of different filter types but also the effect of different distribution of frequency boundaries.

2. Materials and methods

2.1. FFT filterbank implementation

The standard FFT filterbank in the ACE strategy which has 128 points was implemented in MATLAB and made use of the Nucleus MATLAB Toolbox (NMT) v4.31 developed by Cochlear Corp (Swanson and Mauch, 2006). This toolbox includes functions for the conversion of the acoustic signal into electrical stimulation patterns. The patterns can then be used to synthesize vocoder output signals or directly streamed to a subject's implant.

The processing for the standard FFT filterbank in the NMT was performed on a circular buffer with a size of 128 samples. The buffer shift is defined based on the division of sampling rate by channel stimulation rate in the strategy (channel stimulation rate is the total implant rate divided by the number of selected channels). For instance, for a typical 900 Hz channel stimulation rate, the buffer shift is equal to 18 samples. A Hanning window with the same size as the buffer size was used before the standard FFT filterbank. Since the input signal was real, the output of the FFT filterbank had Hermitian symmetry, thus half of the FFT bins were discarded and a total of 64 bins were combined into 22 channels (Swanson, 2008).

The Gammatone filterbank frame-based implementation used a longer frame (256 samples) compared to the standard FFT filterbank, thus for the comparison the FFT frame size was also increased to 256 samples. This reduced the FFT frequency spacing to 62.5 Hz and 128 bins were used to combine into 22 channels. The FFT output was recomputed for the changes in the bin assignments to equalize the frequency response and the cutoff frequencies were kept the same as for the standard FFT filterbank. The remaining processing steps were unchanged.

Although the frame size of the standard FFT filterbank was changed to 256 samples, it was still not possible to get the same frequency resolution as the Gammatone filterbank in the lowest frequency channel. Thus, the FFT frequency resolution was increased to 512 points which gives a 31.25 Hz frequency spacing and a total of 256 bins for combining into 22 channels. In this way, we can explore the effect of changing the frequency mapping but at the same time the frequency resolution was increased from 256 to 512 points. In order to separate the effects of changing the cutoff frequencies and the frequency resolution, two conditions were considered: the FFT 512 points with the standard FFT frequency mapping and with the Gammatone frequency mapping (FFTGT). FFGT used the frame size of 512 samples which imposed longer delay compared to the Gammatone filterbank with 256 frame size (the frame sizes of 256 and 512 samples with a sampling rate of 16 KHz are equal to 16 msec and 32 msec delays). Thus, four different implementations of the FFT filterbank were compared with the Gammatone filterbank: the standard FFT, FFT256, FFT512 and FFTGT. All these implementations were operated using the NMT v4.31. It is worth mentioning that moving from the standard FFT filterbank with 128 points to 512 points (FFT512) increased the frequency resolution but at the cost of lower temporal resolution and longer delay.

2.2. Gammatone filterbank implementation

Filters in the Gammatone filterbank were 4th order all-pole IIR using Hohmann's implementation in the frequency domain

(Hohmann, 2002). It was reported that when the filter's order is in the range of 3 to 5, it gives a good approximation of the human auditory filters and is very similar to that of rounded-exponential (roex) filters (Hohmann, 2002; Holdsworth et al., 1988; Patterson et al., 1987; Slaney, 1993; Yin et al., 2011). This value is not in contradiction to the value proposed by Cosentino et al. (Cosentino et al., 2014), where the filter order was defined with the aim of optimizing the speech intelligibility in CIs rather than based on physiological characteristics of the cochlea. The bandwidths of the filters were set to the ERB scale (Moore and Glasberg, 1983) with the cutoff frequencies ranging from 187.5 Hz to 7937.5 Hz. This range was chosen to have the same covered frequency range along the cochlea as the standard FFT filterbank in the ACE strategy; although it would also be possible to choose lower values for the lowest cutoff frequency if so desired. The lowest cutoff frequency of the standard FFT filterbank is 187.5 Hz since an equivalent bandwidth of 3 dB for a Hanning window that is used before the FFT filterbank is 1.5 bins (1.5 bins with the sampling rate of 16 KHz and 128 points FFT is equivalent to 187.5 Hz) (Fearn, 2001). The highest cutoff frequency was set below the Nyquist frequency.

The absolute value of the complex output of the filter channels represents an approximation to the Hilbert envelope which can be used as an estimator for the further processing of the envelope (Hohmann, 2002). The frame-based implementation of the Gammatone filterbank and the way that the complex output of filter channels was combined to get one envelope value for each channel per frame imposed the use of longer buffer size. Thus, the same buffer as the standard FFT filterbank was used for the Gammatone filterbank but with a size of 256 samples. In each data frame, one value is selected for each channel and this was done by taking the arithmetic average over the signal envelope values of each channel. In the standard FFT filterbank implementation, some weights were inserted into the extracted envelopes of the channels to equalize the maximum gain of all channels (Laneau, 2005). These weights were defined for the Gammatone filterbank from the response of each channel to a sinusoid at its center frequency and then multiplied with the extracted envelope of that channel.

The implementation mentioned above replaced the filtering and envelope extraction stages in the NMT. The rest of the processing which includes selection of channels with the highest amplitudes, compression of the extracted envelope to the electric dynamic range and amplitude modulation of biphasic pulses were kept unchanged (Zeng et al., 2008). The different processing stages for the standard FFT and the Gammatone filterbanks are depicted in Fig. 1.

2.3. Stimuli

Digitally synthesized complex tones were used in this study since they are representative of acoustic tones in a real world (Nimmons et al., 2008). The synthetic tones were derived from clarinet notes in the RWC Music Database, instrument number 31, variation 1, normal articulation and mezzo dynamics (Goto, 2004). The design of the synthetic clarinet notes was done based on the spectral profile of a clarinet and similar to the approach described by Nimmons et al. (Nimmons et al., 2008). However, instead of a linear decay, a cosine ramp was applied to the beginning and end of each complex tone. Temporal modulation transfer functions are level dependent in NH and CI subjects and the linear decay may have a perceptual effect on the detection of envelope fluctuations (Milczynski et al., 2009). Each clarinet note was generated with the following harmonic relation; $F0$ represents the fundamental frequency:

$$F0\,(100\%) + 3F0\,(72\%) + 5F0\,(34\%) + 6F0\,(8.5\%) + 7F0\,(17\%) \tag{1}$$

Since the mapping of the cutoff frequencies in the Gammatone filterbank leads to more resolution in the lower frequency channels (Laneau et al., 2004), stimuli were selected in octaves 3 and 4. For the melodic contour identification, each note was 250 msec in duration with a 10 msec cosine ramp to the beginning and end to reduce any transient spectral splatter. A 50 msec interval was applied between notes and a root note (the lowest note in the melody) was selected C3 (130.81 Hz). The interval between successive notes in each contour was varied from 1 to 2 semitones for NH subjects and 1 to 3 semitones for CI subjects (Galvin et al., 2007). 3 semitones interval was not tested with NH subjects since it was relatively easy for them.

Stimulus duration in psychophysical experiments typically ranges between 200 msec and 500 msec (Fearn, 2001). In the JND experiment which was tested with only NH subjects; each synthetic clarinet note had duration of 500 msec with 50 msec cosine ramp to the beginning and end and 300 msec of silence between presented tones. The same relation was used to produce the synthetic clarinet notes in octaves 3 and 4. All semitones in one octave were generated according to:

$$f = 2^{x/12} f_{ref} \tag{2}$$

where f is the frequency of the target note, x is the number of semitones relative to the root note and f_{ref} is the frequency of the root note which is 130.81 Hz in octave 3 and 261.62 Hz in octave 4 (tuned to A4 = 440 Hz) (Galvin et al., 2007).

All the stimuli for the both experiments with NH subjects were processed with a neural-based vocoder with 10′000 neurons which incorporates refractory and the spread of excitation functions based on electrically evoked compound action potential (ECAP) recordings (El Boghdady et al., 2016). The stimuli for the MCI experiment with CI subjects were tested with direct streaming via Nucleus Implant Communicator (NIC) and an L34 research processor to their implants using the subjects' own clinical maps.

2.4. Subjects

6 NH participants (3 females, 3 males) aged between 31 and 52 years (mean age 39.83 years; standard deviation 9.37 years) volunteered to participate in the MCI and the JND experiments. The audiograms were measured for both ears and all of the participants had pure tone hearing thresholds less than 20 dB HL from 250 Hz to 4 KHz on both ears. 4 participants were experienced in playing musical instruments.

17 CI subjects participated in the MCI experiment; 10 of them were able to distinguish the melodic contour patterns and completed the experiment. The 10 CI participants (5 females, 5 males) aged between 28 and 64 years (mean age 53 years; standard deviation 13.12 years). The demographic details of the CI participants are presented in Table 1; all of the subjects were using the ACE strategy. All experiments were carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained from all subjects.

2.5. Experiments

2.5.1. Procedure
An example of the electrical stimulation patterns (electrodograms) from the standard FFT filterbank and the Gammatone filterbank applied to D3# (155.56 Hz) clarinet tone is shown in Fig. 2. The electrodograms depict differences in spectral patterns between both schemes and the discrimination of the harmonics is better represented with the Gammatone filterbank compared to the standard FFT filterbank.

As already mentioned, other conditions such as FFT256 and FFT512 were considered to explore the effect of increasing the
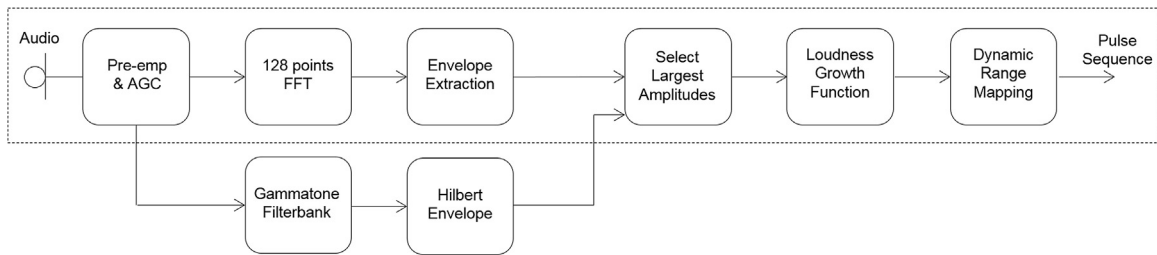
Fig. 1. Processing schemes of the standard FFT filterbank in the ACE strategy (components enclosed by dashed rectangle) and the Gammatone filterbank.

Table 1
Demographic details of CI subjects participated in this study.

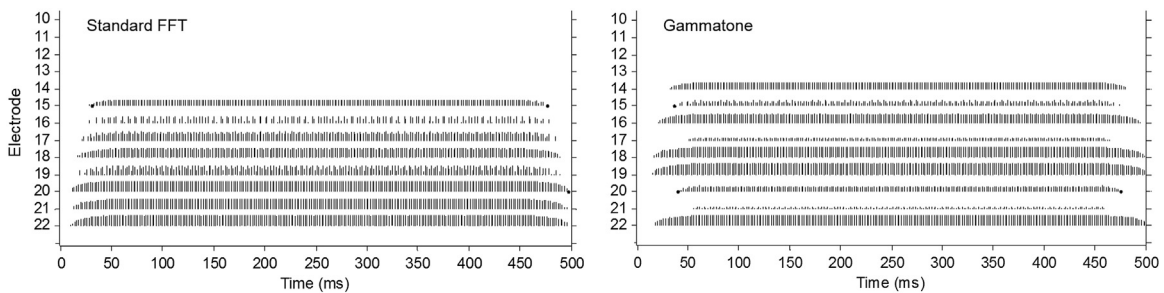| Subject	Age (yrs)	Gender	Etiology	CI experience (yrs)	Stimulation rate per channel (Hz)
S1	37	M	Unknown	4	900
S2	56	F	Traumatic	12	1200
S3	28	M	Congenital	9	900
S4	64	F	Unknown	9	720
S5	63	M	Traumatic	10	500
S6	40	F	Traumatic	7	900
S7	64	M	Traumatic	4	900
S8	55	M	Traumatic	5	720
S9	61	F	Unknown	9	900
S10	62	F	Toxic	2	900

Fig. 2. Example electrodograms for the D3# (155.56 Hz) clarinet tone. The figures show the stimulation patterns obtained with the standard FFT and the Gammatone filterbanks. The x-axis represents time in msec and the electrode numbers (in apex-to-base order) are on the y-axis. Each electrodogram is truncated to the range of active electrodes.
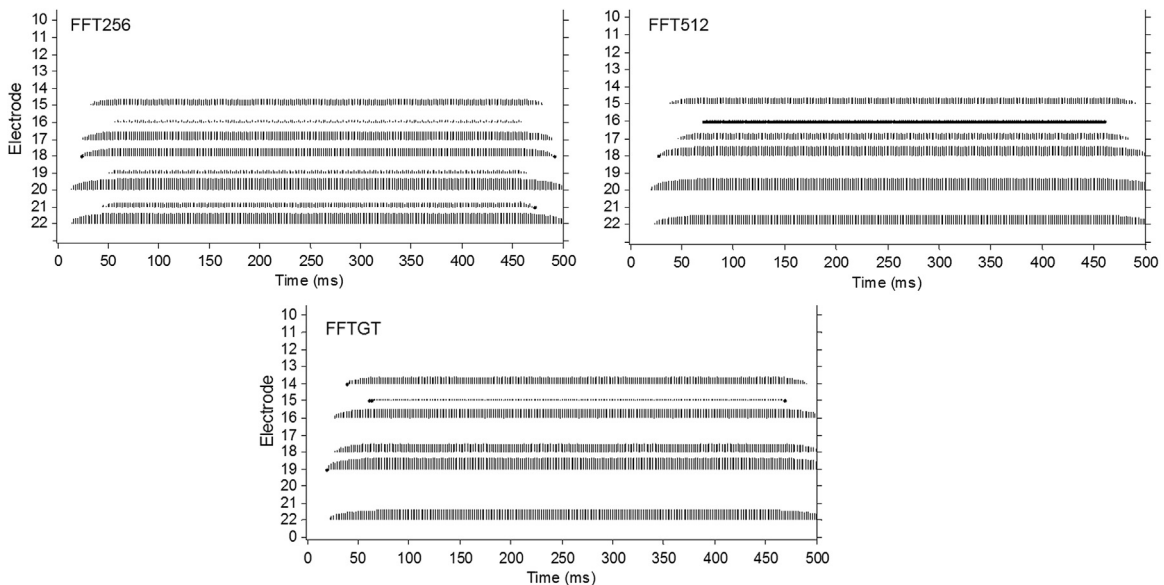

Fig. 3. Three different filterbanks were applied to the D3# (155.56 Hz) clarinet tone; FFT256, FFT512 and FFTGT. The x-axis represents time in msec and the electrode numbers (in apex-to-base order) are on the y-axis and each electrodogram is truncated to the range of active electrodes.
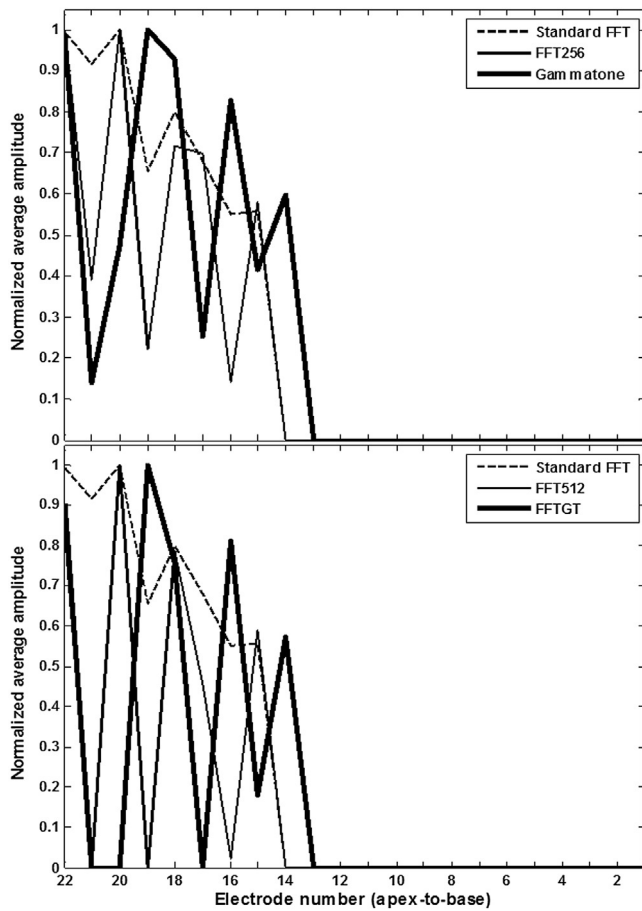
Fig. 4. The average channel amplitudes for the D3# (155.56 Hz) clarinet tone when processed with the standard FFT (dashed line), the FFT256 (solid thin line in the top graph), the Gammatone (solid thick line in the top graph), the FFT512 (solid thin line in the bottom graph) and the FFTGT (solid thick line in the bottom graph) filterbanks. The electrodes are shown in the apex-to-base order.
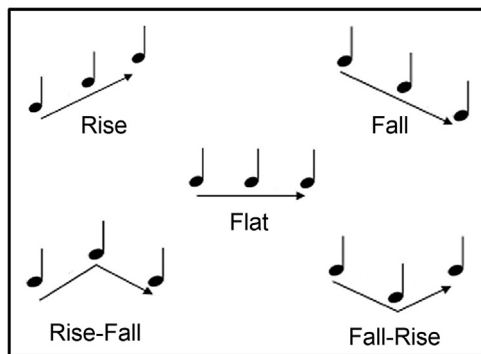

Fig. 5. Five melodic contour patterns used in the MCI test. Each pattern consists of three tones and the interval between successive tones in each pattern can be one to three semitones.

frequency resolution. The FFT filterbank with the Gammatone frequency mapping (FFGT) was also added to the mentioned conditions to investigate the effect of frequency boundaries; this new frequency mapping was possible with 512 points FFT filterbank. The electrodograms for these new conditions applied to the D3# (155.56 Hz) clarinet tone are shown in Fig. 3 to compare with the standard FFT and the Gammatone filterbanks.

The above electrodograms were summarized in Table 2 which shows the resolved harmonics of the D3# (155.56 Hz) clarinet note in different channels. The numbers in the table are the electrode numbers the same as it had been shown for the electrodograms; with the highest number for the apical electrode. In order to define the resolved harmonics in each channel, five stimuli with the same frequencies as each harmonic or fundamental frequency and the same scaling for the amplitudes in the original stimulus (D3# clarinet note) were produced and then were used as audio inputs to the coding strategy with five different filterbanks. This led to 25 (5 stimuli × 5 filterbanks) electrodograms which could show the activated channels for each stimulus. Note that no activation is shown for 6F0 harmonic because of the small amplitude (8.5%) which was defined for this harmonic.

Table 2 shows that with the standard FFT, the Gammatone or even the FFT256 filterbanks one specific harmonic can activate more than one channel compared to the FFT512 and the FFTGT filterbanks. As is well known, electrical current spreads out widely along the cochlea and excites a wide range of populations of auditory nerve fibers which leads to a decrease in the selectivity and the number of effective channels (Undurraga et al., 2012). Thus, spatial spread of the electric field has a major impact on the spectral resolution of CI users and decreases the excitability of the affected neural population. When the number of activated channels for one specific frequency is smaller (as with the FFT512 and the FFTGT filterbanks), it may help to reduce spread and to improve the performance of the CI recipients.

Fig. 4 shows the normalized average channel amplitudes of the 22 channels for the clarinet D3# (155.56 Hz) stimulus processed with the standard FFT, Gammatone, FFT256, FFT512 and FFTGT filterbanks. To extract the average amplitudes values, only the steady state part of the electrodograms from 50 msec to 450 msec was considered and the transient parts were removed. The standard FFT filterbank (dashed line) is shown as a reference for the comparison in the top and the bottom graphs of Fig. 4. Conditions with the Gammatone frequency mapping (the Gammatone and the FFTGT filterbanks) are shown in a solid thick line. According to the different frequency mapping of the FFT and the Gammatone filterbanks, the positions of the peaks as well as the troughs are not at the same electrodes and shifts towards the base for the Gammatone filterbank. The peakedness (width of a peak) of the Gammatone filters is narrower compared to the standard FFT filterbank which may be the reason of the better resolved harmonics. This peakedness is improved for the standard FFT filterbank with increasing the frequency resolution to 512 points.

2.5.2. Melodic contour identification (MCI)

The MCI test designed by Galvin et al. contained nine different patterns each consisting of five tones (Galvin et al., 2007). However it has been shown that the large number of response choices, makes the test exhausting and more demanding for the participants (Omran et al., 2010). Furthermore, having five tones in each pattern provides more cues for the participants to distinguish the patterns and makes it easier for them to select the correct answer. Thus, to avoid the aforementioned problems, five different patterns with three tones in each pattern were used as illustrated in Fig. 5.

The test was carried out using MACarena software (Lai and Dillier, 2002) and the stimuli were delivered via a single loudspeaker at 65 dB SPL with the NH subjects seated directly facing the loudspeaker and a distance of 1.5 m. For CI subjects, the stimuli were delivered with direct streaming via NIC and the L34 research processor to their implants using the subjects' clinical map. A touchscreen with corresponding figures of the patterns (Fig. 5) was used for the response interface. In each trial the subjects were presented with one of the five patterns and were asked to press the

Table 2
Resolved harmonics for the D3# (155.56 Hz) clarinet tone with different filterbanks.

Filterbank	F0	3F0	5F0	6F0	7F0
Standard FFT	22	21, 20, 19	19, 18, 17	–	16, 15
Gammatone	22	20, 19, 18	16, 15	–	14
FFT256	22	21, 20, 19	18, 17	–	15
FFT512	22	20	18, 17	–	15
FFTGT	22	19, 18	16	–	14

button with the figure of the corresponding pattern. The root note was selected as C3 (130.81 Hz) and the interval between successive tones in each melodic contour pattern was either one or two semitones for NH subjects and varied from 1 to 3 semitones for CI subjects. Each interval set was tested separately and the order of the patterns was randomized. At the beginning of each set, training was provided and the subjects could listen to the five patterns as many times as they wanted.

For NH subjects, each pattern was repeated 5 times which resulted in 25 trials (5 patterns × 5 repetitions) per interval set. Five different processing filterbanks were tested which ended up with 250 patterns (25 trials × 2 interval sets × 5 filterbanks) per subject. For CI subjects, each pattern was repeated 10 times which resulted in 50 trials (5 patterns × 10 repetitions) per interval set. Two different processing filterbanks (the standard FFT and the Gammatone filterbanks) were tested which ended up with 300 patterns (50 trials × 3 interval sets × 2 filterbanks) per subject. According to the time limitation and the fact that CI subjects were not available for more time-consuming tests, they were tested with only two processing filterbanks to find out whether there is a same trend for CI subjects like NH subjects or not.

2.5.3. Just noticeable difference (JND)

Frequency discrimination as a function of frequency can be measured in different ways; difference limens for frequency or difference limens for change (Sek and Moore, 1995). A JND test which measures difference limens for change may be more useful compared to a pitch discrimination test which measures difference limens for frequency, as a measure of music perception. It was reported that the CI subjects can perceive a change in pitch reasonably accurately but the pitch direction which is tested in pitch discrimination tests is perceived as ambiguous or even in the opposite direction (Fearn, 2001).

The JND test in this study was conducted in an adaptive 2-down, 1-up, 4 Alternative Forced-Choice task (4AFC) and tested with only NH subjects (Fearn, 2001; Gfeller et al., 2002). In each trial the subjects were asked to indicate the tone that was different in pitch from three other tones. The reference tone (the tone that was presented three times in each trial) was selected as the clarinet C3 (130.81 Hz) note for octave 3 and the clarinet C4 (261.62 Hz) note for octave 4. The probe tone (the different tone in each trial) could be varied and was selected from the rest of the 11 semitones in the tested octave. If the response was correct in two trials in a row, the difference between the probe and the reference was decreased but if the subject responded incorrectly, this difference was increased (Kollmeier et al., 1988; Levitt, 1971). Thus two runs (octaves 3 and 4) were conducted for each filterbank which resulted in 10 runs (2 runs × 5 filterbanks) per subject.

The setup for the JND test was the same as the MCI test; the MACarena software (Lai and Dillier, 2002) was used and the stimuli were presented via loudspeaker at 65 dB SPL facing directly in front of the subjects. The loudness of the stimuli was roved ±3 dB to reduce any cues for pitch discrimination that may be caused by differences in the loudness and the participants were advised to listen to the pitch differences and not the loudness (Fearn, 2001; Gfeller

et al., 2002). No feedback or training was provided. 10 reversals were obtained in each run and the last 8 reversals were averaged to get the JND level.

3. Results

3.1. NH subjects

3.1.1. MCI experiment

The MCI results for 6 NH participants are shown in Fig. 6 (one semitone interval) and Fig. 7 (two semitones interval) as confusion matrices; the subjects' responses are shown on the x-axis while the presented patterns are on the y-axis. The diagonal of the matrices shows the mean percentage correct responses and the amount of darkness corresponds to the number of correct responses. For instance, in Fig. 6, for the FFT filterbank, the "Flat" pattern is confused with "Fall-Rise", "Fall" and "Rise" patterns which leads to only 46.7% correct responses and lighter color for the diagonal compared to the other patterns. The average percentage correct responses are shown under the matrix for each of the filterbanks.

ANOVA analysis with three within subject factors (filterbank, interval, and pattern) was performed on the scores taken from all NH participants. The analysis showed a highly significant effect of filterbank ($p < 0.001$). A pairwise comparison using the Bonferroni correction showed that the performance with the Gammatone filterbank was significantly better than the standard FFT filterbank ($p < 0.01$). This significance was also observed for the pairwise comparison of other filterbanks with the standard FFT filterbank.

For each interval the ANOVA with two within subject factors (filterbank and pattern) was performed. A statistically significant effect of filterbank ($p < 0.05$) was obtained for one semitone interval in which a pairwise comparison with the standard FFT filterbank, showed significantly better scores of the FFT512 and the FFTGT filterbanks ($p < 0.01$) and significant scores of the FFT256 and the Gammatone filterbanks ($p < 0.05$). The average responses of NH subjects for each interval with 5 different filterbanks are summarized in Fig. 8.

We further examined whether the subjects improved their performance with any of the filterbanks. We performed the ANOVA with two within subject factors (interval and pattern) for each filterbank. No significant effect for any of the factors was obtained for the FFT512 and the FFTGT filterbanks. The interval factor was significant ($p < 0.05$) for the FFT256 filterbank. Both factors (interval and pattern) were significant for the Gammatone and the standard FFT filterbanks ($p < 0.01$) and the interaction of these two factors was only significant for the Gammatone filterbank ($p < 0.01$).

3.1.2. JND experiment

The average JND result for 6 NH participants is shown in Fig. 9. The x-axis represents the two octaves which were tested while the y-axis shows the threshold which was achieved for distinguishing the differences in pitch in semitone. For instance, number one on the y-axis means that the subjects could distinguish the tones
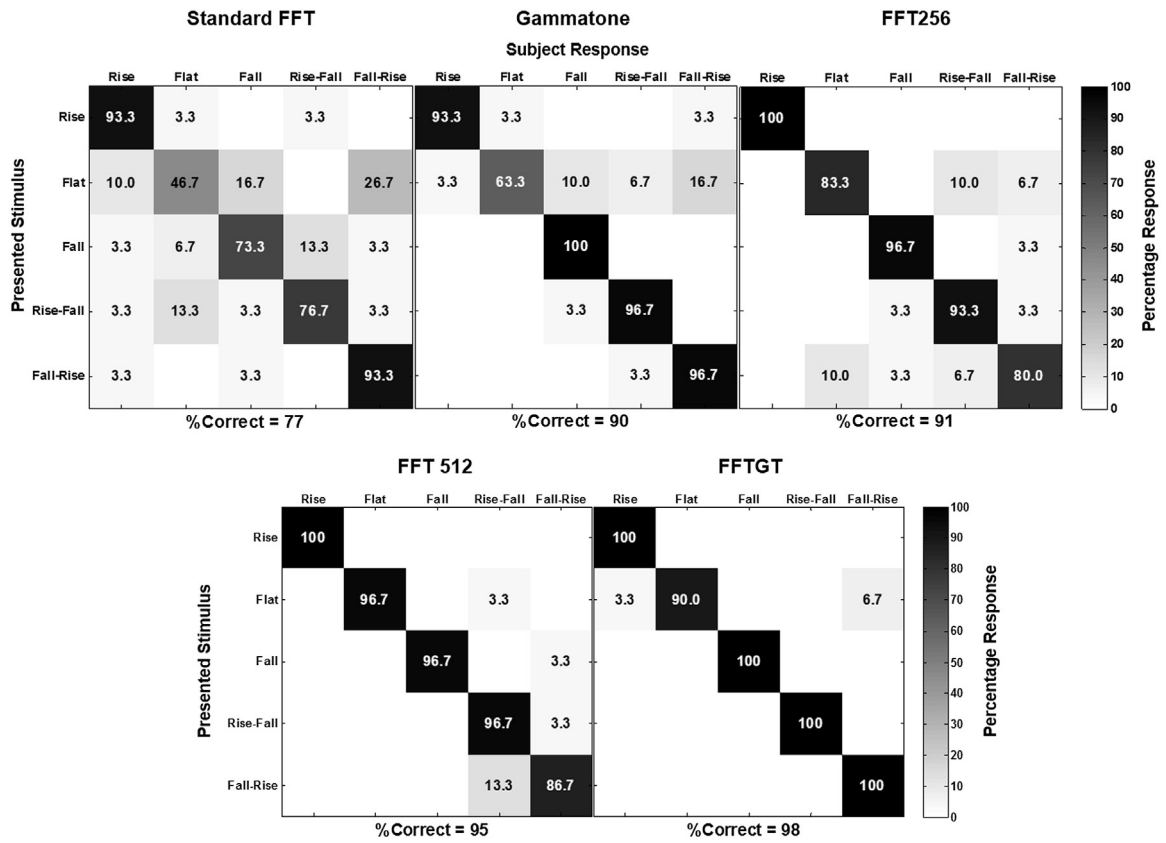
Fig. 6. Confusion matrices for the MCI test with 1 semitone interval. The average responses of 6 NH participants for all trials with one semitone interval were considered for each filterbank. The x-axis represents the subjects response and the presented stimuli are on the y-axis.
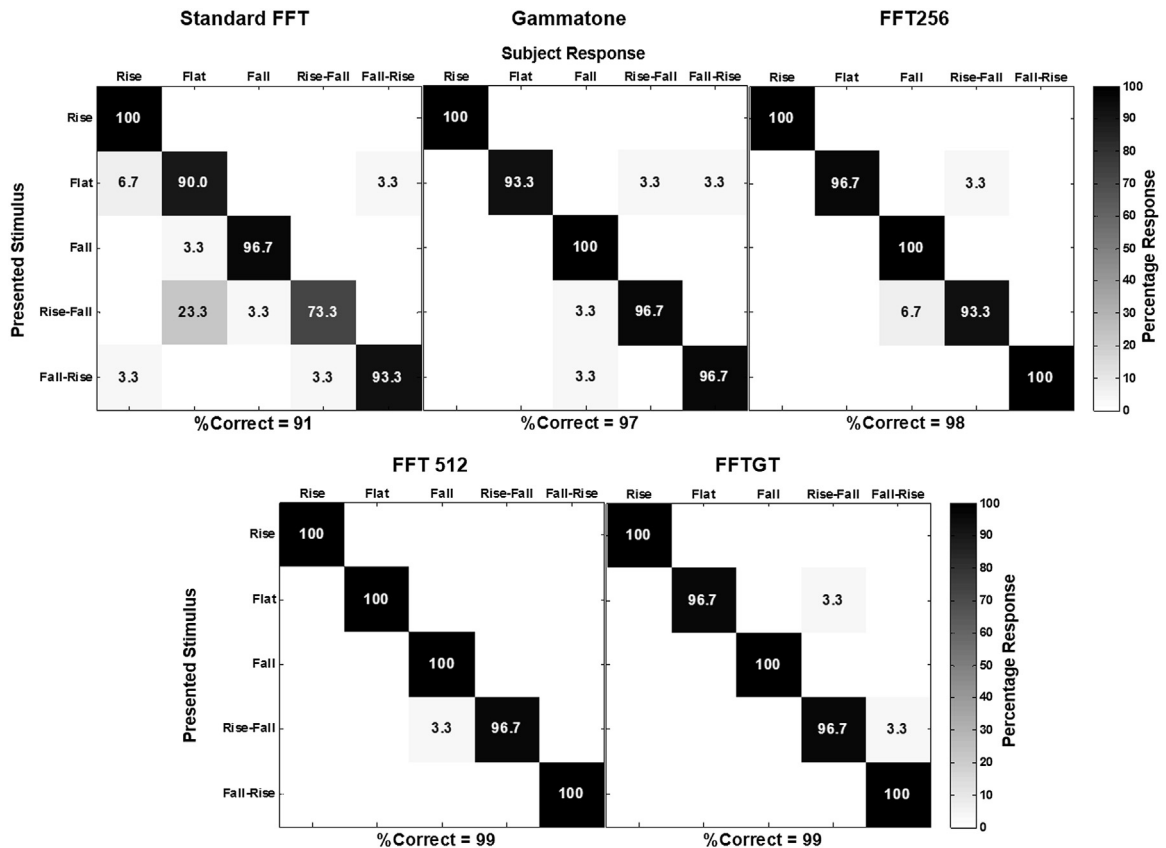

Fig. 7. Confusion matrices for the MCI test with 2 semitones interval. The average responses of 6 NH participants for all trials with two semitones interval were considered for each filterbank. The x-axis represents the subjects response and the presented stimuli are on the y-axis.
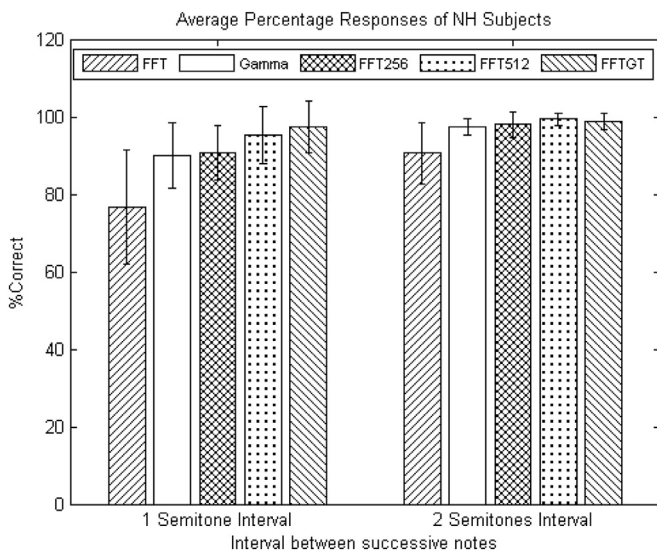
Fig. 8. Average percentage responses of NH subjects with the MCI experiment for either 1 or 2 semitones interval between successive notes in each contour. FFT represents the standard FFT filterbank and Gamma represents the Gammatone filterbank. The x-axis represents the two intervals which were tested while the y-axis shows the percent correct scores of NH subjects. Error bars indicate standard deviation.
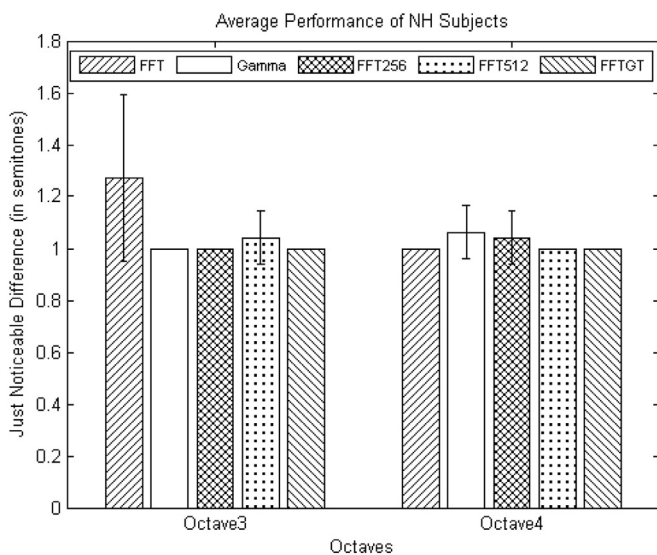

Fig. 9. Average performance of 6 NH participants for the JND test. The x-axis represents the octaves which were tested and the y-axis shows the threshold which was achieved for distinguishing the differences in pitch in semitone. The standard deviations are shown as error bars; FFT represents the standard FFT filterbank and Gamma represents the Gammatone filterbank.

with one semitone difference. The standard deviations are shown as error bars in the figure.

ANOVA analysis with two within subject factors (filterbank and octave) was performed on the threshold scores taken from the NH subjects. The analysis showed a significant effect of filterbank ($p < 0.05$) but no significant effect for octave. The interaction of the two factors had a significant effect ($p < 0.01$). A pairwise comparison using the Bonferroni correction showed no significant effect of filterbank.

3.2. CI subjects

The MCI results for 10 CI subjects with 1 to 3 semitones interval between successive notes in each contour are shown in Fig. 10 as confusion matrices. The x-axis represents the subject's response while the y-axis shows the presented melodic contour patterns and the average percentage correct responses for each filterbank are shown under the matrix. The first column in the figure shows the results for the standard FFT filterbank and the second column shows the results for the Gammatone filterbank and each row represents the results for different interval between successive notes in each contour.

ANOVA analysis with three within subject factors (filterbank, interval and pattern) was performed on the scores of CI subjects. The analysis showed a significant effect of filterbank ($p < 0.05$) and highly significant effect of interval and pattern ($p < 0.001$) but their interaction was not significant. A pairwise comparison using the Bonferroni correction showed a significant effect for Gammatone ($p < 0.05$). For each interval the ANOVA with two within subject factors (filterbank and pattern) was performed. A statistically significant effect of filterbank was obtained only for one semitone interval and a pairwise comparison of the Gammatone filterbank with the standard FFT filterbank showed a significant effect for the Gammatone filterbank ($p < 0.05$).

In order to analyze the performance of the CI subjects with any of the filterbanks, the ANOVA with two within subject factors (interval and pattern) for each filterbank was performed. In this case interval and pattern factors were highly significant for the standard FFT filterbank ($p < 0.001$) and the interval factor was significant for the Gammatone filterbank ($p < 0.05$). The average result of 10 CI subjects for 3 tested intervals with 2 filterbanks is shown in Fig. 11.

4. Discussion

A number of studies evaluated filter frequency boundaries for vowel recognition and F0 discrimination (Fourakis et al., 2004; Geurts and Wouters, 2004; Laneau et al., 2004). However, little is known about the effect of filter cutoff frequencies on musical signals that have a dynamic F0 contour (Kasturi and Loizou, 2007). Apart from that, CI users have much greater difficulty than NH under realistic and demanding situations such as music (Nie et al., 2005; Wilson and Dorman, 2008; Wilson et al., 2005). Therefore, the focus of this study was on MCI and JND tests which can provide insights for improving music perception in CI recipients. The speech recognition in quiet as well as in noise were beyond the scope of this work and will be tested in future studies.

The results from the MCI experiment showed that the choice of filterbank has a significant effect. However, if the performance of the NH subjects were considered with any of the filterbanks, the FFT512 and the FFTGT filterbanks showed already a good performance with one semitone interval between successive notes which led to no significant differences for these two filterbanks. The significant performance difference of the Gammatone and the FFT256 filterbanks in comparison with the standard FFT filterbank showed the importance of the frequency resolution rather than the filter type for the performance of the NH subjects. This is consistent with the Cosentino et al. study which showed that more physiologically-inspired filters in the CI speech processor do not necessarily improve performance (Cosentino et al., 2014). However, an increase in frequency resolution could not be observed between the FFT256 and the FFT512 filterbanks, probably because of a ceiling effect with the FFT512. The importance of the frequency boundaries was investigated and no significant effect was found since the FFT512 showed already a good performance in one semitone interval between successive notes.
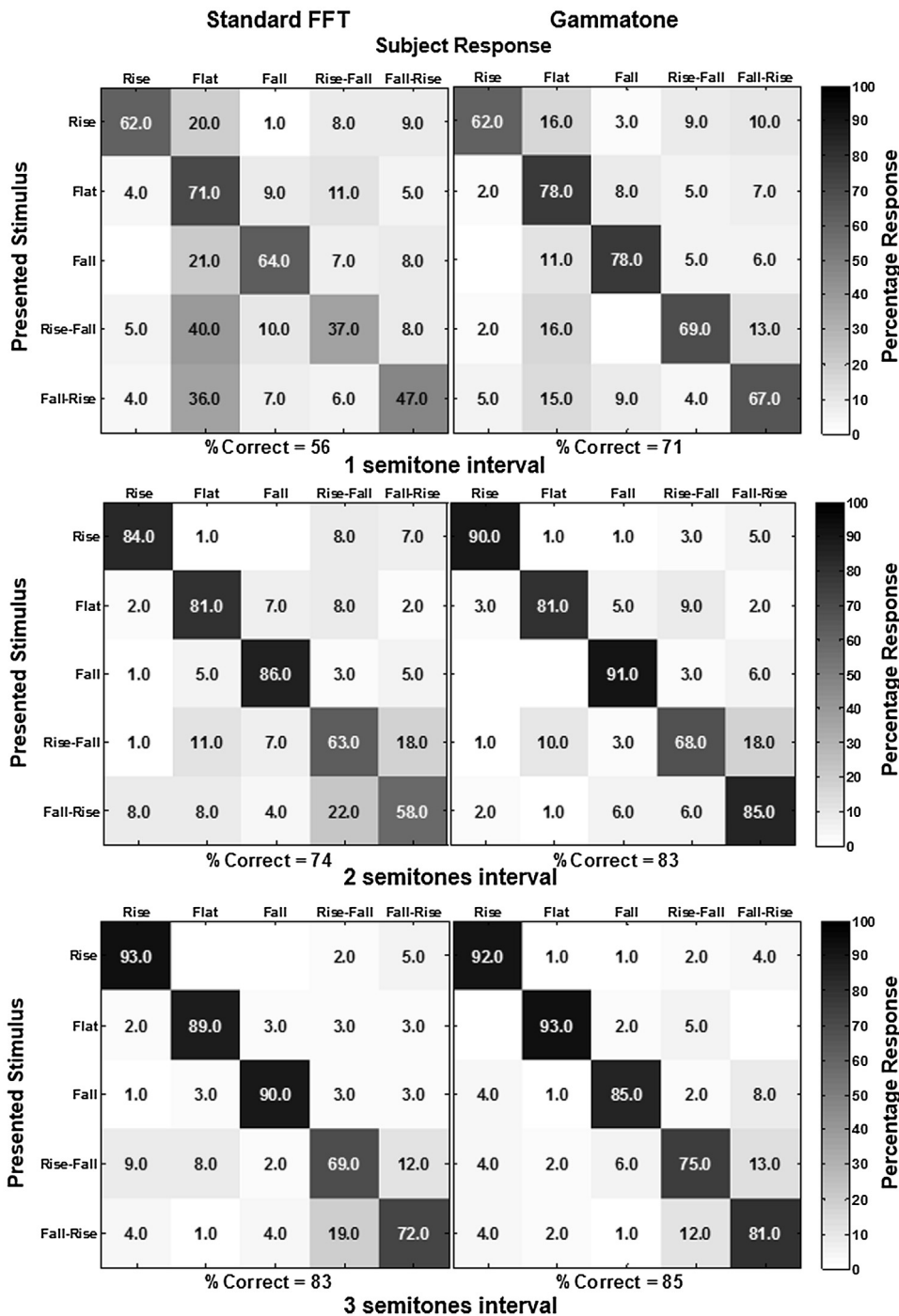
Fig. 10. Confusion matrices for 10 CI participants. The first column shows the results for the standard FFT filterbank and the second column shows the results for the Gammatone filterbank. First row represents the results for 1 semitone interval, second row shows the results for 2 semitones interval and third row shows the results for 3 semitones interval.

The confusion matrices revealed a poor performance of the NH subjects for the "Flat" pattern with the standard FFT filterbank in one semitone interval between successive notes. However, this performance was improved in the higher frequency resolution of the FFT filterbank; the higher the resolution the better the performance. The results of the filterbanks for the two semitones interval were improved and the subjects showed good performance for most of the filterbanks except the standard FFT filterbank for the "Rise-Fall" pattern which had an average percentage response of 73.3%. The performance of the NH subjects with the vocoded stimuli was bet-

ter than the CI subjects for 1 semitone and 2 semitones intervals in the MCI experiment. For instance, the CI subjects had the average percentage response of 56% for the standard FFT filterbank in 1 semitone interval between successive notes while this score was 77% for the NH subjects. This shows that vocoded stimuli with the neural-based vocoder are not the same as real CI simulations. Thus, the results of NH subjects with the vocoder can be affected by some parameters of the vocoder which are discussed below.

First, the percentage of neural survival was chosen as 100% in the neural-based vocoder. This means it was assumed in the vocoder
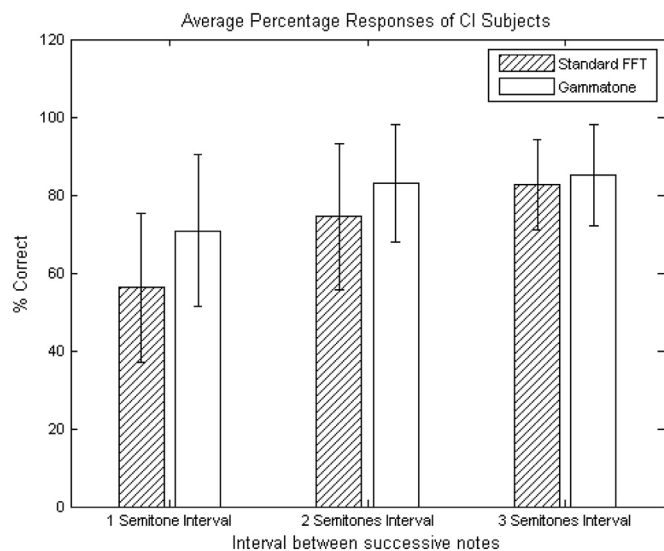
Fig. 11. Average percentage responses of 10 CI subjects with the MCI experiment for 1 to 3 semitones interval between successive notes in each contour. The standard deviations are shown as error bars. The x-axis represents the three intervals which were tested while the y-axis shows the percent correct scores of CI subjects.

that there were no dead regions or irregular patterns of neural survival in the cochlea. However, CI subjects may have some dead regions which can lead to the shifted or split excitation patterns (Zhu et al., 2012). Secondly, the absolute refractory period was set to a random value between 1 μsec and 300 μsec (El Boghdady et al., 2016). This is in contrast to the absolute refractory value used in other studies (Cohen, 2009b; Miller et al., 2001; Morsnowski et al., 2006). The spatial spread of neural excitation was also considered with the same profile shape for all the electrodes in the array with its peak at the stimulated electrode and decrement towards neighboring electrodes. However, it was shown that this profile can be different for different electrode locations (Cohen, 2009a). Apart from that, the function for spread of neural excitation was a simplified version of the excitation pattern and was not exactly the same as the excitation patterns that can be extracted from ECAP recordings (Hughes and Stille, 2010). This pattern can also be different from one CI subject to another which was not the case for NH subjects where the same excitation profile was used for the vocoded sounds. Another difference between NH and CI subjects' results can arise from their different pitch sensitivity and different musical experience. Finally, one should keep in mind that in the noise-band vocoder the noise bands are different from deterministic pulse trains which are used in the CIs (Laneau et al., 2006). The effect of these parameters on the performance of NH subjects need to be investigated in future studies.

The FFT512 filterbank had the same frequency resolution in the lowest frequency channel as the Gammatone filterbank and as it was mentioned before, it was not possible to get the same frequency resolution in the lowest frequency channel with the lower resolutions of the FFT filterbank (the FFT256 or the standard FFT filterbank). The FFT512 filterbank used the Hanning window and the frame size of 512 samples which imposed 32 msec delay with a typical sampling rate (16 KHz) of CI processors. However, this delay was 16 msec for the Gammatone filterbank since the frame size of 256 samples was used for it. Apart from that when a specific harmonic activates more channels (as shown in Table 2), it can deteriorate discrimination of other harmonics on adjacent channels and increase the channel interaction. Therefore, the FFT512 filterbank had a better discrimination of harmonics and performance

compared to the Gammatone filterbank but at the cost of longer delay and lower temporal resolution.

For the MCI experiment with CI subjects, 17 participants took part in the test and 10 of them could distinguish the melodic contour patterns. All of 10 CI subjects who participated in the study were using the ACE strategy and no other variation such as SPEAK (spectral peak coding) strategy since earlier studies showed a preference of CI recipients for the ACE strategy compared to the SPEAK strategy (Fu et al., 2004; Pasanisi et al., 2002; Skinner et al., 2002). Based on the information provided by CI recipients, it seems that CI users without musical background may need additional training to be able to perform the experiment. The results from the MCI experiment with CI subjects showed a statistically significant effect for the Gammatone filterbank. The smaller the semitone interval between successive notes was the more benefit from the Gammatone filterbank was achieved. Thus CI subjects showed 15% improvement for 1 semitone interval, 9% for 2 semitones interval and 2% for 3 semitones interval for the Gammatone filterbank. Their performance was improved particularly for the "Rise-Fall" pattern which was difficult for them to distinguish with the standard FFT filterbank. CI subjects got 37% correct for this pattern with the standard FFT filterbank in 1 semitone interval while the percentage correct response was increased to 69% with the Gammatone filterbank (Fig. 10).

The results from the JND experiment with NH subjects showed the importance of the filterbank for the performance of the subjects. There was no significant effect for the octaves which were tested. However, there was an improvement for the standard FFT filterbank performance from octave 3 to octave 4.

The default "power-sum" option for an envelope extraction was used in the NMT from Cochlear Corp. Electrodograms with "power-sum" and "vector-sum" for synthetic clarinet notes in octave 3 and octave 4 were compared for the ACE coding strategy. RFcap (Lai and Dillier, 2013) was used for the comparison and the electrodograms were found to be quite similar from C3 to A3 notes in both versions of the envelope extraction. This means that the results of the MCI experiment will not be affected by changing from "power-sum" to "vector-sum" for the envelope extraction (in the MCI experiment the largest interval between successive notes is 3 semitones interval, thus for the "Rise" pattern with the root note of C3 the highest note will be F3#). However, the envelope extraction mode may affect the results of JND experiment, especially for octave 4, and should be investigated further.

This study was similar to the Laneau et al. study (Laneau et al., 2004) in some aspects with a few differences. First, the Gammatone filterbank implementation was an all-pole design of the filterbank which could reduce the computational effort of implementation by approximately 50% and would be beneficial for CI processors. Secondly, the covered cutoff frequency range along the cochlea was matched to the covered cutoff frequency range in the ACE coding strategy. Apart from that, the cutoff frequencies distribution of the FFT and the Gammatone filterbanks was matched. In order to separate the effects of different cutoff frequencies and filter type in our study, conditions such as the FFT512 and the FFTGT filterbanks were added. In addition, the effect of increasing the frequency resolution of the FFT filterbank on the participants' performance was investigated. Finally, this study was performed with 10 CI participants who were using the ACE coding strategy as their own clinical coding strategy. The Laneau et al. study (Laneau et al., 2004) included only 4 CI participants whereby 2 of them were using the ACE strategy and the other 2 the SPEAK strategy as their clinical coding strategy.

5. Conclusions

In the present study, we investigated a frequency domain implementation of an all-pole IIR Gammatone filterbank in con-

junction with the ACE coding strategy, the clinical standard strategy implemented in speech processors of Cochlear Corp. Significant improvement for NH and CI participants was demonstrated in the MCI test with the Gammatone filterbank. But it was also shown that this improvement may be due to the usage of a longer frame size for the Gammatone filterbank and not just due to the frequency boundaries of the Gammatone filters. However, the frequency resolution and channel spacing with the Gammatone filterbank can be adapted to the resolution of the auditory system.

The total delay of the Gammatone filterbank can be made smaller than the delay of the FFT filterbank with the same frequency resolution at low frequencies. This delay makes the implementation impractical for a real-time processing of CI processors since the auditory information will fall out of synchronization with visual information and interferes with lip reading. In addition to that, the results from this study showed that the FFT filterbank performance was improved with increasing frequency resolution from 128 points to 512 points. This improvement came at the cost of decreasing temporal resolution and higher delay compared to the Gammatone filterbank.

In this study the covered frequency range for the Gammatone filterbank was chosen to match the covered frequency range of the standard FFT filterbank in the ACE strategy. However, the Gammatone filterbank also allows to specify the lowest cutoff frequency at will, whereas the FFT filterbank's lowest cutoff frequency is restricted by the bandwidths of the bins. It may be advantageous to set the Gammatone filterbank's lowest cutoff frequency to even lower values and thereby enlarge the covered frequency range. Whether this results in an improvement for the CI recipients needs to be investigated in future studies.

Conflict of interest statement

All authors declare no actual or potential conflicts of interest.

Acknowledgments

References

Cohen, L.T., 2009a. Practical model description of peripheral neural excitation in cochlear implant recipients: 2 Spread of the effective stimulation field (ESF), from ECAP and FEA. Hear. Res. 247, 100–111.

Cohen, L.T., 2009b. Practical model description of peripheral neural excitation in cochlear implant recipients: 5: refractory recovery and facilitation. Hear. Res. 248, 1–14.

Cosentino, S., Falk, T.H., McAlpine, D., 2014. Marquardt T. cochlear implant filterbank design and optimization: a simulation study. IEEE/ACM Trans. Audio Speech Lang. Process. 22, 347–353.

El Boghdady, N., Kegel, A., Lai, W.K., Dillier, N., 2016. A neural-Based vocoder implementation for evaluating cochlear implant coding strategies. Hear. Res. 333, 136–149.

Fearn, R.A., 2001. Music and Pitch Perception of Cochlear Implant Recipients. University of New South Wales.

Fourakis, M.S., Hawks, J.W., Holden, L.K., Skinner, M.W., Holden, T.A., 2004. Effect of frequency boundary assignment on vowel recognition with the nucleus 24 ACE speech coding strategy. Am. Acad. Audiol. 15, 281–299.

Fu, Q.-J., Hsu, C.-J., Horng, M.-J., 2004. Effects of speech processing strategy on chinese tone recognition by nucleus-24 cochlear implant users. Ear Hear. 25, 501–508.

Galvin III, J.J., Fu, Q.-J., Nogaki, G., 2007. Melodic contour identification by cochlear implant listeners. Ear Hear. 28, 302–319.

Geurts, L., Wouters, J., 2004. Better place-coding of the fundamental frequency in cochlear implants. Acoust. Soc. Am. 115, 844–852.

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

Goto, M., 2004. Development of the RWC music database. Proceedings of the 18th International Congress on Acoustics (ICA), 553–556.

Hohmann, V., 2002. Frequency analysis and synthesis using a Gammatone filterbank. Acta Acust. United Acust. 88, 433–442.

Holdsworth, J., Patterson, R., Nimmo-Smith, I., Rice, P., 1988. Implementing a GammaTone Filter Bank.

Hughes, M.L., Stille, L.J., 2010. Effect of stimulus and recording parameters on spatial spread of excitation and masking patterns obtained with the electrically evoked compound action potential in cochlear implants. Ear Hear. 31, 679–692.

Kasturi, K., Loizou, P.C., 2007. Effect of filter spacing on melody recognition: acoustic and electric hearing. J. Acoust. Soc. Am. 122, EL29–EL34.

Kollmeier, B., Gilkey, R.H., Sieben, U.K., 1988. Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. J. Acoust. Soc. Am. 83, 1852–1862.

Lai, W., Dillier, N., 2002. A flexible computer-based speech testing environment. Proceedings of the 7th International Cochlear Implant Conference.

Lai, W., Dillier, N., 2013. RFcap A software analysis tool for multichannel cochlear implant signals. Cochlear Implants Int. 14, 107–116.

Laneau, J., Wouters, J., Moonen, M., 2004. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J. Acoust. Soc. Am. 116, 3606–3619.

Laneau, J., Moonen, M., Wouters, J., 2006. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. Acoust. Soc. Am. 119, 491–506.

Laneau, J., 2005. When the Deaf Listen to Music-Pitch Perception in Cochlear Implants (Als Doven Naar Muziek Luisteren-Toonhoogtewaarneming Met Een Cochleair Implantaat). Katholeike Universiteit, Leuven.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.

Lyon, R., 1997. All-pole models of auditory filtering. Divers. Audit. Mech., 205–211.

Milczynski, M., Wouters, J., Van Wieringen, A., 2009. Improved fundamental frequency coding in cochlear implant signal processing. J. Acoust. Soc. Am. 125, 2260–2271.

Miller, C.A., Abbas, P.J., Robinson, B.K., 2001. Response properties of the refractory auditory nerve fiber. JARO 02, 216–232.

Moore, B.C., Glasberg, B.R., 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. 74, 750–753.

Morsnowski, A., Charasse, B., Collet, L., Killian, M., Muller-Deile, J., 2006. Measuring the refractoriness of the electrically stimulated auditory nerve. Audiol. Neurotol. 11, 389–402.

Nie, K., Stickney, G., Zeng, F.-G., 2005. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans. Biomed. Eng. 52, 64–73.

Nimmons, G.L., Kang, R.S., Drennan, W.R., Longnion, J., Ruffin, C., Worman, T., Yueh, B., Rubinstein, J.T., 2008. Clinical assessment of music perception in cochlear implant listeners Otology & neurotology: official publication of the American Otological Society. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. 29, 149–155.

Nogueira, W., Büchner, A., Lenarz, T., Edler, B., 2005. A psychoacoustic NofM-type speech coding strategy for cochlear implants. EURASIP J. Appl. Signal Process. 18, 3044–3059.

Omran, S.A., Lai, W., Dillier, N., 2010. Pitch ranking, Melody contour and instrument recognition tests using two semitone frequency maps for Nucleus Cochlear Implants. EURASIP J. Audio Speech Music Process., 1–16.

Pasanisi, E., Bacciu, A., Vincenti, V., Guida, M., Berghenti, M.T., Barbot, A., Panu, F., Bacciu, S., 2002. Comparison of speech perception benefits with SPEAK and ACE coding strategies in pediatric Nucleus CI24 M cochlear implant recipients. Int. J. Pediatr. Otorhinolaryngol. 64, 159–163.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., Rice, P., 1987. An efficient auditory filterbank based on the gammatone function. Paper Presented at a Meeting of the IOC Speech Group on Auditory Modelling at RSRE.

Sek, A., Moore, B.C., 1995. Frequency discrimination as a function of frequency, measured in several ways. J. Acoust. Soc. Am. 97, 2479–2486.

Sit, J.-J., Simonson, A.M., Oxenham, A.J., Faltys, M.A., Sarpeshkar, R., 2007. A low-Power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. IEEE Trans. Biomed. Eng. 54, 138–149.

Skinner, M.W., Holden, L.K., Whitford, L.A., Plant, K.L., Psarros, C., Holden, T.A., 2002. Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. Ear Hear. 23, 207–223.

Slaney, M., 1993. An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. Apple Computer. Perception Group, TechRep, 1-41.

Swanson, B., Mauch, H., 2006. Nucleus MATLAB Toolbox 4.20, Software User Manual. Cochlear Ltd., Australia, pp. 1–27.

Swanson, B.A., 2008. Pitch Perception with Cochlear Implants.

Undurraga, J.A., Carlyon, R.P., Macherey, O., Wouters, J., Van Wieringen, A., 2012. Spread of excitation varies for different electrical pulse shapes and stimulation modes in cochlear implants. Hear. Res. 290, 21–36.

Wilson, B.S., Dorman, M.F., 2008. Cochlear implants: a remarkable past and a brilliant future. Hear. Res. 242, 3–21.

Wilson, B.S., Schatzer, R., Lopez-Poveda, E.A., Sun, X., Lawson, D.T., Wolford, R.D., 2005. Two new directions in speech processor design for cochlear implants. Ear Hear. 26, 73S–81S.

Yin, H., Hohmann, V., Nadeu, C., 2011. Acoustic features for speech recognition based on gammatone filterbank and instantaneous frequency. Speech Commun. 53, 707–715.

Zeng, F.-G., Rebscher, S., Harrison, W., Sun, X., Feng, H., 2008. Cochlear implants: system design, integration, and evaluation. IEEE Rev. Biomed. Eng. 1, 115–142.

Zhu, Z., Tang, Q., Zeng, F.-G., Guan, T., Ye, D., 2012. Cochlear-implant spatial selectivity with monopolar, bipolar and tripolar stimulation. Hear. Res. 283, 45–58.