Analysis of the drilling sound component from expert performance in a maxillo-facial surgery

Pablo F. Hoffmann[1], *Florian Gosselin*[2], *Farid Taha*[3]

[1]Acoustics, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg Ø, Denmark
[2]CEA, LIST, Laboratoire d'Interfaces Sensorielles, 18 route du Panorama, BP6, Fontenay-aux-Roses, F-92265 France
[3]Service de Chirurgie Maxillo-Faciale CHU Nord, Place Victor Pauchet 80054 Amiens, France
pfh@es.aau.dk, florian.gosselin@cea.fr, taha.farid@chu-amiens.fr

ABSTRACT

Auditory displays can have a great potential in surgical simulators that aim at training skills associated to the correct interpretation of auditory information. Here, we present preliminary results in the analysis of the sound produced by the drilling procedure in a maxillo-facial surgery when performed by expert surgeons. The motivation of this work is to find relevant acoustic parameters that allow for an efficient synthesis method of auditory displays so that they can effectively convey information on expert surgical drilling.

1. INTRODUCTION

Efforts to improve medical training in surgery are increasingly focusing on the use of virtual reality (VR) and multimodal technologies due to their potential to significantly enhance learning of complex medical or surgical procedures [1]. With respect to the multimodal aspect substantial attention has been paid to the visual and haptic modalities [2]. Relatively little or no attention has been given to the auditory modality. Perhaps, this is a consequence of many simulators implemented for surgical procedures in which sound may not play a significant functional role.

Surgical procedures involving drilling are expected to have a significant auditory component. A surgery that has recently received considerable attention in the development of training simulators is temporal bone dissection [3, 4, 5]. In this surgery, dissection begins with drilling of the temporal bone in the mastoid region. Simulators of this surgery have modeled the drilling sound as the sum of few sinusoids harmonically related [3, 4]. For multimodal interaction the frequency of the sinusoids is changed proportional to the force applied to the haptic interface [3]. In addition to force, the type of drill burr is also used to modulate the frequency of the harmonics [4].

In [6] a speech codec algorithm is used to synthesize the drilling sound during temporal bone dissection. The purpose of this study was to analyze the spectral signature of the sound and also to assess the potential of this codec for encoding this signature. Results showed that this codec could not provide an adequate spectral resolution to discriminate the different frequency components of the drilling sound. A more recent study has also examined the spectral features of the drill-bone contact during temporal bone dissection [7]. This study found evidence of a harmonic relation between the sinusoidal components. It was also observed that the spectrum changed in a consistent manner as a function of bone structure. Spectral peaks were slightly shifted down for the thicker structure. This is consistent with the notion that during drilling, the resistance offered by the bone structure to the drill will be higher for a thicker structure, and this will produce a reduction in the frequency revolutions of the drill motor, which, in turn, will cause a decrease in the energy at high frequencies.

Another surgical procedure in which the sound of drilling may play an important role is the maxillo-facial surgery called Epker Osteotomy. In this surgery, surgeons have to split the mandible by drilling the bone at the junction of the mandibular ramus and body. The drilling needs to be sufficiently deep to make the splinting easier. It is critical that the surgeon stops drilling before reaching the nerve area, which is hidden in the spongy bone underlying the cortical part of the bone. The main risk here is to damage this nerve, responsible for the sensitivity of the teeth and half of the lower lip. Damage to this nerve is irreversible and is very handicapping for the patient. For these reasons the Epker Osteotomy surgery remains very stressful for expert surgeons, even after years of practice. In this context, the advantage of a surgical simulator is that it allows trainees to proceed in a safe environment where mistakes can be made many times without compromising the patient’s safety.

Perception of bone compliance appears to be an important landmark related to the prospective and fine motor control skills required to avoid damaging of the nerve in a maxillo-facial surgery. In the framework of the European project SKILLS (www.skills-ip.eu), a multimodal training simulator is under development with the goal of implementing effective training protocols to enhance learning of this surgery. Among the possibilities to exploit are the multimodal cues available for the detection of changes in bone compliance, i.e. changes in haptic feedback, changes in bone color, and changes in drilling sound.

In the present study we report on a preliminary analysis of the drilling sound recorded during maxillo-facial surgery. Section 2 describes the capturing system used to record the drilling sound. Sections 3 reports on the sinusoidal model and psychoacoustic principles used to explore the possibilities for an efficient synthesis. Finally, section 4 provides an insight onto some points for future work.

2. PLATFORM AND DATA ACQUISITION

A multimodal data acquisition system was implemented for a capturing campaign conducted in the anatomy laboratory of the Rouen University Hospital in France. Multimodal data was acquired for three expert surgeons while they were performing a complete maxillo-facial surgery on cadavers. Each surgeon performed the surgery on one side of the head of two different cadavers (one left mandible and one right mandible for each surgeon).

The multimodal platform was composed of sensors for capturing position and orientation of the tools and surgeons’ forearm, arm and head; forces and torques between the tool and surgeons’ hand; acceleration of the tools; EMG activity from the surgeons’ right arm, as well as picture, video and

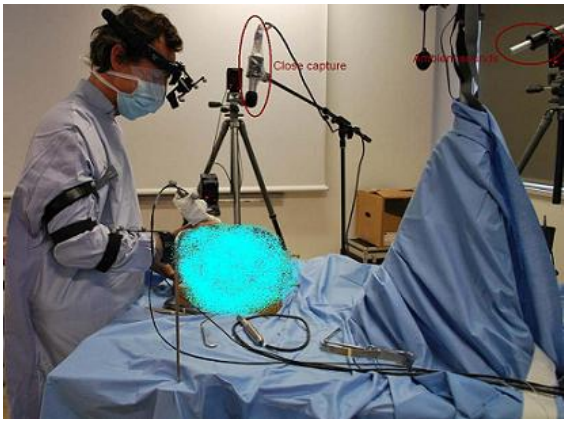

Figure 1. *Setup for multimodal capturing. For audio capturing, microphones are indicated by red ellipses.*

audio. The multimodal capturing setup is depicted in Figure 1.

For audio capturing two Superlux CM-H8K condenser microphones were employed (20 mm diameter, 150 mm length). One microphone was placed on a boom at approximately 40 cm above the region of drilling and pointing directly to the zone of drill-bone contact with its directivity pattern set to super cardioid. This microphone will be referred to as the close-to-source microphone. The second microphone was located at a farther distance from the drilling, and its purpose was capturing of the acoustic surrounding (omni-directional pattern).

The recording system consisted of a laptop equipped with a Firewire audio interface (Edirol FA66). The analogue output of the microphones was connected to the input of the A/D converter of the audio interface. Digital audio was recorded as 24-bit floating point samples at 192 kHz sampling rate. The open-source software audacity was used for recording and storing of audio files.

Figure 2 shows an excerpt of a recording in which two instances of drill-bone contact can be observed. In addition to the time-domain representation a spectrogram shows how the energy is distributed in different frequency regions. In general, it is possible to observe that most of the energy is concentrated between 800 Hz to 12 kHz. High-energy regions can be observed for narrowband components at about 1.5 and 8 kHz. To better illustrate the energy distribution around these frequency components Figure 3 shows the spectrum for the two drill-bone contact instances for times 0.29 and 0.78 s respectively. It can also be observed that several other spectral peaks are distributed in between 1.5 and 8 kHz as well as at lower frequencies (500 and 800 Hz approximately) with a decrease in energy towards frequencies higher than 10 kHz. This is representative of the spectral signature observed on other drilling excerpts from the same surgeon as well as for the other two surgeons.

3. SOUND ANALYSIS AND SYNTHESIS

Only close-to-source recordings were used in the analysis. Audio recordings were downsampled to 48 kHz and re-quantized as 16-bit samples. The approach to synthesize the drilling sound was based on a sinusoidal model [8].

The sinusoidal model approximates an input signal $x(n)$ as a sum of sinusoids whose parameters vary in time. The approximated signal is given by

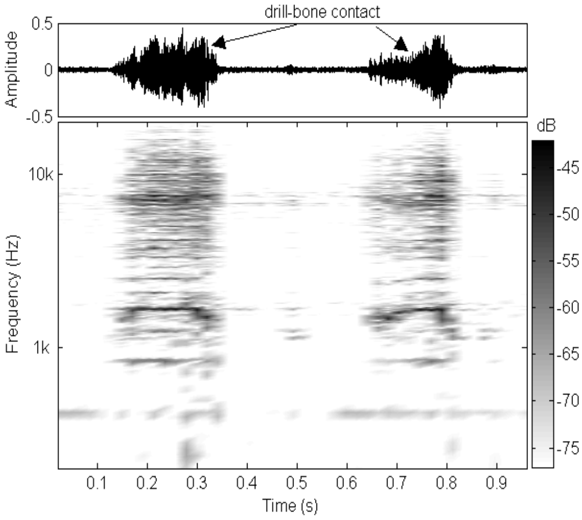

Figure 2. *Excerpt of audio recording that shows two drilling instances as indicated by the arrows. A spectrogram is plotted below to illustrate how the spectral energy is distributed during drilling.*

$$\hat{x}(n) = \sum_{m=0}^{M-1} a_m(n) \cdot \cos(2\pi \cdot f_m(n) + \phi_m(n)) \quad (1)$$

where f_m, a_m and ϕ_m are the parameters of the model and correspond to frequency, amplitude and phase variation of the *m-th* sinusoid. The task of the model is to find these three parameters for a set of sinusoids such that their sum can provide an accurate representation of the original signal.

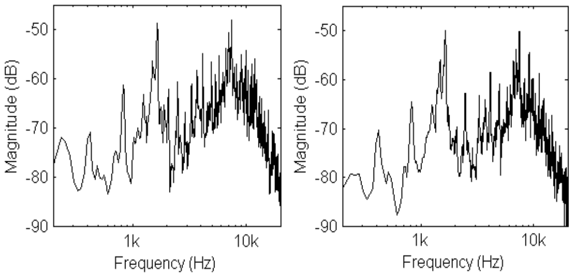

Figure 3. *Power spectrum of the drilling. The spectrum was calculated from the audio recording portions indicated by the arrows in Figure 2.*

3.1.Parameter estimation

Sinusoidal parameters $\{f_m, a_m, \phi_m\}$ were extracted using a sliding-window framework and an analysis-by-synthesis method based on the matching pursuits algorithm [9]. The analysis was carried out on windowed segments of the original drilling signal. A 256-coefficient Hanning window $w(n)$ was employed with 50% overlap between segments. For each windowed segment $x_w(n) = x(n) \cdot w(n)$, individual sinusoids were estimated sequentially. That is, the estimation algorithm started by comparing the current windowed segment to a dictionary of complex exponentials

$$g(k,n) = \hat{w}(n) \cdot e^{j2\pi \frac{k}{K} n} \quad (2)$$

with $n = 0 \ldots N-1$ corresponding to the time index in the windowed segment ($N = 256$), and $k = 0 \ldots K-1$ to the frequency index. The complex exponential (sinusoid) showing the largest correlation with current segment was selected and subtracted from the same segment to form a residual. This procedure was then repeated on the residual signal and iterated 100 times in order to estimate 100 sinusoids (M=100).

Complex exponentials *g(k,n)* were set to span a frequency range between 200 Hz and 15 kHz with a resolution of 20 Hz ($K = 741$). The weighting function in eq. 2

$$\hat{w}(n) = \frac{w(n)}{\|w(n)\|}$$

corresponded to a normalized version of the analysis windows, which is required for a correct synthesis. Correlations were computed via the inner product between the K=741 complex exponentials and the current residual signal. The index *kmax* corresponding to the largest inner product α was used to find the value of the frequency parameter. The inner product α is a complex value, and thus the amplitude and phase parameters were derived from α by

$$a_{k\max} = 2 \cdot \frac{\|w(n)\|}{\sum_{n=0}^{N-1} w(n)} |\alpha| \qquad (3)$$

and

$$\phi_{k\max} = \angle \alpha \qquad (4)$$

M=100 was arbitrarily chosen according to informal listening of the resulting synthesized signal for different numbers of iterations. The error signal was computed by

$$e(n) = x_w(n) - w(n)\hat{x}(n) \qquad (5)$$

and the full output signal was reconstructed by overlap-add of consecutive synthesized windowed segments.

Although the signal synthesized with the first 100 sinusoids was comparable to the original signal, it was considered redundant in the sense that not all 100 sinusoids may be audible if one takes into account the masking characteristics of the human ear. Thus, in order to reduce the order of the sinusoidal model and still keep an acceptable perceptual quality, masking curves were constructed and those sinusoids whose magnitudes were below the curve were removed. Similar approaches have been reported in [10, 11] for perceptually weighted matching pursuits sinusoidal coding, and for efficient additive synthesis [12].

3.2. Frequency masking model

For each of the 100 sinusoids an excitation pattern was computed using the procedure described in [13]. In order to incorporate the dependency of the shape of excitation pattern on the sound pressure level (SPL) of the signal, absolute SPLs from the different sinusoids were required. Because information about the real SPLs was not available, it was assumed that a sinusoid of normalized amplitude 1 was equivalent to approximately 96 dB SPL considering a 16-bit quantization. This procedure is usually employed in audio coding algorithms [14].

After the individual excitation patterns were computed their corresponding levels were shifted down by 10 dB. This offset represents the level difference between the masker and the masking threshold. The overall masking curve was constructed as the sum of the powers of the individual excitation patterns. In addition, the curve for the absolute hearing threshold *A(f)* was added to the overall masking curve. This curve was computed using the approximation proposed in [15] (as referenced in [14]). This approximation includes an approximation to the transfer function of the outer and middle ear, and is given by

$$A(f) = 3.64(\frac{f}{1000})^{-0.8} - 6.5e^{-0.6(\frac{f}{1000}-3.3)^2} + 10^{-3}(\frac{f}{1000})^4 \qquad (6)$$

Once the overall masking curve was computed, all the sinusoids whose amplitudes were below the level of the masking curve were discarded. Figure 4 shows the result from this procedure for a given windowed segment. In this case, a total of 36 sinusoidal components were kept (open circles).

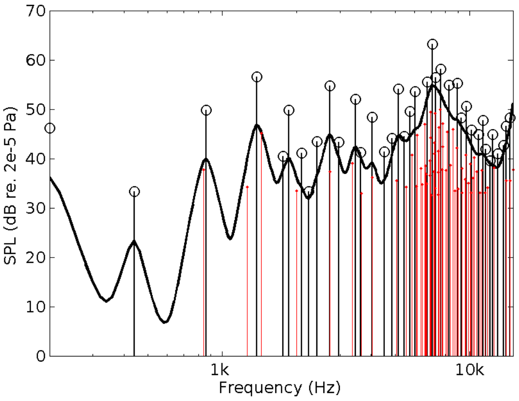

Figure 4. *Amplitude of 100 sinusoids and its associated masking curve (thick curve). Open circles represent those sinusoids whose amplitudes are above the masking level. Red lines represent those sinusoids whose amplitudes are below the masking level.*

An example of the synthesized signals as well as the respective error signals obtained from a drilling-sound excerpt is shown in Figure 5. Observe that the error of the low order model is larger than that of the model based on 100 components. In spite of this increase in error informal verifications have shown that the synthesized signal is not easily distinguished from the original one. We believe that the error is not perceptually relevant since psychoacoustics principles have been used as model reduction methods. This may attain additional relevance considering that the major goal when synthesizing the drilling sound is not physical fidelity (i.e. the synthesized sound is physically equivalent to original sound) but functional fidelity. That is, acoustic information deemed significant for the correct execution of the drilling procedure should be rendered properly.

To observe whether the number of reduced sinusoids depended on the surgeon technique, four drilling excerpts of 8192-sample length were selected from the audio recordings of each surgery. The number of perceptually relevant sinusoids was collected for each windowed segment and pooled across the four excerpts. Table 1 summarizes the percentages of removed sinusoids. Between 68% and 70% of the components were removed. That is, the use of approximately 30 sinusoids appears to be sufficient for a perceptually motivated synthesis of the drilling sound in maxillo-facial surgery. Furthermore, the small differences

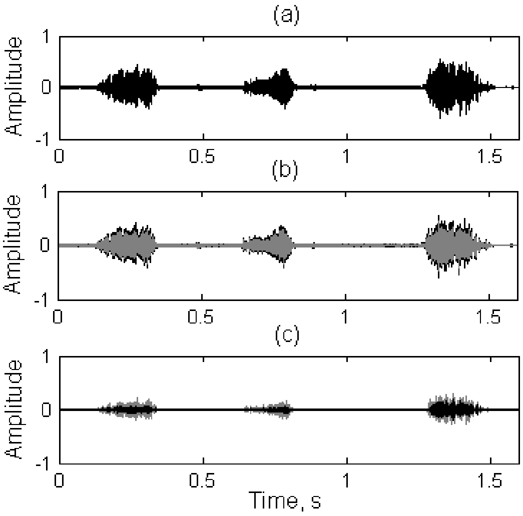

Figure 5. *(a) Original signal, (b) Synthesized signal based on 100 sinusoidal components (dark) and signal with reduced model order (gray), (c) respective error signals*.

Surgeon	**Reduced components (%)**
S1	70
S2	69
S3	68

Table 1. Percentage of removed sinusoidal components for drilling audio signals captured from recordings of different surgeons.

between the expert surgeons may suggest that an auditory display designed to convey expert performance on drilling, can be based on a general approximation. However, formal listening tests are necessary to validate these hypotheses.

4. FUTURE WORK

Studying the harmonic relation of the sinusoids may reveal additional information that can serve to further reduce the order of the drilling sound model. Further work will also be conducted to enable the model to control mimicking of tissue relations, to respond according to changes in force and in bone compliance, and also to account for surgeon distance and orientation. Proper integration of visual, haptic and auditory information is expected to provide an effective multimodal display for the enhancement of drilling skills in maxillo-facial surgery, and probably for other surgical procedures.

5. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the SKILLS Integrated Project (IST-FP6 #035005, http://www.skills-ip.eu) funded by the European Commission. Mads G. Christensen is greatly acknowledged for fruitful discussions on aspects of sound analysis. CEA-LIST staff is greatly acknowledged, in particular C. Mégard, F. Ferlay, F. Periquet, S. Bouchigny and J. Le Garrec. We thank the surgeons for their cooperation and availability: Dr. F. Taha from CHU Amiens, Dr. P. Delcampe from CHU Rouen, Dr. C. D'Hauthuille from CHU Nantes. Thanks to Mr. B. Beloncle for his contribution.

6. REFERENCES

[1] R. M. Satava and S. B. Jones, "Virtual environments for medical training and education," *Presence*, vol. 6, no. 2, pp. 139–146, 1997.

[2] A. Liu, F. Tendick, K. Cleary, and C. Kaufmann, "A survey of surgical simulation: Applications, technology, and education," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 6, pp. 599–614, 2003.

[3] G. J. Wiet, D. Stredney, D. Sessanna, J. A. Bryan, D. B. Welling, and P. Schmalbrock, "Virtual temporal bone dissection: An interactive surgical simulator," *Otolaryngol Head Neck Surg*, vol. 127, no. 1, pp. 79–83, 2002.

[4] D. Morris, C. Sewell, N. Blevis, F. Barbagli, and K. Salisbury, "A collaborative virtual environment for the simulation of temporal bone surgery," in *Proceedings of MICCAI (Medical Image Computing and Computer-Aided Intervention) VII*, vol. 3217 of *Springer-Verlag Lecture Notes in Computer Science*, (Rennes, France), pp. 319–327, Sept. 26–30 2004.

[5] D. Morris, C. Sewell, F. Barbagli, and K. Salisbury, "Visuohaptic simulation of bone surgery for training and evaluation," *Virtual and augmented reality supported simulators*, pp. 48–57, 2006.

[6] J. L. Mercade, J. Connell, U. Y. Noma, P. G. O'Sullivan, and N. P. Shine, "Application of the 3G AMR speech compression algorithm to drill signature analysis in temporal bone surgery," *Irish Signals and Systems Conference*, pp. 209–314, 2006.

[7] N. P. Shine, P. G. O'Sullivan, J. Connell, P. Rulikowski, and J. Barrett, "Digital spectral analysis of the drill-bone acoustic interface during temporal bone dissection: a qualitative cadaveric pilot study," *Otol Neurotol*, vol. 27, pp. 728–733, August 2006.

[8] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, pp. 744–754, August 1986.

[9] M. M. Goodwin, "The STFT, sinusoidal models, and speech modifications," in *Springer Handbook of Speech Processing* (J. Benesty, M. M. Sondhi, and Y. Huang, eds.), ch. 12, pp. 229–256, Springer Berlin Heidelberg, 2007.

[10] T. S. Verma and T. H. Y. Meng, "Sinusoidal modeling using frame-based perceptually weighted matching pursuits," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 9, pp. 981–984, 1999.

[11] R. Heusdens, R. Vafin, and W. B. Kleijn, "Sinusoidal modeling using psychoacoustic-adaptive matching pursuits," *IEEE Signal Processing Lett.*, vol. 9, pp. 262–265, August 2002.

[12] M. Lagrange and S. Marchand, "Real-time additive synthesis of sound by taking advantage of psychoacoustics," in *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, (Limerick, Ireland), December 6–8 2001.

[13] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103–138, 1990.

[14] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*. Norwell, Massachusetts, USA: Kluwer Academic Publishers, 2nd ed., 2003.

[15] E. Terhardt, "Calculating virtual pitch," *Hearing Research*, vol. 1, pp. 155–182, March 1979.