

身体姿勢を用いた直感的サウンドインタラクションの検討

下道 雄太^{1,a)} 入江 英嗣¹ 坂井 修一¹

概要: 本研究では直感的な身体の動きから自然な音の連なりを作る手法を検討する。すなわち、自由かつ自然に身体を動かした結果、それがそのまま心地の良い音の連なりとして反映されるような体験を創出することを目標とする。姿勢の検知には骨格推定ライブラリ openpose を用いた。取得した関節座標から、自然な音の連結になるよう MIDI ノートナンバーを計算し、OpenSoundControl プロトコルによって PureData へ送信して和音を鳴らす。この際、どのような身体の動きからどのような音の変化が生まれると心地良いかを探るために、身体の動きと音の変化を対応させた 3 通りの異なるシステムを実装し、実際に被験者に体験してもらうことでフィードバックを得た。

A Study of Intuitive Sound Interaction Using the Body Posture

SHIMOMICHI YUTA^{1,a)} IRIE HIDETSUGU¹ SAKAI SHUICHI¹

Abstract: In this study, we tried to explore a way of generating natural musical sequences from intuitive body movement. Our goal is to create an experience in which free and natural movement is transformed into a sequence of pleasant sounds directly. We used openpose, a skeleton estimation library, to detect the posture. The MIDI note number is calculated from the joint coordinates obtained from openpose so that it may be natural musical sequences, and transmitted to PureData by the OpenSoundControl protocol to sound a chord. In order to reveal what kind of correspondence of body movement to the sound makes us feel pleasant, we implemented three different systems and obtained feedback from the subjects.

1. はじめに

人間からのアクションに応じて音によるリアクションがあるシステムのことをサウンドインタラクションと呼ぶ。サウンドインタラクションの中でも最も代表的なものは楽器であるが、一般に楽器の演奏には習熟が必要であるほか、多くの場合高価であり、メンテナンスにもコストがかかるなど初心者にとってそのハードルは決して低くはない。

一方で、物理的に楽器を演奏するのではなく、身体の動きから直接音を生み出そうという試みがこれまでに数多く模索されてきた。特に、近年では 2010 年に Microsoft から発売された Kinect をはじめ、人間の姿勢や動作の情報をリアルタイムに取得する技術が広く一般に普及したことから、人間の身体の動きを用いた様々な作曲システムやバーチャルインストゥルメント (VMIs: Virtual Musical Instruments)

が提案されている。これらのシステムでは関節の座標や速度、加速度、手の形などパフォーマンスの姿勢や動作に応じてリアルタイムに音を変化させることで音楽を作っていくことが出来る。

このように体の一部または全体をコントローラとして特定の曲を演奏しようとした場合、作りたい音に応じて取るべき動作はおおよそ一意に決まり、パフォーマンスは楽譜通りの正しい演奏をすることが求められる。また、動作と音を機械的にマッピングした結果、本来の人間の身体の動きに馴染まない不自然な動きを要求されることもある。新しい楽器としてコンピュータを使った音楽の可能性を模索する場合、これらは必ずしも欠点というわけではないが、演奏するためにはそのシステムに習熟する必要がある。つまり、初心者への高いハードルという問題を多くの VMIs もまた抱えている。

そこで本研究では、身体の動きとより直感的に結びつい

¹ 東京大学大学院 情報理工学系研究科

^{a)} shimomichi@mtl.t.u-tokyo.ac.jp

た音を生成するサウンドインタラクションの可能性を模索する。ここでは楽器のように特定の曲の演奏を目指すのではなく、自由かつ自然に身体を動かした結果、それがそのまま心地の良い音の連なりとして反映されるような体験を創出することを目標とする。

2. 関連研究

身体の動きから音を作る際のセンシングの仕方として、Todd Winkler[1] は空間センサと身体センサという2つのアプローチを挙げている。空間センサは身体の部位や関節の空間的位置を用いるもので、例としてはテルミンや John Cage による《Variations V》, M.Mathews らの《RADIO DRUM》[5] などがある。一方身体センサは各部位の相対的な位置関係や角度、速度、加速度なども用いるもので、例としては YAMAHA の MIBURI や J.Pradiso らのダンシングシューズ [6], 橋本らのデータグローブ [7], [8] などが挙げられる。

次に、Microsoft の Kinect などによって関節座標を取得して実現された VMI には以下のようなものがある。T.Berg ら [9] は全身 15 個の関節それぞれに FM 音源を割り当て、各関節の x, y, z 座標を用いて各音源の音高、音色、音量を操作した。The Synapse for Kinect プロジェクト [10] では Ableton Live など OSC(Open Sound Control) メッセージを受け取ることの出来るアプリケーションと連携することで、様々なパラメータを柔軟かつリアルタイムに制御することが出来る。M.Yoo ら [11] らの提案では全身計 24 個の関節を 5 つの部位に分け、それぞれに音源を割り当てた。Gillian と Paradiso の提案した《Digito》[12] は手の位置と指のジェスチャーによって操作する。A.Hollenbach ら [13] の提案では部屋の中の移動と手の位置、指のジェスチャーを組み合わせて音を操作する。J.Fernandez らの提案した GeKiPe(Geste, Kinect, Percussions)[16] では、Kinect による関節座標と手に装着したデータグローブから取得した加速度や角度などのデータを組み合わせて、曲の進行に合わせて動的に音のパラメータに割り当てている。

3. 提案

3.1 直感的な音の生成

ここまで見てきた提案の多くでは、思い通りに音を奏でて音楽にするためにはそのシステムに習熟する必要があるか、はじめから訓練されたダンサーが踊ることを前提としていた。また、身体の動きと音のパラメータを機械的に割り当てた結果、直感的な動きから乖離した窮屈な動作が要求されたり、実際に演奏できる音楽の可能性が制限されてしまうこともある。

そこで、本研究では身体の動きとより直感的に結びついたサウンドインタラクションの可能性を模索する。すなわち、楽譜通りの演奏を意識することなく、直感的な身体

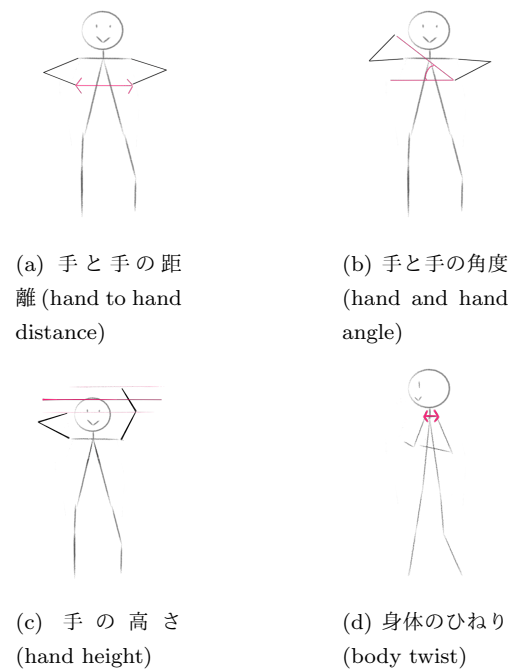


図 1 身体姿勢
Fig. 1 Body Posture

動きから自然で音楽的な音の連なりを生み出せるようなシステムの実現である。ここでは、楽器のように楽譜通りに特定の曲を演奏することは考えず、初心者でも簡単に実行できるような自由な身体の動きがそのまま音楽となるように、身体の動きと音を結びつける事を考える。

したがって、このサウンドインタラクションが目指すのは、人に聴かせるためのパフォーマンスを行うことではなく、より内省的に自分自身と向き合う音楽であると言える。自分自身の内側にある自覚的、無自覚的な音楽的欲求が、身体の動きとシステムを通じて実際の音として具現化される体験を通して、心の充実を得ることがこのサウンドインタラクションの最終的な目標となる。

3.2 身体の動きと音の変化の対応付け

初心者でも簡単に実行可能な身体の動きとして、手を上げる、身体を開く、縮める、ひねるといった動作が挙げられる。本研究ではまず、このような単純な身体の動きに対して、和音遷移、音高や音の数、音色の変化などの音楽的要素を割り当てることで、身体の動きに応じて音が変化することでどのように感じるか、どのような動きにどのような音の変化が対応すると心地よいと感じるのかを調べた。

ここでは、身体の動きと音の変化を対応付けた三通りのシステムを考案し、それぞれ A:Triad, B:Chords, C:Timbre として区別する。

3.2.1 A:Triad

2章で挙げた関連研究の多くでは、関節の座標など身体の1つのパラメータに対して単音を割り当てていた。それに対してこの A:Triad では、手と手の距離、手と手の角

度という身体のパラメータを用いて、3つの音からなる三和音を操作する。

すなわち、手と手の距離(図1(a))を和音の転回、手と手の角度(図1(b))を回転させることを隣り合う異なる和音への遷移に対応させる。直感的には、手と手を広げた時に和音が転回して音が高くなり、手を回転させることによって和音の構成音に変化する。手を離したり近づけたりする動作は1つの和音の中でアルペジオのような効果を生み、手を回転させることによって和音の種類を選択することが出来る。

3.2.2 B:Chords

B:Chordsのシステムでは、予め定められたコード進行に従って和音が自動で進行していく。今回の実装では、パッヘルベルのカノンによる $C \rightarrow G \rightarrow Am \rightarrow Em \rightarrow F \rightarrow C \rightarrow F \rightarrow G$ のコード進行を採用した。

身体パラメータは手と手の距離(図1(a))を鳴っている音の数、手の高さ(図1(c))を和音の転回に対応させる。ユーザーが操作する和音とは別にベース音が常時演奏されており、手と手の距離が最も短くなっている時、ベース音の上に音が1つだけ鳴っている状態となる。手を広げていくと現在の和音に対応した構成音が追加されていき、最大で6つの音による和音が演奏される。また、手の高さを上げていくと和音が転回して音が次第に高くなり、手を上下に揺らすことでアルペジオのような効果を得ることが出来る。

3.2.3 C:Timbre

C:Timbreのシステムでは身体の動きによって音色を操作する。

C:TimbreでもB:Chordsと同じく音は自動で進行していく。ここではA:Triad, B:Chordsと違って和音ではなく2本のノコギリ波を重ねた単音が演奏される。身体のひねり(肩と肩の距離:図1(d))によって2本のノコギリ波の音高がわずかにずれてモジュレーションがかかり、2つの手の高さの平均(図1(c))によってバンドパスフィルターの周波数が変化する。手を下に下げている時には高周波成分がカットされてほとんどsin波のような音色になり、手を上に上げていくにつれて次第にノコギリ波のバリバリとした音色が変わっていく。手の上下と身体のひねりを組み合わせることで、身体でシンセサイザーを操作しているような音の変化を楽しむことが出来る。

4. 実装

実装にあたっては、関節座標の推定にディープラーニングを用いた骨格推定ライブラリ openpose[2] を、音の生成に PureData[17] を用いた。また、C++によるプログラムと PureData との通信には OpenSoundControl(OSC) プロトコル [18], [19] を使用した。A:Triadにおける和音遷移の方法は、Audanika GmbH の開発した iOS 用 MIDI コントローラ SoundPrism[15] の音の配置を参考にした。



図2 システム構成
Fig. 2 System Configuration

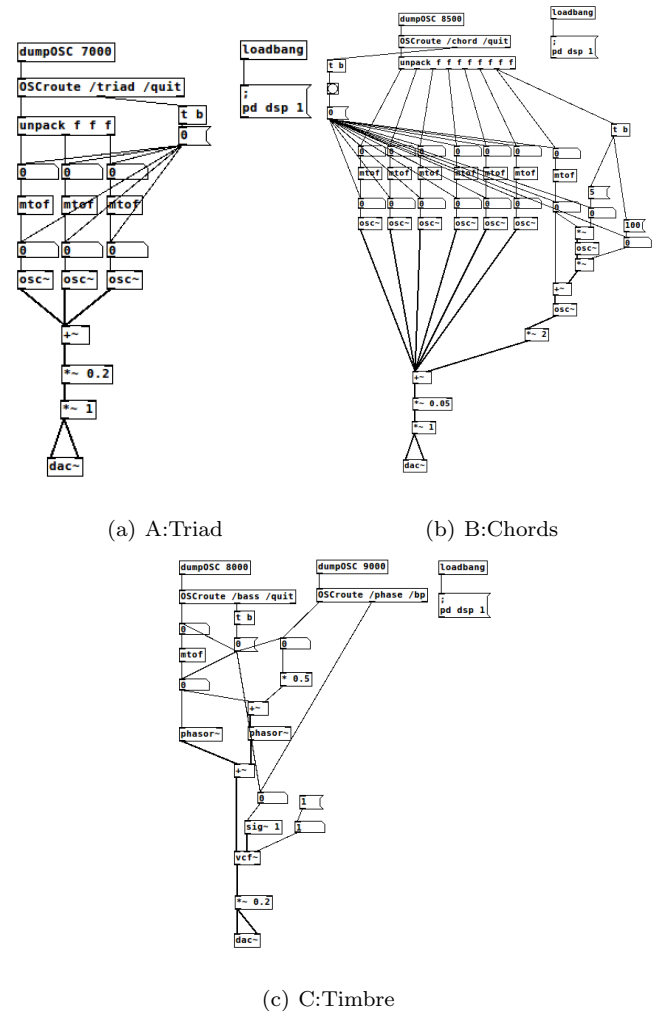


図3 PureData パッチ

Fig. 3 PureData Patch

システム全体の構成は図2のようになっている。まず、単眼のwebカメラからの映像を入力としてopenposeにより関節座標を推定する。次に、実現したいシステムに応じて関節座標からMIDIノートナンバーを計算し、OSC通信によってPureDataに送信する。PureDataではA:Triad, B:Chords, C:Timbreそれぞれについて、図3(a), 図3(b), 図3(c)に示したパッチを用いて音を生成した。

5. 評価

5.1 調査方法

評価にあたってはA, B, Cそれぞれのシステムを6人の被験者に実際に体験してもらい、アンケート調査とフィードバックの聞き取りを行った。アンケートの作成にあつ

ては、以下の2点を本研究の目標として設定し、質問文を作成した。

- 直感的に音に干渉し、操作できているという感覚を得ること
- 自然な身体の動きから、自然な音の連なりが生成されること

この目標がどの程度達成出来ているか、以下の5つの文章について、その同意の程度を1(強くそう思わない)から5(強く思う)の5段階で評価してもらった。

- (1) 身体の動きに合わせて音が変わることによって楽しく感じた。
- (2) 実際に鳴った音は心地よく、自然なものだと感じた。
- (3) 身体の動きと音の変化の対応は直感的な感覚に合っていた。
- (4) 鳴っている音に自ら干渉し、操作しているという感覚があった。
- (5) 音を鳴らすために要求された身体の動きは無理のない自然なものだった。

また質問6として、それぞれのシステムについてどの程度面白い、楽しく感じたか、その程度を直線上に線を引くことで評価してもらった。

5.2 調査結果

質問1から質問6までの回答結果をそれぞれ図4(a)から図4(f)に示す。

質問1から質問5の回答結果の横軸は、解答欄の1(強くそう思わない)から5(強く思う)に対応し、縦軸は被験者6人中の回答数である。

質問6の回答結果(図4(f))では、6人の各被験者(a, b, c, d, e, f)とする。a, c, fは音楽経験者、bはダンス経験者、d, eは音楽未経験者)ごとにA, B, Cそれぞれについて、引いてもらった直線の長さを示している。

またシステムに習熟し、満足するまで任意時間体験してもらったところ、平均の体験時間はA:Triadが87秒、B:Chordsが83秒、C:Timbreが84秒となった。

6. フィードバックと議論

6.1 ユーザの自由度

A:Triadのシステムは身体の動きに合わせて自由に和音が遷移し、特定のテンポやリズムは存在しない。それに対してB:ChordsやC:Timbreのシステムは一定テンポでコードが進行する中で和音や音色を変化させていた。

質問1, 質問6の結果からわかるようにA, B, Cのうち最も「楽しかった」という声が多かったのはB:Chordsであった。これは、自動でコードが進行することで、どのように動いてもある程度曲らしい音が自然と生成されるからである。またB:Chordsを高く評価した人はA:Triadの評価が低くなる傾向にあり、フィードバックとして「自由度

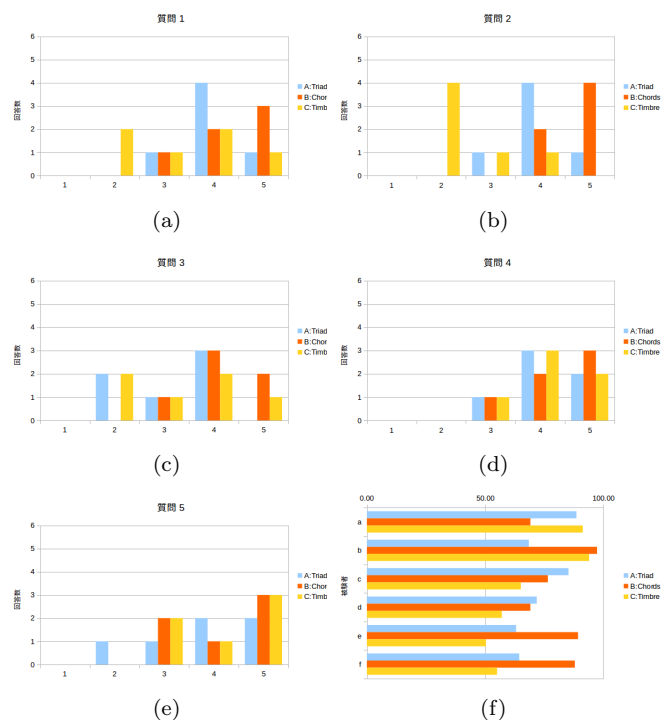


図4 アンケート回答結果
Fig. 4 Questionnaire results

が高すぎるとどうすればよいかわからない」「曲になっている方が楽しい」という声があった。一方で、音楽経験のある被験者からは「もっと複雑なことがしたい」「曲を演奏したい」といった声も聞かれた。

音楽に馴染みのないユーザと音楽経験者の要望を同時に満たすことは難しい。1分半弱という短い平均体験時間はこのシステムが初心者にとっても馴染みやすいものであると同時に飽きやすい、奥深くないという側面も示しており、音楽未経験者にとってもまだ自由度を増やす余地はあると考えられる。生成される音が自然と音楽になるように条件を拘束しながら、音楽経験者も満足出来るような自由度の中間点を探ることが今後の課題となる。

6.2 身体的要素

まず、身体の動きと音の変化の対応について、質問3と質問5の結果からA:Triadのシステムが特に身体の動きと音の変化の対応が自然では無かったという結果となった。これは手の距離と角度という機械的なパラメータをそのまま音の変化に割り当てたため、身体を動かすというよりも、コントローラを操作しているような感覚が勝ったのではないかと考えられる。

これに対してB:Chords, C:Timbreのシステムでは手を上げることと周波数が上がることで、という直感的な対応があった。このように心理的、物理的なイメージを介して音と身体の動きを結びつけることで、より直感にあったインタラクションの実現が期待出来る。

フィードバックの中には「身体の動きと感情、感情と音、というように間接的に身体と音を結びつけることも出来るのではないか」という意見もあった。また、これ以外のフィードバックとして「パーカッションのような動きに対応した音が鳴って欲しい」という声が複数聞かれたほか、「首の傾きや膝の曲げが使えるのではないか」という意見もあった。

6.3 音楽的要素

C:Timbre について「音色が変わるのは面白いが、もっと音色が綺麗だったら良かった」という声が多数聞かれた。これは単純に PureData の実装の問題であり、今後上質で多彩な音色やエフェクトを操作することが出来るようになるれば面白いサウンドインタラクションの実現が期待できる。

また A:Triad, B:Chords について「和音の構成音しか鳴らせないのはつまらない」という声があった。この場合、1つのコードの上で対応するスケールの音を鳴らせるようにするなどの案が考えられる。しかし、連続的な身体の動きからは隣り合った音にしか遷移出来ず、旋律を奏でようとすると音の発音と停止をコントロールして演奏する必要があると言う問題がある。これに対してはペンタトニックスケールのような特殊なスケールを使うなどの方法が考えられる。

この他には音楽的な要素として「リズム要素が欲しい」という声もあった。これも今後の実装の課題となる。

7. 今後の課題

フィードバックから得られた今後の実装の課題としてはまず、リズム要素の追加と音色の改善、打楽器のような身体の動きに対応したインタラクションの導入が挙げられる。また静的な姿勢の変化だけでなく、動いた方向によって異なる和音を鳴らしたり、複数人の姿勢を同時に組み合わせ合わせて合奏を行うなどの実装も考えられる。

身体の動きと音の変化をどう対応付けるかという問題については、物理的、心理的イメージや感情などと、身体の動き、音の変化をそれぞれ結びつけて間接的に身体と音に対応付けるというアプローチで、より直感的なインタラクションが実現出来る可能性がある。

和音の遷移関係については、ある程度自然に音楽になるように条件を拘束しながら、その中でユーザが様々な表現を行えるような自由度を確保することが課題となる。その場合、既存の曲の演奏に干渉するのか、その場で自動生成される音楽がユーザの動きによってインタラクティブに変化していくのか、その方向性はいくつか考えられる。

今後のさらなる展開としてはまず、アンビエントミュージックやテクノなどの音楽ジャンルへの適用が挙げられる。また、実社会への応用として、高齢者等のリハビリにおける身体を動かすモチベーション付けとしてのシステム

や、指などが不自由な障害者でも用意に音楽を楽しむことが出来るインタラクション、フロアの盛り上がりに合わせて動的に曲が変化するクラブミュージック、正しい身体の動きに対して正しく曲が演奏される運動支援システムなどが考えられる。

参考文献

- [1] T. Winkler, "Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music" In *Proceedings of the 1995 International Computer Music Conference*, 1995.
- [2] Z. Cao, T. Simon, S. Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," In *CVPR*, 2017.
- [3] 橋本周司, "ジェスチャーによる音楽制御," 日本ロボット学会誌 Vol14 NO.2 pp204 207, 1996.
- [4] 竹川佳成. "インタラクティブパフォーマンス・新世代楽器," 電子情報通信学会『知識の森』(<http://www.ieice-hbkb.org/>) 2群-9編-2章, 2011.
- [5] B. Boie, M. Mathews and A. Schloss, "The Radio Drum as a Synthesizer Controller," In *Proceeding of International Computer Music Conference (ICMC1989)*, pp.42-45, 1989.
- [6] J. Paradiso, K. Hsiao and E. Hu, "Interactive Music for Instrumented Dancing Shoes," In *Proceeding of International Computer Music Conference (ICMC1999)*, pp.453-456, 1999.
- [7] A. Sato, S. Hashimoto and S. Ohteru, "Singing and Playing in Musical Virtual Space," In *Proceeding of International Computer Music Conference (ICMC)*, pp. 289-292, 1991.
- [8] P. Hartono and S. Hashimoto, "Adaptive Timbre Control Using Gesture," In *Proceeding of ICMC*, pp. 151-158, 1994.
- [9] T. Berg, D. Chattopadhyay, M. Schedel and T.Vallier, "Interactive Music: Human Motion Initiated Music Generation Using Skeltal Tracking By Kinect", In *Proceeding of Conf.Soc.Electro-Acoustic Music United States*, 2012.
- [10] Synapse for kinect
<http://synapsekinect.tumblr.com/post/6307790318/synapse-for-kinect>
- [11] M. Yoo, J. Beak, I. Lee, "Creating Musical Expression using Kinect," In *Proceeding of the International Conference on New Interface for Musical Expression(NIME)*, 2011.
- [12] N.Gillian, J.A.Paradiso, "Digit:A Fine-Grain Gestually Controlled Virtual Musical Instrument", In *NIME*, 2012.
- [13] A.Hollenbach, M.Cox, J.Giegel, "A Dynamic 3D Performance Space for Control of Virtual Musical Instruments", In *Proceeding of ACM International Conference on Interactive Surfaces and Spaces*, 2016.
- [14] 黒田修平, 放地 宏佳, 吉見 真聡, 吉永 努, 入江 英嗣, "相対座標を用いた運動指導システム," マルチメディア、分散協調とモバイルシンポジウム 2013 論文集, pp.1475-1482.
- [15] Audanika GmbH, "SoundPrism Pro Advanced MIDI Controller for iOS"
<http://www.soundprism.com/>
- [16] J. Frenandez, G. Lorieux, T. Koppel, A. Vert, N. Verstraete, P. Spiesser, "GeKiPe, a gesture-based interface for audiovisual performance", In *Proceeding of NIME*, 2017.

- [17] PureData
<https://puredata.info/>
- [18] M. Wright, A. Freed. “Open Sound Control: A New Protocol for Communicating with Sound Synthesizers”, In *Proceeding of International Computer Music Conference*, 1997.
- [19] oscpack
<http://www.rossbencina.com/code/oscpack>