



**HAL**  
open science

# Towards Custom Dilated Convolutions on Pitch Spaces

Rony Abecidan, Mathieu Giraud, Gianluca Micchi

► **To cite this version:**

Rony Abecidan, Mathieu Giraud, Gianluca Micchi. Towards Custom Dilated Convolutions on Pitch Spaces. International Society for Music Information Retrieval Conference (ISMIR 2020), Late-Breaking Demo Session, 2020, Montréal, Canada. Extended Abstracts for the Late-Breaking Demo Session of the 21st Int. Society for Music Information Retrieval Conference. hal-02959676

**HAL Id: hal-02959676**

**<https://hal.science/hal-02959676>**

Submitted on 2 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Towards Custom Dilated Convolutions on Pitch Spaces

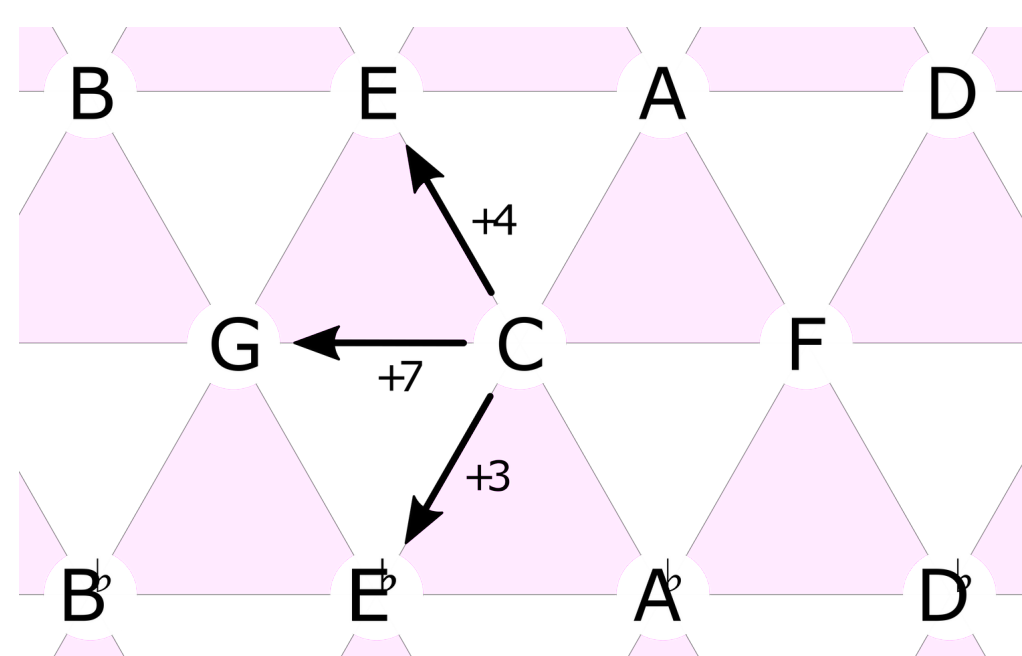
R. Abecidan

M. Giraud

G. Micchi

We benchmark several convolution kernels, in particular, **custom dilated convolutions**.

We test whether convolutions inspired by known pitch spaces like the **Tonnetz** may help to achieve better results on the task of **key detection**.



The **Tonnetz** [2] generalizes several pitch spaces such as the circle of fifths.

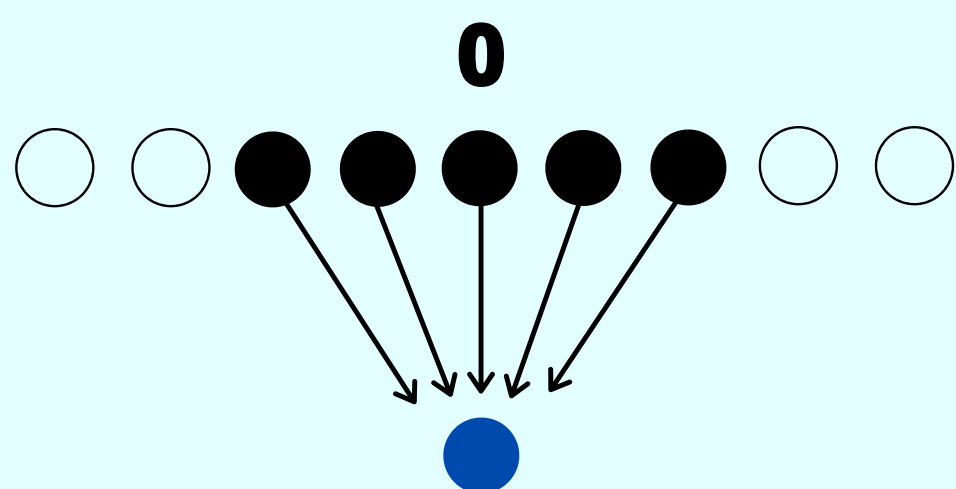
**201 scores** (baroque, classical, romantic) with key and roman numeral labels [5]

Frames : **200**   
 Training set : **40%**  
 (×12 transpositions)  
 Validation set : **30%**  
 Test set : **30%**

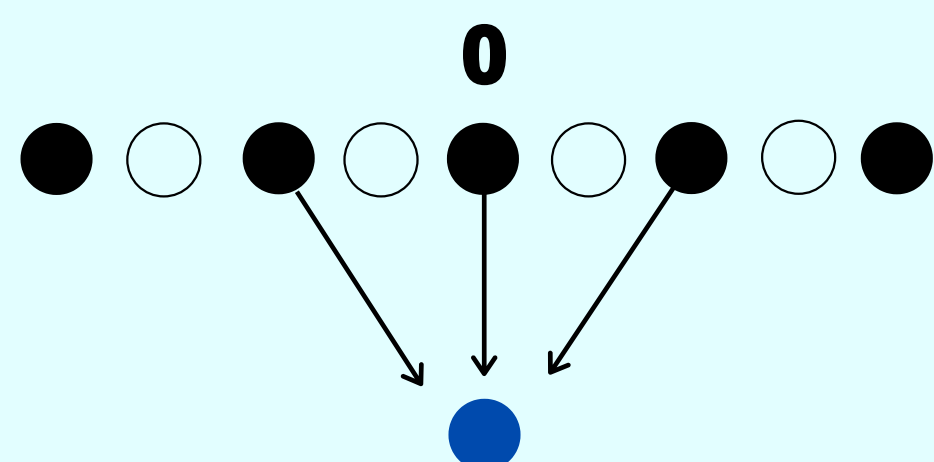
**Network**  
 Convolutions ×3  
 GRU  
 FC

*music21* [4], *pytorch*

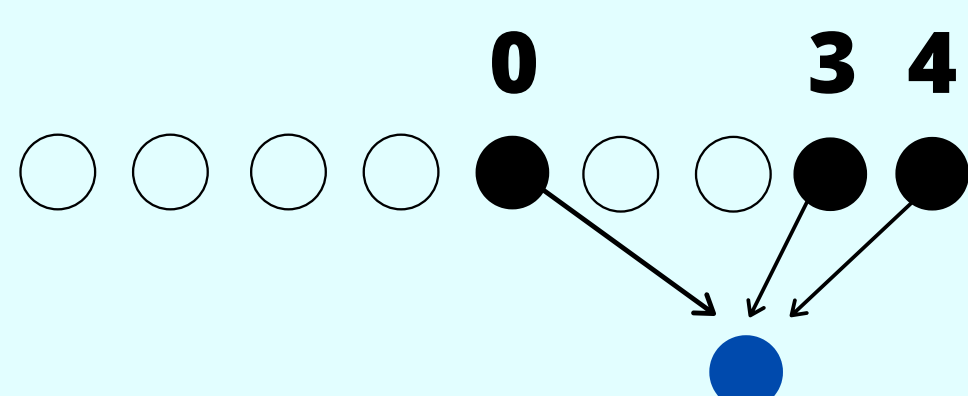
## Usual convolution



## Dilated convolution



## Custom {0, 3, 4} convolution



Some custom dilated convolutions yield better accuracies in key detection. This should be further studied to see how pitch spaces may influence the learning of such music tasks.

*without transposition of the training set*

Kernel	Accuracy	Kernel	Accuracy
$\{0, 6\}^{12}$	45.3%	$\{0, 4, 8\}^8$	42.3%
$\{0, 8\}^{12}$	46.4%	$\{0, 3, 10\}^8$	43.9%
...		...	
$\{0, 10\}^{12}$	50.7%	$\{0, 8, 11\}^8$	53.1%
$\{0, 7\}^{12}$	51.0%	$\{0, 5, 7\}^8$	54.0%

*with transposition of the training set*

Kernel	Accuracy	Kernel	Accuracy
$\{0, 4\}^{12}$	68.1%	$\{0, 2, 7\}^8$	66.4%
$\{0, 8\}^{12}$	68.3%	$\{0, 2, 4\}^8$	67.0%
...		...	
$\{0, 10\}^{12}$	70.4%	$\{0, 3, 10\}^8$	71.3%
$\{0, 3\}^{12}$	71.0%	$\{0, 5, 8\}^8$	71.8%

[1] Briot, Pachet, Deep Learning Techniques for Music Generation, 2019.  
 [2] Tymoczko, A Geometry of Music, 2011.  
 [3] Lattner, Grachten, Widmer, ISMIR 2018  
 [4] Cuthbert, Ariza, ISMIR 2010  
 [5] Micchi, Gotham, Giraud, TISMIR, 2020.