FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Title of the Dissertation

Márcio Duarte

WORKING VERSION

Programa Doutoral em Engenharia Informática

Supervisor: Name of the Supervisor

July 31, 2008

Abstract

Here goes the abstract written in English. In Portuguese

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed vehicula lorem commodo dui. Fusce mollis feugiat elit. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec eu quam. Aenean consectetuer odio quis nisi. Fusce molestie metus sed neque. Praesent nulla. Donec quis urna. Pellentesque hendrerit vulputate nunc. Donec id eros et leo ullamcorper placerat. Curabitur aliquam tellus et diam.

Ut tortor. Morbi eget elit. Maecenas nec risus. Sed ultricies. Sed scelerisque libero faucibus sem. Nullam molestie leo quis tellus. Donec ipsum. Nulla lobortis purus pharetra turpis. Nulla laoreet, arcu nec hendrerit vulputate, tortor elit eleifend turpis, et aliquam leo metus in dolor. Praesent sed nulla. Mauris ac augue. Cras ac orci. Etiam sed urna eget nulla sodales venenatis. Donec faucibus ante eget dui. Nam magna. Suspendisse sollicitudin est et mi.

Fusce sed ipsum vel velit imperdiet dictum. Sed nisi purus, dapibus ut, iaculis ac, placerat id, purus. Integer aliquet elementum libero. Phasellus facilisis leo eget elit. Nullam nisi magna, ornare at, aliquet et, porta id, odio. Sed volutpat tellus consectetuer ligula. Phasellus turpis augue, malesuada et, placerat fringilla, ornare nec, eros. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Vivamus ornare quam nec sem mattis vulputate. Nullam porta, diam nec porta mollis, orci leo condimentum sapien, quis venenatis mi dolor a metus. Nullam mollis. Aenean metus massa, pellentesque sit amet, sagittis eget, tincidunt in, arcu. Vestibulum porta laoreet tortor. Nullam mollis elit nec justo. In nulla ligula, pellentesque sit amet, consequat sed, faucibus id, velit. Fusce purus. Quisque sagittis urna at quam. Ut eu lacus. Maecenas tortor nibh, ultricies nec, vestibulum varius, egestas id, sapien.

Donec hendrerit. Vivamus suscipit egestas nibh. In ornare leo ut massa. Donec nisi nisl, dignissim quis, faucibus a, bibendum ac, diam. Nam adipiscing hendrerit mi. Morbi ac nulla. Nullam id est ac nisi consectetuer commodo. Pellentesque aliquam massa sit amet tellus. Vivamus sodales aliquam leo.

Agradecimentos

Aliquam id dui. Nulla facilisi. Nullam ligula nunc, viverra a, iaculis at, faucibus quis, sapien. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Curabitur magna ligula, ornare luctus, aliquam non, aliquet at, tortor. Donec iaculis nulla sed eros. Sed felis. Nam lobortis libero. Pellentesque odio. Suspendisse potenti. Morbi imperdiet rhoncus magna. Morbi vestibulum interdum turpis. Pellentesque varius. Morbi nulla urna, euismod in, molestie ac, placerat in, orci.

Ut convallis. Suspendisse luctus pharetra sem. Sed sit amet mi in diam luctus suscipit. Nulla facilisi. Integer commodo, turpis et semper auctor, nisl ligula vestibulum erat, sed tempor lacus nibh at turpis. Quisque vestibulum pulvinar justo. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam sed tellus vel tortor hendrerit pulvinar. Phasellus eleifend, augue at mattis tincidunt, lorem lorem sodales arcu, id volutpat risus est id neque. Phasellus egestas ante. Nam porttitor justo sit amet urna. Suspendisse ligula nunc, mollis ac, elementum non, venenatis ut, mauris. Mauris augue risus, tempus scelerisque, rutrum quis, hendrerit at, nunc. Nulla posuere porta orci. Nulla dui.

Fusce gravida placerat sem. Aenean ipsum diam, pharetra vitae, ornare et, semper sit amet, nibh. Nam id tellus. Etiam ultrices. Praesent gravida. Aliquam nec sapien. Morbi sagittis vulputate dolor. Donec sapien lorem, laoreet egestas, pellentesque euismod, porta at, sapien. Integer vitae lacus id dui convallis blandit. Mauris non sem. Integer in velit eget lorem scelerisque vehicula. Etiam tincidunt turpis ac nunc. Pellentesque a justo. Mauris faucibus quam id eros. Cras pharetra. Fusce rutrum vulputate lorem. Cras pretium magna in nisl. Integer ornare dui non pede.

O Nome do Autor

"You should be glad that bridge fell down.
I was planning to build thirteen more to that same design"

Isambard Kingdom Brunel

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

ADT	Abstract Data Type
ANDF	Architecture-Neutral Distribution Format
API	Application Programming Interface
CAD	Computer-Aided Design
CASE	Computer-Aided Software Engineering
CORBA	Common Object Request Broker Architecture
UNCOL	UNiversal COmpiler-oriented Language
Loren	Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Sed vehicula lorem commodo dui
WWW	*World Wide Web*

Chapter 1

Introduction

1.1 Context

The current developments of tools are advancing such existing models, further lead to modeling new sounds, providing possibilities to generate audio efficiently and faster but at the same time interesting results from the trained data set of samples in audio domain. [2]
Deep learning with audio shifts the focus to next level of real-time synthesis of sound by creating completely new-sounding sounds. [2]

1.2 Motivation

Today, digital technologies and advanced computational features, e.g. deep learning and artificial intelligence (AI) tools are shaping our relationships with music as well as enabling new possibilities of utilising new musical instruments and interfaces. [2]

1.3 Objectives

1.4 Dissertation Structure

Para além da introdução, esta dissertação contém mais x capítulos. No Capítulo 2, é descrito o estado da arte e são apresentados trabalhos relacionados. No Capítulo **??**, ipsum dolor sit amet, consectetuer adipiscing elit. No Capítulo **??** praesent sit amet sem. No Capítulo 5 posuere, ante non tristique consectetuer, dui elit scelerisque augue, eu vehicula nibh nisi ac est.

Chapter 2

State of the Art

2.1 Background

2.2 Related Work

2.3 Existing Technologies

2.3.1 WaveNet

[1]
A single WaveNet can capture the characteristics of many different speakers with equal fidelity, and can switch between them by conditioning on the speaker identity. When trained to model music, we find that it generates novel and often highly realistic musical fragments.
The dataset consisted of 44 hours of data from 109 different speakers.
Because the model is not conditioned on text, it generates non-existent but human language-like words in a smooth way with realistic sounding intonations. This is similar to generative models of language or images, where samples look realistic at first glance, but are clearly unnatural upon closer inspection. The lack of long range coherence is partly due to the limited size of the model's receptive field (about 300 milliseconds), which means it can only remember the last 2-3 phonemes it produced.
Finally, we observed that the model also picked up on other characteristics in the audio apart from the voice itself. For instance, it also mimicked the acoustics and recording quality, as well as the breathing and mouth movements of the speakers.
Even with a receptive field of several seconds, the models did not enforce long-range consistency which resulted in second-to-second variations in genre, instrumentation, volume and sound quality. Nevertheless, the samples were often harmonic and aesthetically pleasing, even when produced by unconditional models.
When applied to TTS, WaveNets produced samples that outperform the current best TTS systems in subjective naturalness. Finally, WaveNets showed very promising results when applied to music audio modeling and speech recognition.

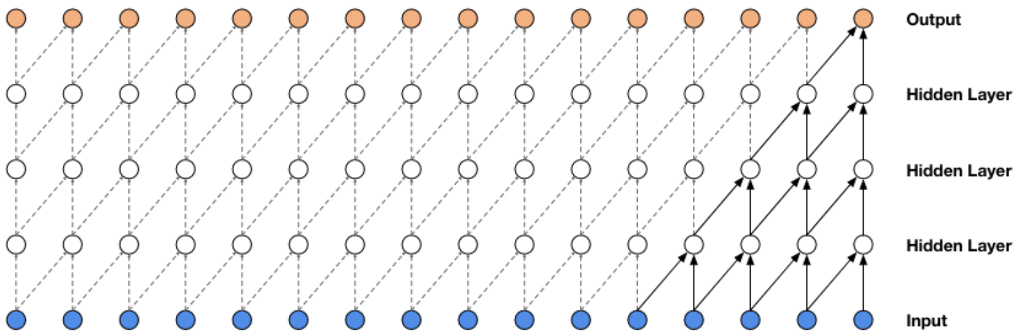
Previously, autoregressive models like WaveNet (used in NSynth) represented the state of the art in neural audio synthesis. These models are good at learning the characteristics of sounds over very short time periods (local latent structure) but struggle with longerterm features (global latent structure). They are also very slow, since they have to generate waveforms one sample at a time. In contrast, GANs are capable of modeling global latent structure, as well as synthesising more efficiently. [2]

2.4 Approaches to Related Problems

2.4.1 WaveNet

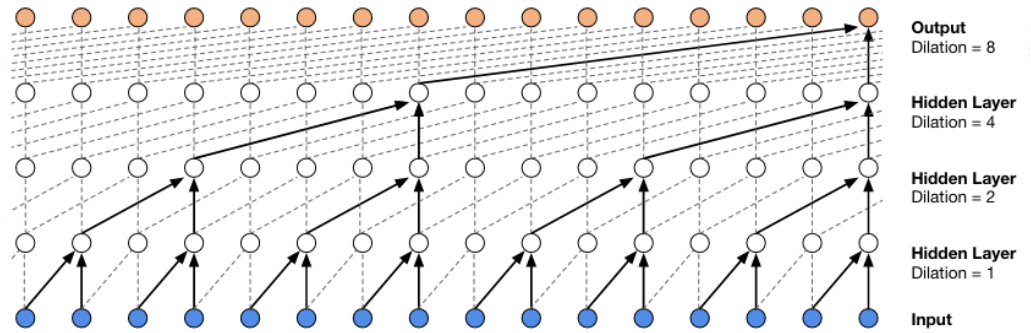
By using causal convolutions, we make sure the model cannot violate the ordering in which we model the data: the prediction p (xt+1 | x1, ..., xt) emitted by the model at timestep t cannot depend on any of the future timesteps xt+1, xt+2, . . . , xT as shown in Fig. 2.1. [1]

Figure 2.1: Causal convolution.



One of the problems of causal convolutions is that they require many layers, or large filters to increase the receptive field.
A dilated convolution (also called ‘ a trous, or convolution with holes) is a convolution where the filter is applied over an area larger than its length by skipping input values with a certain step. Fig. 2.2 shows a dilated convolution.
Stacked dilated convolutions enable networks to have very large receptive fields with just a few layers, while preserving the input resolution throughout the network as well as computational efficiency.
By conditioning the model on other input variables, we can guide WaveNet’s generation to produce audio with the required characteristics. For example, in a multi-speaker setting we can choose the speaker by feeding the speaker identity to the model as an extra input. Similarly, for TTS we need to feed information about the text as an extra input.
if a model is not conditioned on text, it will generate human-like sounds without any meaning behind it

Figure 2.2: Dilated convolution.



these models also capture extra sounds such as the breathing and background noises

Chapter 3

Problem

audio is 1-D data.[1]
raw audio is typically stored as a sequence of 16-bit integer values (one per timestep). [1]

Chapter 4

Solution

It is common to use a deconvolutional neural network (DNN) as generator and a convolutional neural network (CNN or ConvNet) as the discriminator. [2]

4.1 Tasks

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Integer hendrerit commodo ante. Pellentesque nibh libero, aliquam at, faucibus id, commodo a, velit. Duis eleifend sem eget leo. Morbi in est. Suspendisse magna sem, varius nec, hendrerit non, tincidunt quis, quam. Aenean congue. Vivamus vel est sit amet sem iaculis posuere. Cras mollis, enim vel gravida aliquam, libero nunc ullamcorper dui, ullamcorper sodales lectus nulla sed urna. Morbi aliquet porta risus. Proin vestibulum ligula a purus. Maecenas a nulla. Maecenas mattis est vitae neque auctor tempus. Etiam nulla dui, mattis vitae, porttitor sed, aliquet ut, enim. Cras nisl magna, aliquet et, laoreet at, gravida ac, neque. Sed id est. Nulla dapibus dolor quis ipsum rhoncus cursus.
Etiam nisi est, dignissim sodales, fermentum id, pulvinar ac, eros. Duis id orci. Nam pretium nisl ac augue. Ut adipiscing magna eget est. Curabitur varius. Nulla facilisi. Pellentesque sit amet neque ac dui accumsan blandit. Donec mauris felis, egestas sit amet, convallis ac, dignissim quis, dolor. Maecenas cursus tortor vel leo. Quisque tristique. Nunc augue odio, tincidunt in, dapibus sed, ultricies sit amet, lorem. In hac habitasse platea dictumst. Praesent iaculis, lacus hendrerit tempor sodales, libero tellus aliquet orci, ut rhoncus massa lectus quis erat. Pellentesque quis dolor nec tortor rhoncus convallis. Aliquam erat volutpat. Fusce placerat, magna eu imperdiet lobortis, augue massa blandit turpis, a consectetuer quam arcu sit amet risus. Suspendisse potenti. Praesent sapien metus, interdum vitae, fermentum id, faucibus ut, lorem. Nunc iaculis purus id tortor. Aenean risus pede, laoreet ac, tristique sed, lobortis in, turpis (see Figure 4.1b).
Vestibulum et lorem in ligula viverra pharetra. Curabitur quis purus in urna facilisis bibendum. Pellentesque at arcu accumsan velit bibendum ornare. Praesent massa. Quisque dolor. In libero. Vestibulum ac diam id leo feugiat blandit. Donec porta, tellus ac pellentesque molestie, felis mauris viverra lacus, sed dignissim purus justo eu justo. Proin iaculis, nunc eu volutpat volutpat, libero purus rutrum enim, id euismod lacus lorem nec augue. Donec hendrerit lacinia ante.

(a) UP at the left (b) UP at the right

Figure 4.1: Two Figures side by side

Integer mollis vulputate orci. In pellentesque, metus pharetra elementum pharetra, est purus bibendum turpis, eu pretium sapien libero convallis odio. Cras sodales bibendum risus. Sed mattis nulla non leo. Nulla nunc. Phasellus egestas sodales massa. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Etiam mi.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Quisque purus sapien, interdum ut, vestibulum a, accumsan ullamcorper, erat. Mauris a magna ut leo porta imperdiet. Donec dui odio, porta in, pretium non, semper quis, orci. Quisque erat diam, pharetra vel, laoreet ac, hendrerit vel, enim. Donec tristique luctus risus. Fusce dolor est, eleifend id, elementum sit amet, varius vitae, neque. Morbi at augue. Ut sem ligula, auctor vitae, facilisis id, pharetra non, lectus. Nulla lacus augue, aliquam eget, sollicitudin sed, hendrerit eu, leo. Suspendisse ac tortor. Mauris at odio. Etiam vehicula. Nam lacinia purus at nibh. Aliquam fringilla lorem ac justo. Ut nec enim. Nunc ornare, eros eu facilisis tristique, nisl lorem lacinia risus, non ullamcorper tellus urna et eros. Quisque eleifend tempus metus. Nunc ipsum.

Phasellus ullamcorper justo id risus. Nunc in leo. Mauris auctor lectus vitae est lacinia egestas. Nulla faucibus erat sit amet lectus varius semper. Praesent ultrices vehicula orci. Nam at metus. Aenean eget lorem nec purus feugiat molestie. Phasellus fringilla nulla ac risus. Aliquam elementum aliquam velit. Aenean nunc odio, lobortis id, dictum et, rutrum ac, ipsum. Aenean tellus magna, lacinia eget, bibendum ut, interdum sit amet, ipsum. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Mauris felis lacus, dapibus sit amet, pretium feugiat, aliquet non, purus. Aliquam elementum, diam quis porttitor gravida, sem sapien iaculis nulla, ut pharetra odio felis a metus. Nulla lacus ipsum, tristique ut, dapibus sed, mollis et, justo. Vivamus non ipsum sed ligula placerat ultrices. Maecenas dictum leo adipiscing mauris. Vestibulum tristique, lacus a consequat suscipit, nunc dui sollicitudin arcu, non interdum libero est eget tortor. Ut eget neque quis leo tempor dictum.

Quisque ullamcorper. Aliquam vel magna. Sed pulvinar dictum ligula. Sed ultrices dolor ut turpis. Vivamus sagittis orci malesuada arcu venenatis auctor. Proin vehicula pharetra urna. Aliquam egestas nunc quis nisl. Donec ullamcorper. Nulla purus. Ut suscipit lacus vitae dui. Mauris semper. Ut eget sem. Integer orci. Nam vitae dui eget nisi placerat convallis.

Sed id lorem. Proin gravida bibendum lacus. Sed molestie, urna quis euismod laoreet, diam dolor dictum diam, vitae consectetuer leo ipsum id ante. Integer eu lectus non mauris pharetra viverra. In feugiat libero ut massa. Morbi cursus, lorem sollicitudin blandit semper, felis magna pellentesque lacus, ut rhoncus leo neque at tellus. Sed mattis, diam eget eleifend tincidunt, ligula

eros tincidunt diam, vitae auctor turpis est vel nunc. In eu magna. Donec dolor metus, egestas sit amet, ultrices in, faucibus sed, lectus. Etiam est enim, vehicula pharetra, porta non, viverra vel, nunc. Ut non sem. Etiam nec neque. Sed rhoncus, justo id imperdiet pharetra, mi tellus accumsan neque, vitae volutpat tortor enim in odio. Nunc porta justo a lorem. Nulla hendrerit odio vitae dolor. Suspendisse eu nisl.

Proin vehicula pharetra urna. Aliquam egestas nunc quis nisl. Donec ullamcorper. Nulla purus. Ut suscipit lacus vitae dui. Mauris semper. Ut eget sem. Integer orci. Nam vitae dui eget nisi placerat convallis.

Chapter 5

Conclusions

Proin sed justo eu sapien eleifend elementum. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vivamus quam lacus, pharetra vel, aliquam vel, volutpat sed, nisl.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam non felis sed odio rutrum ultrices. Donec tempor dolor. Vivamus justo neque, tempus id, ullamcorper in, pharetra non, tellus. Praesent eu orci eu dolor congue gravida. Sed eu est. Donec pulvinar, lectus et eleifend volutpat, diam sapien sollicitudin arcu, a sagittis libero neque et dolor. Nam ligula. Cras tincidunt lectus quis nunc. Cras tincidunt congue turpis. Nulla pede velit, sagittis a, faucibus vitae, porttitor nec, ante. Nulla ut arcu. Cras eu augue at ipsum feugiat hendrerit. Proin sed justo eu sapien eleifend elementum. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vivamus quam lacus, pharetra vel, aliquam vel, volutpat sed, nisl.

Nullam erat est, vehicula id, tempor non, scelerisque at, tellus. Pellentesque tincidunt, ante vehicula bibendum adipiscing, lorem augue tempor felis, in dictum massa justo sed metus. Suspendisse placerat, mi eget molestie sodales, tortor ante interdum dui, ac sagittis est pede et lacus. Duis sapien. Nam ornare turpis et magna. Etiam adipiscing adipiscing ipsum. Fusce sodales nisl a arcu. Cras massa leo, vehicula facilisis, commodo a, molestie faucibus, metus. Suspendisse potenti. Duis sagittis. Donec porta. Sed urna. Maecenas eros. Vivamus erat ligula, pharetra sit amet, bibendum et, fermentum sed, dolor. Nullam eleifend condimentum nibh. Integer leo nibh, consequat eget, mollis et, sagittis ac, felis. Duis viverra pede in pede. Phasellus molestie placerat leo. Praesent at tellus a augue congue molestie. Proin sed justo eu sapien eleifend elementum. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aliquam tempor tristique risus. Suspendisse potenti. Fusce id eros. In eu enim. Praesent commodo leo. Nullam augue. Pellentesque tellus. Integer pulvinar purus a dui convallis consectetuer. In adipiscing, orci vitae lacinia semper, sapien elit posuere sem, ac euismod ipsum elit tempus urna. Aliquam erat volutpat. Nullam suscipit augue sed felis. Phasellus faucibus accumsan est.

Aliquam felis justo, facilisis sit amet, bibendum ut, tempus ac, dolor. Sed malesuada. Nunc non

massa. In erat. Nulla facilisi. Phasellus blandit, est in accumsan cursus, libero augue elementum leo, vitae auctor mauris nisl ac tortor. Cras porttitor ornare elit. Fusce at lorem. Sed lectus tortor, vestibulum id, varius a, condimentum nec, lectus. Maecenas in nisi et magna pretium aliquam. Pellentesque justo elit, feugiat nec, tincidunt a, dignissim vel, ipsum. Sed nunc. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam tempus rhoncus leo. Donec neque quam, cursus sit amet, ultricies varius, semper non, pede. Donec porttitor. Sed aliquet feugiat elit.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Phasellus tellus pede, auctor ut, tincidunt a, consectetuer in, felis. Mauris quis dolor et neque accumsan pellentesque. Donec dui magna, scelerisque mattis, sagittis nec, porta quis, nulla. Vivamus quis nisl. Etiam vitae nisl in diam vehicula viverra. Sed sollicitudin scelerisque est. Nunc dapibus. Sed urna. Nulla gravida. Praesent faucibus, risus ac lobortis dignissim, est tortor laoreet mauris, dictum pellentesque nunc orci tincidunt tellus. Nullam pulvinar, leo sed vestibulum euismod, ante ligula elementum pede, sit amet dapibus lacus tortor ac nisl. Morbi libero. Integer sed dolor ac lectus commodo iaculis. Donec ut odio.

References

[1] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio, September 2016. arXiv:1609.03499 [cs].

[2] Koray Tahiroğlu, Miranda Kastemaa, and Oskar Koli. Al-terity: Non-Rigid Musical Instrument with Artificial Intelligence Applied to Real-Time Audio Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Proceedings of the International Conference on New Interfaces for Musical Expression, pages 337–342. International Conference on New Interfaces for Musical Expression, July 2020.

Appendix A

Lorem Ipsum

Depois das conclusões e antes das referências bibliográficas, apresenta-se neste anexo numerado o texto usado para preencher a dissertação.

A.1 O que é o *Lorem Ipsum*?

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum [**?**].

A.2 De onde Vem o Lorem?

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet...", comes from a line in section 1.10.32.
The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from "de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham.

A.3 Porque se usa o Lorem?

It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using “Content here, content here”, making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for “lorem ipsum” will uncover many web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like).

A.4 Onde se Podem Encontrar Exemplos?

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don’t look even slightly believable. If you are going to use a passage of Lorem Ipsum, you need to be sure there isn’t anything embarrassing hidden in the middle of text. All the Lorem Ipsum generators on the Internet tend to repeat predefined chunks as necessary, making this the first true generator on the Internet. It uses a dictionary of over 200 Latin words, combined with a handful of model sentence structures, to generate Lorem Ipsum which looks reasonable. The generated Lorem Ipsum is therefore always free from repetition, injected humour, or non-characteristic words etc.