

Väinö-Waltteri Granat

**TITLE**

Tarkentava alaotsikko

Bachelors's Thesis  
Faculty of Inforamtion Technology and Communication Sciences  
Examiner: Prof. Joni Kämäräinen  
May 2020

## ABSTRACT

Väinö-Waltteri Granat: Title  
Bachelors's Thesis  
Tampere University  
Information technology  
May 2020

---

The abstract is a concise, self-containing one page description of the work: what was the problem, what was done, and what are the results. Do not include charts or tables in the abstract. Put the abstract in the primary language of your thesis first and then the translation when that is needed. International students do not need to include an abstract in Finnish.

Keywords: Deep learning, Tacotron, Speech synthesis, avainsana, avainsana

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# TIIVISTELMÄ

Väinö-Waltteri Granat: Kuvaava otsikko  
Kandidaatintyön  
Tampereen yliopisto  
Tietotekniikka  
Toukokuu 2020

---

Tiivistelmä on suppea, 1 sivun mittainen itsenäinen esitys työstä: mikä oli ongelma, mitä tehtiin ja mitä saatiin tulokseksi. Kuvia, kaavioita ja taulukoita ei käytetä tiivistelmässä. Laita työn pääkielillä kirjoitettu tiivistelmä ensin ja käännös sen jälkeen. Suomenkieliselle kandidaatintyölle pitää olla myös englanninkielinen nimi arkistointia varten.

Avainsanat: Syväoppiminen, Tacotron, Puhesynteesi, keyword, keyword

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

## PREFACE

Tämä dokumenttipohja on laadittu Tampereen yliopiston tekniikan alan opinnäytetöitä varten. Mallipohja perustuu aikaisemmalle Tampereen teknillisen yliopiston pohjalle, mutta se on päivitetty vuonna 2019 toimintansa aloittavaa Tampereen yliopistoa varten. Jo alkuun on hyvä todeta, että tätä pohjaa käyttämällä saa julkaisua varten tuotettua suoraan arkistointikelpoisen (PDF/A-1b) dokumentin.

Alkusanoissa esitetään opinnäytetyön tekemiseen liittyvät yleiset tiedot. Tapana on myös esittää kiitokset työn tekemiseen vaikuttaneille henkilöille ja yhteisöille. Alkusanat eivät kuulu arvioinnin piriin, mutta niissä ei silti ole sopivaa moittia tai kritisoida ketään. Alkusanojen pituus on enintään 1 sivu. Alkusanojen lopussa on päivämäärä, jonka jälkeen työhön ei ole enää tehty korjauksia.

Tampereella, 18th May 2020

Väinö-Waltteri Granat

## CONTENTS

1. Introduction . . . . .	1
2. Audio synthesis . . . . .	2
2.1 Deep neural network . . . . .	2
3. Viittaustekniikat . . . . .	3
4. Conclusions . . . . .	4
References . . . . .	5
Appendix A: Esimerkkilite . . . . .	6

## LIST OF SYMBOLS AND ABBREVIATIONS

DNN    Deep neural network

ML     Machine learning

$\mathbb{R}$      reaaliluvut

TTS    Text to speech

## 1. INTRODUCTION

Speech is the most important form of human communication. This is why it's not surprising that humans have been developing machines to produce artificial speech as far back as 18th century [1]. Such machines allow communication in situations where one of the participants is not able to produce audible speech for the rest. This might be because of physical distance, physiological inability or the speaker might not be a human in the first place. Modern artificial speech solutions are most often created as computer programs called text to speech applications.

This study is done in concurrency to a larger project of deep speaking avatar which seeks to create a human like digital avatar, that anyone can interact with by speaking to. Goal of this study is to produce a text to speech system for that avatar by using state of the art deep learning methods. The system implements a pretrained text to speech neural network called Tacotron2 [2].

The structure of this thesis is as follows.

## **2. AUDIO SYNTHESIS**

Audio synthesis is about synthesising audio.

### **2.1 Deep neural network**



### **3. VIITTAUSTEKNIIKAT**

## **4. CONCLUSIONS**

## REFERENCES

- [1] Schroeder, M. R. A brief history of synthetic speech. eng. *Speech Communication* 13.1 (1993), pp. 231–237. ISSN: 0167-6393.
- [2] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R., Saurous, R. A., Agiomyrgiannakis, Y. and Wu, Y. *Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions*. 2018. arXiv: 1712.05884 [cs.CL].

## **APPENDIX A: ESIMERKKILIITE**

Tämä teksti toimii esimerkkinä liitteiden muodostamiseen tässä dokumenttipohjassa. Vähän pidempi saa siitä kokonaisen kappaleen näköisen.