

Johannes Kvam

Application of Machine Learning in Tissue Characterization

Doctoral thesis
for the degree of PhD

Trondheim, December 2018

Norwegian University of Science and Technology
Medical Faculty
Department of Circulation and Medical Imaging

NTNU

Norwegian University of Science and Technology

Doctoral thesis
for the degree of PhD

Medical Faculty
Department of Circulation and Medical Imaging

© 2018 Johannes Kvam. All rights reserved

ISBN (printed version)
ISBN (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU,

Printed by NTNU-trykk

Bruk av maskinl ring i karakterisering av vev

Karakterisering av vev handler om   skille ulike vevstyper basert p  en eller annen egenskap som vevet innehar. Innen medisin har ofte sykt og friskt vev ulike fysiske egenskaper. Dermed vil man kunne b de diagnostisere og overv ke sykdommer dersom man kan m le disse egenskapene. Innen husdyrproduksjon til kjøtt, er den riktige sammensetningen av fett- og muskelvev viktig for smaksopplevelsen. Ved   skille fett- og muskelvev, samt   kunne m le dette nøyaktig, vil man kunne m te markedets forventninger til en god spiseopplevelse av kjøtt.

I begge disse eksemplene er det et sterkt behov for   kunne gj re slike m linger *in vivo* og ikke-invasivt. Innen medisin gj res invasive m linger typisk ved biopsi. Slike pr ver for rsaker ofte ubehag, smerte og i noen tilfeller, skade. I tillegg kan m lingene v re upresise, ettersom det kan v re vanskelig   ta pr ve p  riktig sted. Dette kan f re til feil diagnostisering som vil v re en stor p kjenning for pasienten. I kjøttindustrien er dyrevelferden viktig og dyrene skal ikke utsettes for un dvendig ubehag. Kjøttkvalitet kan dermed ikke kan m les *in vivo*, men etter at dyret er slaktet; *post mortem*. Konsekvensen er at den genetiske framgangen blir redusert, ettersom man m  beslutte om genene til et dyr skal f res videre basert p  m lt kjøttkvalitet hos s sknene.

Det er tre typiske avbildningsmetoder som brukes for ikke-invasiv bildedannelse: ultralyd (US), r ntgen computertomografi (CT) og magnetresonanstomografi (MR). Av disse tre er ultralyd den billigste og mest portable metoden. Ulempen ved ultralyd er at det er mange potensielle st ykilder som kan v re  deleggende for det resulterende bildet. Denne st yen representerer ofte ikke den underliggende anatomien og gj r at ultralyd, sammenlignet med de andre metodene, har den d rligste bildekvaliteten.

Gjennom de siste  rene har medisinsk bildeanalyse gjennomg tt en revolusjon. Den  kende populariteten til kunstig intelligens og maskinl ring har gjort at disse metodene brukes mer og mer i automatisk bildeanalyse. Dyp l ring, en gren av maskinl ring, har vist seg   v re spesielt suksessrik, da disse metodene kan gjenkjenne komplekse m nstre i bilder. Disse metodene lærer seg selv hvordan bilder skal tolkes ved   trene p  data der fasiten er kjent. For at disse metodene skal bli robuste, kreves dermed store mengder data. I tillegg m  det eksistere en sammenheng mellom dataene og fasiten som er observerbar. Som konsekvens blir ogs  kvaliteten p  dataen viktig.

I dette arbeidet unders kes potensialet for bruk av dyp l ring for karakterisering av vev, og ved hvilket punkt disse metodene begrenses av datagrunnlaget. Vi observerer at dyp l ring fungerer godt i CT, hvor bilde- og datakvaliteten er h y. Dette lar oss lage en robust metode for automatisk beskrivelse av CT volumdata ved bruk av dyp l ring. En tilsvarende metode blir ogs  fors kt p  ultralydbilder. Målet er   bestemme fettinnhold fra ultralydbilder og resultatene er konkurrerende

med gullstandardmetoder. Likevel observeres det i denne applikasjonen en øvre grense hvor disse metodene ikke lenger fungerer. For å unngå denne grensen er det et behov for andre ultralydavbildningsmetoder slik at datakvaliteten økes.

Konvensjonell ultralyd måler en spesifikk akustisk egenskap hos materialer. Ettersom denne har vist seg å være begrenset er det et behov for å undersøke muligheten for å måle andre egenskaper ved bruk av ultralyd. Materialets ulineære oppførsel viser seg å være en akustisk parameter med stor variasjon for bløte vev, og spesielt sensitiv til fettinnhold. Det vises at denne parameteren kan måles ved bruk av en ny ultralydmetode som baserer seg på å sende ut to lydbølger, med forskjellige frekvenser, samtidig. Dataene fra denne metoden, sammenlignet med konvensjonell ultralyd, har mer informasjon og potensielt høyere kvalitet. Dette indikerer at metoden kan forbedres ytterligere ved bruk av dyp læring. Metoden vil dermed kunne bidra til å oppnå robust vevskarakterisering ved bruk av ultralyd.

Abstract

Tissue characterization involves classifying tissue types based on some distinguishable property of the tissue. There are a wide variety of applications where distinguishing one type of material from another provides benefit and insight. One such application is in medicine where the identification of both pathological and healthy tissue can provide added value in diagnosis and further treatment planning. Another example is in the meat industry where tissue composition is strong predictor of meat quality and eating experience. In both applications there is a strong need for non-invasive characterization methods. In the medical sector, invasive methods can cause discomfort or even harm, and are often hard to conduct as navigation can be challenging. Correspondingly, in the meat industry, unnecessary harm or discomfort should be minimized and characterization done *post-mortem* limits genetic progress as selection for breeding must be determined through siblings.

Performing *in vivo* imaging is typically done with three main modalities; Ultrasound (US), X-ray Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Out of these three, ultrasound is the most portable and inexpensive method, albeit with the poorest overall image quality.

In recent years, the field of medical imaging analysis has undergone a revolution. With the increasing popularity of artificial intelligence and machine learning, more and more of these techniques are being utilized in automatic image analysis. Deep learning has proven to be especially successful, a data-driven approach able to distinguish complex features in data.

This work investigates the potential for using deep learning in tissue characterization applications, and at what point the performance of the models becomes limited by the underlying data. It is found that for segmentation of CT volumes, deep learning achieves a high performance allowing automatic labeling of volumes and enabling robust Atlas segmentation. The same approach is used in an attempt to quantify fat content from pulse-echo ultrasound images, producing results comparable to state-of-the-art approaches. However, for high fat content the results deteriorate, producing a limit on the range of applicability due to decreasing im-

age quality.

To produce robust tissue characterization with ultrasound, investigation of other acoustic parameters is necessary. In particular it is found that the nonlinear bulk elasticity of soft tissues has a large variation compared to other acoustic parameters. In addition, the parameter is especially sensitive to fat content. Further, it is shown that this parameter can be measured using a dual frequency approach. The presented approach has the characteristic hallmarks of an approach in which deep learning can be successful, and the potentially increased data quality can provide a more robust tissue characterization method. However, as deep learning is data-driven, the main challenge will be the procurement of realistic training data, consequently becoming an interesting area of further study.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* (Ph.D.) at the faculty of Medicine and Health Sciences at the Norwegian University of Science and Technology (NTNU). This work was funded through the Norwegian Research Council's industrial Ph.D. program with Norsvin SA as the industrial partner in collaboration with the Norwegian University of Science and Technology, University of Oslo (UIO) and SURF technology AS. The work has been carried out partly at the Department of Circulation and Medical Imaging (NTNU) and partly at the Department of Informatics (UIO). The main supervisor was Bjørn A.J. Angelsen (NTNU / SURF) with co-supervisors Jørgen Kongsro (Norsvin), Sverre Holm (UIO), Alfonso Rodriguez-Molares (NTNU) and Anne C. Elster (NTNU).

Acknowledgements

As I have had the privilege of working with so many different groups and people over the course of my work, there are a lot of people to thank.

Firstly, I would like to sincerely thank Bjørn Angelsen, working together so closely, not just over the course of the PhD, has been an absolute honor. His brilliant mind and creativity is profoundly awe inspiring and has been a critical source of inspiration. I would also like to thank Alfonso Rodriguez-Molares, his multitude of skills and deep knowledge is just incredible and working with you was always tremendously fun. A special thank you to Anne C. Elster for including me in the High-Performance Computing lab and for all the delicious pizza!

From Norsvin I would like to give a profound thank you to my advisor Jørgen Kongsro. His perpetual encouragement has been incredibly motivating. His including persona ensured I was taken in to the Norsvin family which both broadened my horizons and gave me many new friends. I would also like to thank Lars Erik Gangsei, working with you and Jørgen was fun, demanding and hugely rewarding. I hope we get many more opportunities in the future.

From the University of Oslo I would like to sincerely thank Sverre Holm, whose enthusiasm and guidance has been crucial for the completion of this work.

In addition, a huge thank you is warranted to Anne Schistad Solberg. Our discussions and your feedback has been a tremendous help and has been highly appreciated.

To my colleagues, but more importantly very good friends; Ole Martin Brende, Ola Finneng Myhre and Stian Solberg. Thank you for being you. Your humor, encouragement and support has made these past years incredible. An added special mention of Stian is warranted. Digging deep into the realm of nonlinear acoustics with you has been fun, and our close collaboration over the last year instrumental for the completion of this thesis.

A special thank you to my partner, Karoline. Thank you for your love, support and patience. You kept me grounded and ensured I never got too out of balance. You always believed in me, even when I didn't. I am looking forward to returning to normal with you.

Finally, my family and friends; thank you for continued love and support. I am looking forward to spending much more time with you now.

Contents

I	Thesis	1
1	Introduction	3
1.1	Acoustic Wave Propagation & Scattering	4
1.1.1	Ultrasound Tissue Characterization	7
1.1.2	Estimation of the acoustic nonlinear response	8
1.2	SURF imaging	10
1.2.1	Received signal model	12
1.2.2	Transducer design & Transmit beam considerations	13
1.2.3	Estimation of tissue nonlinearity with SURF	14
1.3	Machine Learning	16
1.3.1	Neural Networks	16
1.3.2	Challenges in Neural Networks	24
1.4	Thesis Summary	27
1.5	Other contributions	30
II	Papers	39
A	The use of deep learning to automate the segmentation of the skeleton from CT volumes of pigs	41
A.1	Introduction	41
A.2	Materials And Methods	43
A.2.1	3D Through 2D Projections	43
A.2.2	Task Description	45
A.2.3	CNN Architecture	49
A.3	Experimental Setup	52
A.3.1	Data Annotation	52
A.3.2	Training, Validation and Testing	53
A.3.3	Performance Evaluation	53
A.4	Results And Discussion	54

A.4.1	Segmentation Of The Spine	55
A.4.2	Classification Of The Spine	55
A.4.3	Segmentation Of Limbs	55
A.4.4	Classification Of Limbs	56
A.4.5	Segmentation Of Ribs	57
A.4.6	Implications	57
A.5	Conclusion	58
B	<i>In vivo</i> Prediction of Intramuscular Fat using Ultrasound and Deep Learning	63
B.1	Introduction	63
B.2	Materials & Methods	64
B.2.1	Data	64
B.2.2	Convolutional Neural Network	65
B.3	Results	66
B.4	Discussion	67
B.5	Conclusion	68
C	Exploiting Ballou’s Rule for Improved Tissue Classification	71
C.1	Introduction	71
C.2	Background	72
C.2.1	Nonlinearity parameter B/A	72
C.2.2	Tissue characterization based on β_n	74
C.3	Modeling of β_n in soft tissue	75
C.3.1	Regression model	75
C.3.2	Thermodynamic Model	77
C.4	Results	82
C.5	Discussion	85
C.6	Conclusion	88
D	Dual frequency transducer design for suppression of multiple scattering	93
D.1	Introduction	93
D.2	Radiation Apertures	95
D.2.1	Optimization metrics	95
D.2.2	High Frequency Field Considerations	96
D.2.3	Low Frequency Field Considerations	96
D.3	Stack Design	98
D.3.1	Single layer isolation section	100
D.3.2	Triple layer isolation section	100
D.4	Results	102

D.5	Conclusion	102
E	SURF Nonlinear Bulk Elasticity Imaging	105
E.1	Introduction	105
E.2	Theory	107
E.3	Estimation Approach	109
E.3.1	Delay estimation	109
E.3.2	Estimation of β_p	110
E.4	Materials & Methods	112
E.4.1	Experimental setup	112
E.4.2	Simulations	114
E.5	Results	114
E.5.1	Simulations	114
E.5.2	Experimental	117
E.6	Discussion	120
E.7	Conclusion	122
III	Complete Index	127
	List of Tables	129
	List of Figures	133

Part I

Thesis

Chapter 1

Introduction

Tissue characterization involves classifying materials based on the differences in their physical properties. In medicine, the properties of healthy and pathological tissue are often dissimilar, enabling tissue characterization to provide an added diagnostic value [1, 2]. Being able to characterize tissue in a non-invasive manner is crucial as invasive methods such as biopsies can cause discomfort and even harm. In addition, as navigation in biopsies is challenging, these can often be unreliable and cause false negatives [3, 4]. Applications of non-invasive tissue characterization is not only limited to the medical sector. For instance, in the meat industry, being able to characterize the amount of fat and muscle is a strong predictor of meat quality [5]. In a commercial breeding program for farmed animals, i.e. pigs, *in vivo* assessment of meat quality is crucial in order to select for improved perception of pork in the live selection candidates. Ultrasound has been used with only moderate success due to low accuracy, especially on lean type animals, but improved imaging and analytical methods, has the potential to improve the accuracy of such measurements.

The most prominent imaging techniques for tissue characterization are magnetic resonance imaging (MRI), X-ray computed tomography (CT) and ultrasound (US). MRI and ultrasound are typically used for imaging of soft tissues, while CT is highly suited for imaging of bone and harder materials. Compared with the other two, ultrasound is the most inexpensive and portable modality. On the other hand, the image quality is typically inferior to that of MRI and CT.

Recently, machine learning techniques such as deep learning [6], have proven capable of solving complex image recognition tasks [7]. Consequently, a revolution in medical imaging analysis is taking place as more and more automated analysis tools become deep learning based [8, 9]. However, as these methods are data-driven, a common understanding is that the performance of these models is limited by the volume and quality of the underlying data used to train these al-

gorithms. This is typically described as, *Garbage in - garbage out*, meaning that the performance of the models is only as good as its input data. The question hence becomes how do these types of algorithms fare in tissue characterization applications. And to what degree these advanced processing algorithms is able to compensate for quality of the input data.

This thesis seeks to investigate the potential for use of machine learning in tissue characterization applications, and at what point the performance of the models become data limited. A special emphasis is given to the detection of fat using ultrasound. As the acoustic nonlinear response of fat is significantly higher than other tissue types, a dual frequency approach is showcased to potentially increase data quality.

1.1 Acoustic Wave Propagation & Scattering

The equations describing acoustic wave propagation stem from three governing equations; the continuity equation, the momentum equation and the equation of state. These are given as [10–12],

$$\text{Continuity: } \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \underline{u}) = 0 \quad (1.1a)$$

$$\text{Momentum: } \rho \left(\frac{\partial \underline{u}}{\partial t} + \underline{u} \nabla \underline{u} \right) + \nabla p = 0 \quad (1.1b)$$

$$\text{State: } p = p(\rho). \quad (1.1c)$$

Here $p(\underline{r}, t)$ is the instantaneous, local acoustic pressure as a function of space vector, \underline{r} , and time t , and $\underline{u}(\underline{r}, t)$ is the particle velocity. The instantaneous mass density is given by $\rho(\underline{r}, t) = \rho_0(\underline{r}, t) + \rho_1(\underline{r}, t)$, with ρ_0 as the equilibrium density for $p = 0$ and ρ_1 is the excess density due to the acoustic compression.

For acoustic waves in fluids and solids, the state equation in (1.1c), is commonly defined through a second order Taylor series expansion of the acoustic pressure in relation to the mass density. This is given as, [10, 11]

$$p = A \frac{\rho_1}{\rho_0} + \frac{B}{2} \left(\frac{\rho_1}{\rho_0} \right)^2 = A \frac{\rho_1}{\rho_0} \left(1 + \frac{B}{2A} \left(\frac{\rho_1}{\rho_0} \right) \right), \quad (1.2)$$

$$A = \rho_0 \left(\frac{\partial p}{\partial \rho} \right)_{0,s} = \frac{1}{\kappa} \quad B = \rho_0^2 \left(\frac{\partial^2 p}{\partial \rho^2} \right)_{0,s},$$

where subscript 0 denotes differentiation around equilibrium density ρ_0 and subscript s denotes the assumption of isentropic conditions. The isentropic compressibility is given by $\kappa(\underline{r})$ and $A(\rho_1/\rho_0)$ hence describes the linear bulk elasticity of the material. Consequently, the term $(B/2A)(\rho_1/\rho_0)$ describes the deviation from linear elasticity, where B/A is commonly known as Beyer's nonlinearity

parameter [10, 11]. Equation (1.2) describes nonlinearity up to second order, but can be expanded up higher orders to produce higher orders of nonlinearity, *e.g.* C/A [13, 14], for typical ultrasonic pressures in fluids and solids, these effects are negligible.

To produce a wave equation from our second order equation of state in (1.2) it is beneficial to adopt the Lagrange spatial description. In this case the coordinate \underline{r} refers to the location of the material point in the unstrained material, *i.e.* equilibrium, and $\underline{\psi}(\underline{r}, t)$ describes the instantaneous, local displacement of a material point from its equilibrium, produced by particle vibrations in the wave. In this case, the continuity equation in (1.1a), takes the form

$$-\nabla \underline{\psi} = \frac{\rho_1}{\rho} = \frac{\rho_1}{\rho_0 + \rho_1}, \quad \frac{\rho_1}{\rho_0} = -\frac{\nabla \underline{\psi}}{1 + \nabla \underline{\psi}} \approx -\nabla \underline{\psi}(1 - \nabla \underline{\psi}). \quad (1.3)$$

By inserting into (1.2), and neglecting terms beyond second order in $\nabla \underline{\psi}$, the isentropic state equation becomes

$$p = -A \nabla \underline{\psi} (1 - \nabla \underline{\psi}) + \frac{B}{2} (\nabla \underline{\psi})^2 = -\frac{1}{\kappa} \nabla \underline{\psi} + \frac{\beta_n}{\kappa} (\nabla \underline{\psi})^2, \quad (1.4)$$

where $\beta_n = 1 + B/2A = 1 + B\kappa/2$ is commonly referred to as the coefficient of nonlinearity[12].

Equation (1.1c) and (1.4) describes the isentropic compression, where there is no transformation of elastic energy to heat, *i.e.* no absorption of acoustic energy in the wave propagation. Linear absorption can be introduced by adding a temporal convolution term, $h_{ab} * \nabla \underline{\psi}$, where $h_{ab}(\underline{r}, t)$ is a convolution kernel representing absorption due to deviation from fully isentropic compression.

For the analysis of wave propagation, it is more intuitive to formulate an equation in terms of the acoustic pressure, p . By inverting (1.4) to second order in p and including the absorption term only up to first order in p , we obtain the relation,

$$-\nabla \underline{\psi} = [1 - \beta_p(\underline{r})p(\underline{r}, t)] \kappa(\underline{r})p(\underline{r}, t) + h_{ab}(\underline{r}, t) * \kappa(\underline{r})p(\underline{r}, t). \quad (1.5)$$

Here, the term $\beta_p(\underline{r}) = \beta_n(\underline{r})\kappa(\underline{r})$ describes the nonlinear bulk elasticity of the material. The total attenuation of a propagating acoustic wave is the sum absorption due to heating given by h_{ab} and scattering of the wave.

To develop a full wave equation, it is time to include the momentum equation in (1.1b). If we assume that the displacements are small compared to the curvature of the wave front, we can neglect the convective term, and thus obtain,

$$\rho_0 \frac{\partial^2 \underline{\psi}}{\partial t^2} \approx -\nabla p, \quad \frac{\partial^2 \nabla \underline{\psi}}{\partial t^2} = -\nabla \left(\frac{1}{\rho_0} \nabla p \right). \quad (1.6)$$

The mass density, compressibility, nonlinear bulk elasticity and absorption have spatial variations in many practical materials such as soft tissues. Slow variations mainly influence the forward propagation of the wave, while rapid variations produce scattering of the wave. Consequently, it is beneficial to split the variation in these parameters into a slowly and rapidly moving part, *i.e.*,

$$\rho(\underline{r}) = \rho_a(\underline{r}) + \rho_f(\underline{r}) \quad (1.7a)$$

$$\kappa(\underline{r}) = \kappa_a(\underline{r}) + \kappa_f(\underline{r}) \quad (1.7b)$$

$$\beta_p(\underline{r}) = \beta_{pa}(\underline{r}) + \beta_{pf}(\underline{r}), \quad (1.7c)$$

where subscript *a* and *f* describe the slow and rapid variations respectively.

Combining Eqs. (1.4), (1.6) and (1.7), we obtain a wave equation which includes nonlinear wave propagation and scattering phenomenon,

$$\begin{aligned} & \underbrace{\nabla^2 p(\underline{r}, t) - \frac{1}{c_0^2(\underline{r})} \frac{\partial^2 p(\underline{r}, t)}{\partial t^2}}_{\text{Linear propagation.}} + \underbrace{\frac{\beta_{pa}(\underline{r})}{c_0^2(\underline{r})} \frac{\partial^2 p(\underline{r}, t)^2}{\partial t^2}}_{\text{Nonlinear propagation}} - \underbrace{\frac{h_{ab}(\underline{r}, t)}{c_0^2(\underline{r})} \frac{\partial^2 p(\underline{r}, t)}{\partial t^2}}_{\text{Absorption}} \\ & = \underbrace{\frac{v(\underline{r})}{c_0^2(\underline{r})} \frac{\partial^2 p(\underline{r}, t)}{\partial t^2} + \nabla(\gamma(\underline{r}) \nabla p(\underline{r}, t))}_{\text{Linear sources}} - \underbrace{\frac{\sigma_n(\underline{r})}{c_0^2(\underline{r})} \frac{\partial^2 p(\underline{r}, t)^2}{\partial t^2}}_{\text{Nonlinear sources}}, \quad (1.8) \\ & c_0(\underline{r}) = \frac{1}{\sqrt{\rho_a(\underline{r}) \kappa_a(\underline{r})}}, \quad v(\underline{r}) = \frac{\kappa_f(\underline{r})}{\kappa_a(\underline{r})}, \\ & \gamma(\underline{r}) = \frac{\rho_f(\underline{r})}{\rho_a(\underline{r})}, \quad \sigma_n(\underline{r}) = \beta_{pf}(\underline{r}) + \beta_{pa}(\underline{r})v(\underline{r}). \end{aligned}$$

Here we have neglected $\nabla(1/\rho_a)$, the low amplitude terms $v(\underline{r})\beta_{pf}(\underline{r})$ of σ_n , and the second order term in the absorption. The linear propagation velocity at low field amplitudes is described by $c_0(\underline{r})$. The left hand side of (1.8) describe the spatial propagation of the wave which arise due to slow variations in the material parameters. The right hand side represent scattering sources that arise due to rapid spatial fluctuations in compressibility, $v(\underline{r})$, mass density, $\gamma(\underline{r})$ and nonlinear bulk elasticity, $\sigma_n(\underline{r})$.

The linear propagation terms of (1.8) determine the linear forward spatial propagation of the incident wave with propagation velocity $c_0(\underline{r})$, without the addition of new frequency components. Correspondingly, the linear scattering terms produce local linear scattering of the incident wave with the same frequency content as the incident wave.

The nonlinear propagation term of (1.8) provide an accumulative distortion of the forward propagating acoustic wave given by the nonlinear bulk elasticity of the medium, β_{pa} . This distortion causes the generation of new frequency content,

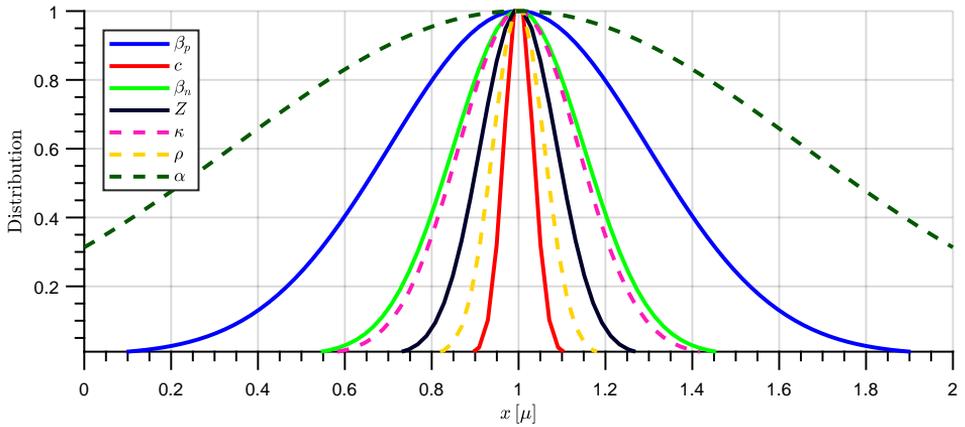


Figure 1.1: Variation of acoustic parameters in soft tissue. Gaussian fit to data in [16].

as the peaks of the wave travel faster than its troughs. This phenomena is the foundation for harmonic imaging. Correspondingly, the nonlinear scattering terms are due to rapid fluctuations in the nonlinear bulk elasticity. This produces a local scattering where the frequency content of the scattered wave is altered compared to the incident wave. A typical example of such scatterers are microbubbles where the large variation in nonlinearity compared to tissue and resonant behavior causes generation of harmonics in the scattered wave.

The absorption term in (1.8) is modeled through its own set of partial differential equations, whose solution can be represented by a convolution kernel, h_{ab} . Acoustic absorption typically follows a power law model, where the energy lost to heat is proportional to the rate of compression. This is formulated as,

$$H_{ab}(\underline{r}, f) = e^{-\alpha(\underline{r})|f|^{b(\underline{r})}}, \quad (1.9)$$

where f is the frequency and H_{ab} is the Fourier transform of the convolution kernel, h_{ab} . The parameters, α , describes the loss linear in frequency, typically specified in terms of dB/cm/MHz. The exponent b specifies the frequency dependence, typically $b \sim 1 - 2$, in soft tissues. If b is a constant, the convolution kernel can be expressed with a fractional derivative [15].

1.1.1 Ultrasound Tissue Characterization

From (1.8) it is clear that there are multiple parameters that can potentially be extracted and utilized in tissue characterization. For ultrasound compression waves, many different estimation schemes have been proposed. Examples of these include, estimation of speed of sound and backscattering coefficients[1, 17], estimation of absorption coefficients [18] and estimation of the coefficient of nonlinearity

[19]. The success of these methods is tied to the precision of the method itself with respect to the physical variation of the parameter itself observed in soft tissues. In other words, a precise method is only beneficial if the variation in the estimated parameter is sufficiently large to allow distinguishing of materials.

In Fig. 1.1 a gaussian fit to the soft tissue data found in [16] is shown. Of all the parameters the variation is clearly largest in the absorption coefficient, α . However, due to the power law relationship in (1.9) this parameter is to a certain degree confounded as separating between α and b requires multiple measurements as b varies significantly in soft tissues. Out of the remaining parameters, the coefficient of nonlinearity, β_n and the compressibility, κ , has a larger variation than speed of sound, c , density, ρ , and the impedance, $Z = \rho c$. The nonlinear bulk elasticity characterized by $\beta_p = \beta_n \kappa$ has a substantially increased variability. This is due to a significant correlation between the coefficient of nonlinearity and compressibility, an observation known as Ballou's rule [11]. In addition, the coefficient of nonlinearity has been shown to be highly sensitive to tissue structure[20] and pathology[21, 22], making it a particularly promising parameter for tissue characterization.

1.1.2 Estimation of the acoustic nonlinear response

The nonlinear response of the medium comes from a pressure dependent propagation velocity,

$$c(p) \approx c_0(1 + \beta_n \kappa p), \quad (1.10)$$

where β_n is the coefficient of nonlinearity, κ is the compressibility and p is the acoustic pressure. The product of $\beta_p = \beta_n \kappa$ is the observable nonlinear response of the material or nonlinear bulk elasticity. As previously mentioned, the pressure dependent sound speed causes the peaks of the acoustic wave to propagate faster than its troughs. This causes an accumulative distortion of the wave with propagation distance, where the most well-known is the generation of harmonics. Tissue characterization based on the nonlinear response of the medium hence involves measuring this distortion and relating it to the coefficient of nonlinearity as,

$$\beta_n = \frac{c_0}{\kappa_0} D_{nl}(p), \quad (1.11)$$

where $D_{nl}(p)$ is the pressure dependent distortion.

A potential key advantage of tissue characterization based on the nonlinear response of the medium is the pressure dependence of the observed nonlinear response. As the reflection coefficient between various soft tissues is low, the magnitude of the backscattered acoustic wave is significantly less than the forward propagating wave. This essentially means that in many cases the propagation of the backscattered wave can be considered linear. Consequently, for estimation of

β_n there is no need for tomographic scanning as the accumulative distortion is only present in the forward propagation.

There are several proposed methods for estimating the coefficient of nonlinearity, β_n , in soft tissues [19]. Typical examples estimation based on sum and difference frequencies using parametric arrays [23], based on the magnitude of the second harmonic[24] or using dual frequency techniques with probe and pulse waves [25–28]. However, the approaches has so far has had little clinical success and Duck attributes this to the lack of robust echo modes [19].

Common for approaches differentiating tissue based on the nonlinear properties of the medium is that they characterize tissue based on the coefficient of nonlinearity, β_n . This is done through some manifestation of (1.11) [29]. However, unless these methods have knowledge of the compressibility and sound speed of the medium, the values of these must be assumed. However, as β_n and κ has a correlation in soft tissues, the nonlinear bulk elasticity $\beta_p = \beta_n \kappa$ has increased variation as seen in Fig. 1.1. Consequently, by characterizing based on the nonlinear bulk elasticity we can potentially increase the sensitivity of the approach and decrease the error due to erroneous assumptions of κ_0 .

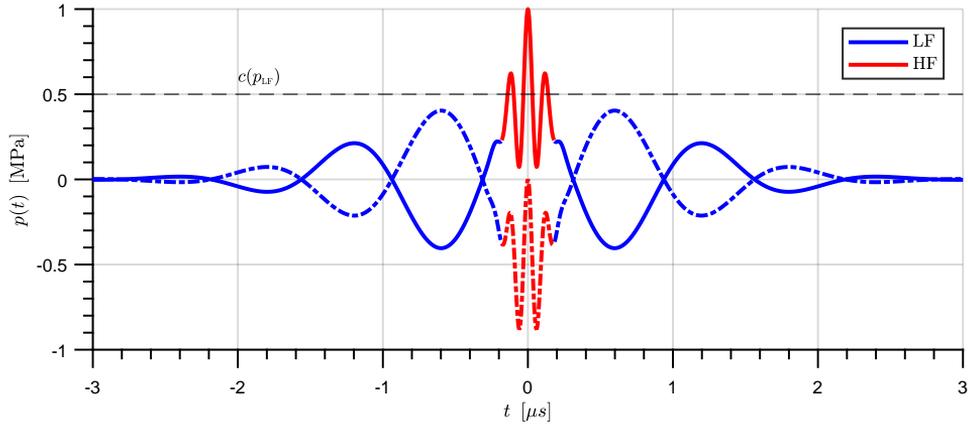


Figure 1.2: A positive and negative SURF pulse composed of a 8 MHz high frequency (HF) imaging pulse and a 0.8 MHz low frequency (LF) manipulation pulse.

1.2 SURF imaging

SURF imaging stands for *Second order Ultrasound Field* imaging and is a dual frequency band imaging technique[30]. The technique entails transmissions of multiple dual frequency pulse complexes widely separated in frequency. A typical frequency separation is a factor $\omega_{\text{HF}}/\omega_{\text{LF}} \sim 10$. An example of a SURF pulse complex is shown in Fig. 1.2.

To understand how the high frequency (HF) and low frequency (LF) interact, it is beneficial to analyze (1.8) for an acoustic pulse complex,

$$p(\underline{r}, t) = p_{\text{LF}}(\underline{r}, t) + p_{\text{HF}}(\underline{r}, t). \quad (1.12)$$

For linear propagation part there is no change as we only obtain a superposition of two wave equations. For the nonlinear terms however we obtain an explicit interaction term,

$$p(\underline{r}, t)^2 = (p_{\text{LF}} + p_{\text{HF}})^2 = p_{\text{LF}}^2 + 2p_{\text{HF}}p_{\text{LF}} + p_{\text{HF}}^2. \quad (1.13)$$

The second term describes the interaction while the first and last terms corresponds to nonlinear self-distortion, *i.e.* the generation of harmonics due to peaks propagating faster than the troughs. For any SURF imaging application, the selection of frequencies is based on which HF frequency will provide optimal imaging within the region of investigation. The LF frequency is then simply chosen $\approx \omega_{\text{HF}}/10$. Consequently, for any given imaging application, the LF typically only propagates on the order of tens of wavelengths, while the HF propagates on the order of hundreds of wavelengths. The relatively short propagation of the LF allows us to approximate its propagation as entirely linear over the imaging region.

The frequency relationship also affects the interaction term in (1.13). As the HF pulse length is much shorter than half the period of the LF, *i.e.* the bandwidth of the HF pulse $B_{\text{HF}} > \omega_{\text{LF}}/2$, as seen in Fig. 1.2, the interaction between HF and LF mainly affects the HF. The wave equation can hence be split into two, one for the LF and one for the HF. For the LF, (1.8) becomes,

$$\begin{aligned} & \underbrace{\nabla^2 p_{\text{LF}} - \frac{1}{c_0^2} \frac{\partial^2 p_{\text{LF}}}{\partial t^2}}_{\text{Linear propagation}} + \underbrace{\frac{\beta_{\text{pa}}}{c_0^2} \frac{\partial^2 p_{\text{LF}}^2}{\partial t^2}}_{\text{Self-distortion}} - \underbrace{h_{\text{ab}} * \frac{1}{c_0^2} \frac{\partial^2 p_{\text{LF}}}{\partial t^2}}_{\text{Absorption}} \\ & = \underbrace{\frac{v}{c_0^2} \frac{\partial^2 p_{\text{LF}}}{\partial t^2} + \nabla(\gamma \nabla p_{\text{LF}})}_{\text{Linear scattering sources}} - \underbrace{\frac{\sigma_n}{c_0^2} \frac{\partial^2 p_{\text{LF}}^2}{\partial t^2}}_{\text{Nonlin. self dist. sources}}. \end{aligned} \quad (1.14)$$

Where, due to the accumulative effect of nonlinearity, and the relatively short propagation distance of the LF, the nonlinear terms in (1.14) are typically neglected. In addition, the LF is typically transmitted as a plane wave further reducing the nonlinear contribution.

As the HF-LF interaction effects affect the HF band, the HF wave equation becomes modified due to the interaction. In this case (1.8) becomes,

$$\begin{aligned} & \underbrace{\nabla^2 p_{\text{HF}} - \frac{1}{c_0^2} \frac{\partial^2 p_{\text{HF}}}{\partial t^2}}_{\text{Linear propagation}} + \underbrace{\frac{2\beta_{\text{pa}} p_{\text{LF}}}{c_0^2} \frac{\partial^2 p_{\text{HF}}}{\partial t^2}}_{\text{Interaction}} + \underbrace{\frac{\beta_{\text{pa}}}{c_0^2} \frac{\partial^2 p_{\text{HF}}^2}{\partial t^2}}_{\text{Self-distortion}} - \underbrace{h_{\text{ab}} * \frac{1}{c_0^2} \frac{\partial^2 p_{\text{HF}}}{\partial t^2}}_{\text{Absorption}} \\ & = \underbrace{\frac{v}{c_0^2} \frac{\partial^2 p_{\text{HF}}}{\partial t^2} + \nabla(\gamma \nabla p_{\text{HF}})}_{\text{Linear scattering sources}} - \underbrace{\frac{2\sigma_n p_{\text{LF}}}{c_0^2} \frac{\partial^2 p_{\text{HF}}}{\partial t^2}}_{\text{Nonlin. inter. sources}} - \underbrace{\frac{\sigma_n}{c_0^2} \frac{\partial^2 p_{\text{HF}}^2}{\partial t^2}}_{\text{Nonlin. self dist. sources}}. \end{aligned} \quad (1.15)$$

Note, by setting $p_{\text{LF}} = 0$, the equation reduces to (1.8). From the first interaction term on the left hand side in (1.15), we see that this term alters the linear propagation as,

$$\nabla^2 p_{\text{HF}} - \left(\frac{1 - 2\beta_{\text{pa}} p_{\text{LF}}}{c_0^2} \right) \frac{\partial^2 p_{\text{HF}}}{\partial t^2} \approx \nabla^2 p_{\text{HF}} - \frac{1}{c_0^2 (1 + 2\beta_{\text{pa}} p_{\text{LF}})} \frac{\partial^2 p_{\text{HF}}}{\partial t^2}, \quad (1.16)$$

where we have utilized the approximation $1 - x \approx 1/(1 + x)$ as $|2\beta_{\text{pa}} p_{\text{LF}}| \ll 1$. Consequently, the fundamental interaction that can be exploited in SURF imaging is the manipulation of the speed of sound by the LF, observed by the HF, *i.e.*,

$$c(p_{\text{LF}}) = \sqrt{c_0^2 (1 + 2\beta_{\text{pa}} p_{\text{LF}})} \approx c_0 (1 + \beta_{\text{p}} p_{\text{LF}}). \quad (1.17)$$

where p_{LF} is the average LF pressure across the HF pulse length. Based on the polarity of the manipulation pressure, p_{LF} , the co-propagating HF pulse propagates

faster or slower, compared to a HF pulse without any co-propagating LF. The wave will hence arrive earlier or later at a spatial point \underline{r} given as,

$$t(\underline{r}) = \int_{\Gamma(\underline{r})} \frac{ds}{c(s, p_{LF}(s))} \approx \int_{\Gamma(\underline{r})} \frac{ds}{c_0(s)} - \int_{\Gamma(\underline{r})} \frac{\beta_p(s) p_{LF}(s)}{c_0(s)} ds \quad (1.18)$$

$$= t_0(\underline{r}) + \tau_x(\underline{r}),$$

where,

$$\tau_x(\underline{r}) = - \int_{\Gamma(\underline{r})} \frac{\beta_p}{c_0} p_{LF}(s) ds \quad (1.19)$$

is an accumulative delay or advancement, called the *nonlinear propagation delay* (NPD). The integration is done along orthogonal trajectories of the HF wavefront, $\Gamma(s)$, for plane waves we have $ds = dz$, and the limits of the integration are simply along the depth axis $[0, z]$ [31]. As mentioned in Sec. 1.1.2, in soft tissues, the magnitude of the backscattered wave is reduced drastically. Consequently, the nonlinear LF manipulation is negligible in the back propagation in soft tissues, causing the nonlinear propagation delay only to accumulate in the forward propagating wave. However, in the presence of strong specular reflectors this approximation is not necessarily valid [32].

Note that if the HF is located at the zero crossing of the LF the average pressure across the HF pulse length is zero. In this case the HF does not experience a pure delay or advancement, but rather a *pulse form distortion* (PFD), as the head and tail of the pulse propagates at different velocities. This causes an expansion or compression of the HF pulse corresponding to a frequency shift[31, 33].

By transmitting multiple SURF pulse complexes with different LF polarities the nonlinear properties of the medium can be investigated. The technique was originally developed for imaging of micro bubbles above resonance, where the LF manipulates the size of the micro bubble [34, 35]. In addition, the approach has shown to be able suppress reverberation noise as characteristic delay of reverberations is lower than a signal originating from investigated depth [36–38].

1.2.1 Received signal model

In order to investigate the nonlinear properties of the medium a comparison must be made between at least two different SURF pulse complexes. For the typical cases of transmitting a SURF complex with a positive LF, zero LF and negative LF, the received signal is modeled as,

$$y_+(t) = \tilde{v}_+(t) *_t x(t - \tau_x) + \tilde{l}_+ *_t n(t - \tau_n) \quad (1.20a)$$

$$y_0(t) = x(t) + n(t) \quad (1.20b)$$

$$y_-(t) = \tilde{v}_-(t) *_t x(t + \tau_x) + \tilde{l}_- *_t n(t + \tau_n), \quad (1.20c)$$

where $x(t)$ is the first-order signal originating from the depth investigated. The noise component, $n(t)$, originates from other acoustic sources in other positions than the depth investigated, but is mapped to the same region due to a similar time of flight. A typical example of such sources are reverberations or multiple scattering sources [38].

The convolutional filter \tilde{v}_{\pm} and the delay τ correspond to pulse form distortion (PFD) and nonlinear propagation delay (NPD) respectively. For simplicity we have assumed that the delay development for the positive and negative LF SURF complexes are the same, but of opposite sign, *i.e.* $\tau_{x,+} = -\tau_{x,-}$. The noise term in (1.20) has a different delay, τ_n , than the first-order signal, τ_x . If the noise component originates from reverberation sources, we have $\tau_x > \tau_n$, if in addition, the delay development is linear we have $\tau_x \approx 2\tau_n$ [36, 38]. The form modification term for the noise, \tilde{l} , resembles the PFD, but also includes a distortion due to the interference of noise components.

Estimating the delay between two polarities can be done through cross correlation. To get the best possible estimate of the first-order delay, the PFD should be negligible, *i.e.* $\tilde{v}_{\pm} \approx 1$ and the signal-to-noise ratio should be high, *i.e.* $x(t) \gg n(t)$. Assuming negligible pulse form distortion the observed delay can be modeled simply as a weighted sum of the first-order and noise delay as,

$$\tau_y(t) = a(t)\tau_x(t) + (1 - a(t))\tau_n(t). \quad (1.21)$$

Here, the parameter a describes the signal to noise ratio, meaning that if $a \approx 1$ we observe the first-order delay. However, if $a \approx 0$, the delay observed is the delay of noise sources in the image.

1.2.2 Transducer design & Transmit beam considerations

Compared to other dual band techniques such as the work by Fukukita[28], pulse form distortion is unwanted in SURF imaging. Consequently, transducer design is of vital importance to produce a uniform manipulation of the observed sound speed. The ideal case would be if the manipulation pressure can be approximated as constant over the entire imaging region, in this case the propagation delay is entirely defined by the nonlinear elasticity of the medium, *i.e.*,

$$\tau_x(z) = -p_{\text{LF}} \int_0^z \frac{\beta_p}{c_0} dz. \quad (1.22)$$

For a homogeneous medium this translates into a linear delay development. To achieve this a high frequency separation between HF and LF is necessary. Myhre *et al.* showed that a high frequency separation between HF and LF is achievable by using a three matching layer structure between HF and LF transducers [39]. However, it is important to keep in mind that as the frequency ratio increases so

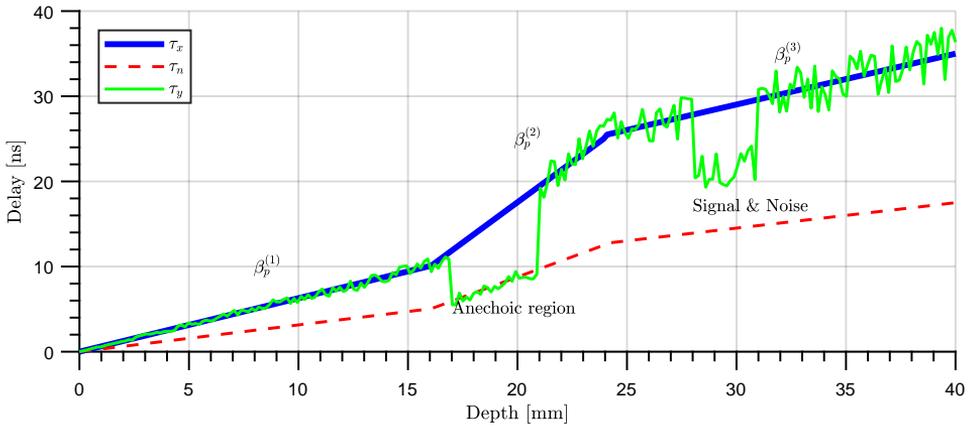


Figure 1.3: Example of development of first-order delay, τ_x , noise delay, τ_n and the measured signal delay τ_y .

does the LF pulse length. Consequently, it is important to ensure that there is a negligible interaction in the backpropagation of the HF.

To ensure spatial uniformity the LF should be a plane wave. However, for a focused HF, this means that the phase relation between HF and LF will vary due to the focusing curvature. This curvature can potentially produce an accumulative PFD, affecting the received signal.

1.2.3 Estimation of tissue nonlinearity with SURF

Evaluating the nonlinearity of the medium directly, β_p , using SURF, has been suggested [30], but has until now, not been pursued. Through differentiation of (1.19), the tissue nonlinearity can be extracted as,

$$\beta_p = \frac{c_0}{p_{\text{LF}}(z)} \frac{\partial \tau_x}{\partial z}, \quad (1.23)$$

under the assumption of a plane wave. However, in a pulse-echo setup, the delay measured between two different SURF pulse complexes, does not necessarily reflect the delay development of the first-order signal, τ_x . As the received signal is composed of a first-order signal and a noise component, as discussed in Sec. 1.2.1, the delay we observe, τ_y , varies depending on the signal-to-noise ratio, a . This variation can, in a general form, be described by (1.21).

A simplified example of a typical observed delay development is given in Fig. 1.3. The development of the first-order delay, τ_x , and noise delay, τ_n , is shown through a medium composed of three regions with different nonlinear bulk elasticity, $\beta_p^{(i)}$. The noise delay in this example follows $\tau_n(z) \approx \tau_x(z/2)$, a typical development of delay for reverberation noise in SURF imaging [36, 38, 40].

The variation in the observed delay in Fig. 1.3 can be rapid due to fluctuations in the signal-to-noise ratio, a , but also due to random speckle variations between two SURF signals. In addition, for large fluctuations in the signal to noise ratio, nearly discontinuous behavior can be observed. This is shown in the anechoic region, where there is no first-order signal, *i.e.* $a \approx 0$, and in region at ~ 30 mm with low echogenicity, where the signal-to-noise ratio is low, $\alpha \approx 0.5$.

Even in this relatively simple situation it is clear that a pure differentiation approach, as in (1.23), is not feasible. Differentiation of a noisy signal amplifies the noise and consequently, overpowers the underlying behavior. In addition, in a practical situation, neither the first-order delay or noise delay are this well-behaved and the inclusion of other acoustic noise sources and beamforming artefacts further increase the variability. Consequently, in order to achieve robust estimation of tissue nonlinearity from the NPD, there is a need for identifying regions with a high signal to noise ratio $a \approx 1$ and perform a more robust fitting between such regions. In this thesis a model-based estimation framework is proposed.

Having analyzed countless delay developments of the form in Fig. 1.3, identifying changes in the gradient has become second nature. With a little experience, one is able to a large degree automatically identify noisy or untrustworthy regions and exclude them from the analysis by visually interpolating between them. However, due to the complicated physics and multitude of different noise sources and artefacts, describing all of these mathematically is challenging. Consequently, the "easy to see, hard to define" nature of the problem indicates that a machine learning approach is highly suited. The main challenge in such an approach would be the generation of realistic training data, as measurement of τ_x , *in vivo* or *in vitro*, is not straightforward. A possibility would be the creation of synthetic training data based on realistic simulations, for instance using k-Wave[41]. In any case, a machine learning approach to modeling of delay development looks promising, and is an exciting avenue for future work.

1.3 Machine Learning

Machine learning (ML) is an umbrella term for various soft-modeling approaches. The idea is simple, based on set of observations with input X and output Y find a mapping function such that,

$$f(X) = Y. \tag{1.24}$$

If we have a set of known observations or training samples (X_T, Y_T) , we can approximate a mapping function $f_T(X_T) = Y_T$. The fundamental idea is that, given a sufficient sample size, X_T spans the distribution of the true variable X and the approximated relation between the input and output should be close to the true relation, $f_T(X) \approx f(X)$. This is called *generalization* which is the fundamental criterion for machine learning methods to succeed.

Machine learning approaches are typically split into two main categories. *Supervised* learning and *unsupervised* learning. In supervised learning we approximate the mapping directly by training our algorithm to find a mapping function between X and Y . Broadly speaking everything from simple methods such as linear regression, logistic regression and support-vector machines can be characterized as supervised machine learning approaches. In unsupervised learning we do not have access to the output Y and hence cannot approximate the mapping function directly. In these techniques we typically try to find characteristic features in our input X which are then likely to be important characteristics in the mapping function $f(X)$. Examples of approaches operating in this fashion are principle component analysis (PCA) and clustering methods. Neural networks is a generalization of these methods and can hence be used in both a supervised and unsupervised manner.

Although machine learning is a broad concept, in the recent years it has become almost synonymous with Deep Learning, a term describing the use of many layered neural networks [6]. This section gives an overview of the fundamentals of neural networks and some of the key challenges in applied deep learning.

1.3.1 Neural Networks

Neural networks are nothing new, the neural network was first algorithmically described in 1950s by psychologist Rosenblatt [42]. The paper inspired engineers, physicists and mathematicians to devote their research to this interesting approach to model the human brain. However, the key issue was that the approach required a substantial amount of computing power and consequently, the application of the early neural networks were limited. The real breakthrough for neural networks came in 2012, when a convolutional neural network called AlexNet [7] won the ImageNet challenge[43], a competition involving classifying millions of images into thousands of distinct classes. AlexNet outclassed all prior competitors by reducing the top-5 classification error from 26% to 15.3%, showing that neural

networks were now able to solve complex applications and essentially started the deep learning revolution. Since 2012 the error rates have steadily declined with even more complex neural networks, even surpassing human performance in 2015 with the introduction of the ResNet [44].

The success of the neural networks have been attributed to the computation power offered by general purpose graphics processing units (GPGPU). The parallel nature of the GPGPU allows complex models to be trained in reasonable amounts of time. In fact, due to the proven performance of neural networks, hardware developers have started making custom made processing units that are specialized in performing matrix operations and convolutions, the core computations in neural networks. An example of such a processor is the tensor processing unit (TPU) [45, 46] based on a systolic array architecture [47].

1.3.1.1 The Perceptron

Before delving into the specifics of neural networks, it is beneficial to understand where the idea comes from, and its biological inspiration. In 1958, psychologist Frank Rosenblatt presented, what was to become the foundation for neural networks, namely the perceptron [42]. The idea was biologically inspired based on the understanding, at that time, of how the human brain perceives and learns. The model was based on the structure of neurons in the brain and how they interact. Today, we know that physiological neurons are much more complex and diverse, and consequently these models are crude simplifications of how these complex systems operate. Consequently, although the inspiration is clear, one should be careful in pushing the brain analogy to far.

In its simplest form a neuron consists of *dendrites*, *synapses*, a nucleus and an *axon*. The dendrites are connected to other surrounding neurons and sense stimuli coming from them. The importance of the stimuli coming from these surrounding neurons is determined by synapses. The cell body, or nucleus, interprets these stimuli and if a certain combination and magnitude of stimuli is observed, the neuron fires, sending out an electrical signal along the axon. Mathematically, this can be described as,

$$f(x) = \begin{cases} 1 & \text{if } \underline{w}^T \underline{x} + b > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (1.25)$$

which is a representation of Rosenblatt's perceptron. Here, the set of stimuli, \underline{x} , is analogous to the signals sensed by the dendrites. The cell body then interprets these stimuli through the linear operation, $\underline{w}^T \underline{x} + b$, where the weights \underline{w} correspond to the synapses, determining the influence of each stimuli. The bias, b , determine the threshold value for the neuron to fire, causing a signal to be transmitted along the axon to other interconnected neurons.

In the human brain there are hundreds of billions of neurons and over the

course of a human life the connections between these are constantly changing and shifting through learning. As children, we learn through positive or negative reinforcement, *i.e.* when we do something good, we get ice-cream, when we do something bad, we get yelled at. Consequently, we learn that some actions have a certain consequence, and the connections and synapses in our brain shifts in order to remember these correlations in the future. The same principle applies to the perceptron in (1.25). Initially, the weights, or synapses, of the perceptron, \underline{w} , are randomly initialized, analogous to a new-born child. In order for perceptron to learn to perceive a specific phenomenon, it must be trained on a set of stimuli \underline{x} with a known desired response d . In order to achieve the desired response, the synapses, or weights, must be changed as,

$$\begin{aligned}\underline{w}(t+1) &= \underline{w}(t) - \epsilon \frac{\partial}{\partial \underline{w}} [d - f(\underline{x}; w_i(t))], \\ &= w_i(t) + \epsilon [d - f(\underline{x}; w_i(t))] \underline{x}.\end{aligned}\tag{1.26}$$

For each time-step t , the weights are updated to increase the likelihood of producing the desired output d , where reinforcement comes from the comparison between produced and desired output. The parameter ϵ , describes how quickly the perceptron should learn the relation between stimuli and output, typically called the *learning rate*.

The initial perceptron in (1.25), today often called a linear binary classifier, had limited representational power as it was only able to solve linearly separable problems. However, it was later showed that the representation power could be increased by creating multiple layers of perceptrons connected to each other. This layered structure became the foundation for neural networks.

1.3.1.2 Mathematical description

The smallest component in the neural network, called a neuron, can be defined in terms of its inputs \underline{x} and output or activation y as,

$$\sigma(b + \underline{w}^T \underline{x}) = y.\tag{1.27}$$

where \underline{w} is a weight vector for each input in \underline{x} , b are biases and σ is an applied non-linearity function. There are many different nonlinearity functions with different characteristics, but the most common ones are the rectified linear unit (ReLU) and the sigmoid function. A neural network is formed by connecting multiple neurons together. These are arranged in *hidden layers* where the output of each neuron in a layer is connected to the input of the neurons in the next layer. Similar to (1.27) the output from each layer is now described as,

$$\sigma(\underline{b} + \mathbf{W}\underline{x}) = \underline{y}\tag{1.28}$$

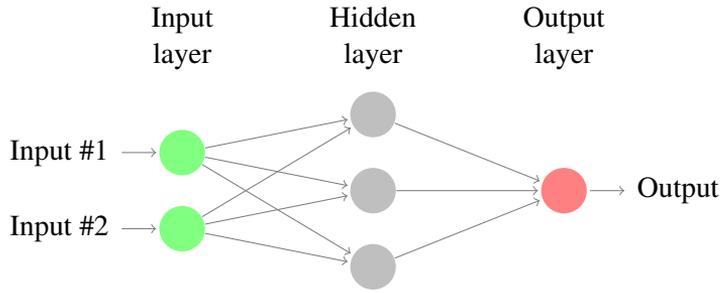


Figure 1.4: Simple neural network

where each row of \mathbf{W} represents a neuron and each entry in \underline{y} represents the activation of a single neuron in the layer. The layers at the start and end of the chain are called input and output layers respectively. An illustration of a simple neural network is shown in Fig. 1.4.

For a two-layer neural network the output of the second layer is given as,

$$\sigma \left[\mathbf{W}^{(2)} \left(\sigma \left[\mathbf{W}^{(1)} \underline{x} + \underline{b}^{(1)} \right] \right) + \underline{b}^{(2)} \right] = \underline{y}^{(2)} \quad (1.29)$$

where superscripts (1) and (2) refer to the first and second layer respectively. Note that if σ is a linear function, the weight matrices can be coalesced into a single matrix and the neural network would only be able to model a linear relationship between input and output. If the operation within a hidden layer is a linear matrix-vector product, as in this example, the layer is called a *fully connected* layer. For fully connected layers, the number of weights is determined by the number of neurons in the hidden layer and the size of the input vector \underline{x} . This means that for each element in \underline{x} there is N weights, where N is the number of neurons in the layer, corresponding to rows of \mathbf{W} . The total number of parameters within this layer is hence $N \times M$, where M is the number of inputs. Consequently, if \underline{x} is an image, a fully connected layer does not exploit spatial invariance, *e.g.* that an apple is an apple independent on its position within the image. In order to achieve spatial invariance, the operation within a neuron must take a different form, *e.g.* a convolution operation which forms the basis for convolutional neural networks. In this case the output from a neuron i is given as,

$$\sigma (\mathbf{W}_i * \underline{x} + b_i) = \underline{y}_i, \quad (1.30)$$

where \underline{x} is the input, which can be multidimensional and $*$ denotes the convolution operation. The convolution kernel, given by \mathbf{W}_i , is also multidimensional and the number of parameters in a convolutional neuron is given as $\prod_d k_d$, where k_d is the kernel size in dimension d . Consequently, the number of parameters in this type of

neuron is significantly less than for a fully connected layer, as the same weights are reused in all spatial locations of the input. The output activation y_i , often called a feature, becomes a filtered version of the input x , typically of the same spatial dimensions. Correspondingly, for a convolutional layer with N neurons, the output becomes N filtered versions of the input x . In deep neural networks, consisting of many convolutional layers, the first few layers typically become high pass filters, detecting edges in different orientations. Due to the hierarchical composition of objects, *i.e.* geometric shapes are a combination of edges, complex polygons is a combination of geometric shapes etc., the features recognized by convolutional layers increase in complexity with depth. Consequently, for image processing, the spatial invariance property of the convolution operation and hierarchical composition of features, the convolutional neural network is superior.

In supervised learning training a neural network involves updating the weights of each neuron by finding an update that reduces the evaluation loss. A typical loss is the *mean-squared-error*,

$$L = \frac{1}{N} \sum_{l=1}^N (y_l - \hat{y}_l)^2, \quad (1.31)$$

where y_l are the ground truth labels and \hat{y}_l are the predicted labels from the neural network for output l . To train the neural network, we perform *backpropagation* by updating the weights in such a manner as to decrease the loss *i.e.*,

$$\mathbf{W}^{(i)} = \mathbf{W}^{(i)} - \epsilon \frac{\partial L}{\partial \mathbf{W}^{(i)}}, \quad (1.32)$$

where $\mathbf{W}^{(i)}$ are the weights in layer i , $\partial L / \partial \mathbf{W}^{(i)}$ is the gradient of the output loss with respect to the weights that layer and ϵ is a step size. In a multi-layered neural network, the gradient of the loss to layer i is calculated by successively applying the chain rule. As the gradients are not numerically calculated all operations in a neural network have to be differentiable in order for backpropagation to work.

The approach depicted in (1.32) is called *gradient descent* involving stepping in the opposite direction of the direction that increases the loss. The weight update scheme is typically referred to as the optimizer which we will discuss further in the next section.

1.3.1.3 Training neural networks

For many applications where one trains a neural network model, the amount of training data is too big to fit in memory. In order to circumvent this we introduce *minibatches*, small chunks of data that we train on. Instead of operating on the entire training set the weights are updated in the same way as in (1.32), but on a batch, or subset of our training data. The size of this batch or batch size is a

tuneable hyperparameter as it can affect training performance. For large batch sizes the weight update will be calculated based on a large amount of samples and it is likely that this will produce an update that will lower the loss in future evaluations. A weight update with a small batch size however can be less globally convergent and yield an update that yields an improvement on the batch but a deterioration on the remaining samples. Large batch sizes hence produce faster convergence. However, small batch training has two main advantages. Firstly, it has potentially a lower computational cost to performance ratio as calculating the gradients from a large batch is more expensive than for a small one. Given that the small batch gradient is similar to the larger batch we can make progress in the training quicker as we only need to calculate the gradients for the minibatch. Secondly, minibatch learning exhibits a more stochastic nature which has been shown to increase performance as larger batches are more prone to getting stuck in local minima.

Training a neural network involves updating the weights to reduce the error in the objective function as discussed in the previous section. When we operate on minibatches the weight update scheme in (1.32) is called *stochastic gradient descent* (SGD). As mentioned previously the core idea is that the gradient of the minibatch approximates the gradient of the entire data distribution. SGD is the simplest optimization strategy, but many alternatives exist[48–50]. The core foundation is the same, the difference is in adaptively choosing how far we step along the gradient in each step by either increasing or decreasing the learning rate. During training we randomly select a minibatch of data to update our weights. When all the training data available has been used to update the weights we call this the completion of an *epoch*.

As initially stated the fundamental idea behind machine learning is that the training data and future input data share the same characteristics. This means that as long as the algorithm performs well on the training samples, we expect the network to also *generalize* to new samples. Essentially this means that the modeled transfer function $f_T(X_T)$ in (1.24) is also valid for any new input data X .

During training we wish to monitor how well our model generalizes. This is typically done by keeping part of our available samples for training in a separate validation set. The validation set simulates unseen data by not being used to update the weights of the neural network. Instead, at the end of each epoch, we evaluate the performance of our model on the validation set to ensure that our model is generalizing. A typical example of such performance monitoring is shown in Fig. 1.5 where both the training loss and validation loss is shown together. We observe that up until 120 epochs, both training and validation loss decrease, indicating that the model is generalizing. However, beyond this point we observe that the training loss continue to decrease, but the validation loss starts increasing. Consequently,

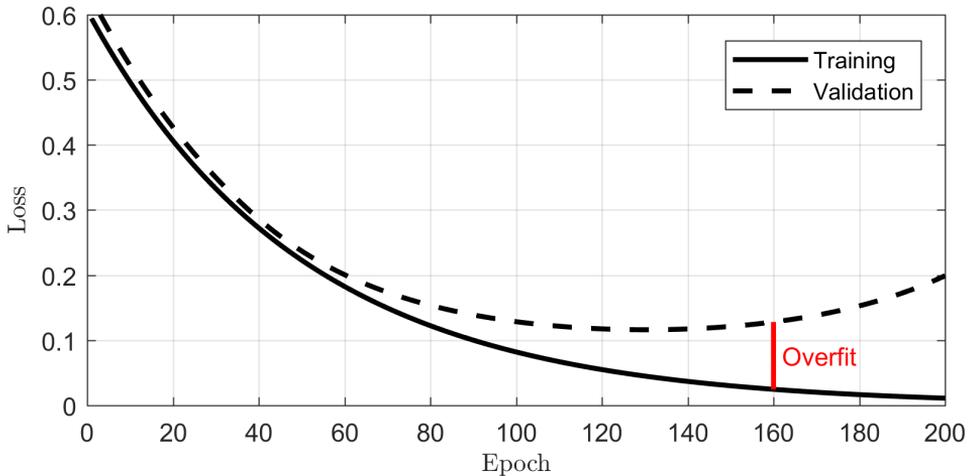


Figure 1.5: Typical training and validation loss development during supervised training.

our model starts to *overfit* to the training data, identifying features of the training data that does not generalize to unseen data.

1.3.1.4 Optimization criteria

To successfully train a neural network a suitable objective function or loss measure is key. As previously mentioned a key characteristic that should be fulfilled is that the function is differentiable. Intuitively optimizing the accuracy, how many samples the neural network is able to correctly model, seems like a sensible objective. However, as the accuracy can vary drastically for small perturbations the function does not have a smooth derivative causing inefficient learning. Consequentially, the smoothness criterion limits us to less general loss formulations but with better convergence properties. In this section a select few loss formulations will be discussed.

The loss formulation depend on whether one is doing regression or classification. For regression tasks, where a continuous output is produced, the *mean-squared-error* in (1.31) is the typical choice.

For classification tasks the choice in loss functions becomes richer. A commonly used loss is the categorical cross entropy loss with softmax,

$$\begin{aligned}
 L &= \frac{1}{N_c} \sum_{i=1}^C -\log \left(\frac{e^{f_{y_i}}}{\sum_{j=1}^N e^{f_j}} \right) \\
 &= \frac{1}{N_c} \sum_{i=1}^C \left(-f_{y_i} + \log \sum_{j=1}^N e^{f_j} \right).
 \end{aligned} \tag{1.33}$$

Here, f_{y_i} is the predicted probability of the correct class and $\sum_j e^{f_j}$ is the sum of the probabilities for all other classes. We see that a zero loss is only found when $f_{y_i} = 1 \forall i$ and $f_j = 0 \forall j \neq i$. This means that during optimization the optimizer both tries to maximize the probability of predicting the correct class but also to minimize the probability of any other class. In scenarios where two or more classes can both be correct, e.g. segmentation tasks with class overlap due to opacity typical in medical imaging, the two losses will counteract each other. One approach is to create special case classes which describe the overlap between classes, however this is in many cases unfeasible due to many potential constellations in overlap.

An approach more suited for these applications is the dice loss. This method relies on optimizing for the intersect of the binary segmentation masks for the labels y_i and predictions p_i relative to the union of the two,

$$L = \frac{1}{N_c} \sum_{i=1}^C 1 - 2 \frac{y_i \cap p_i}{y_i \cup p_i}. \quad (1.34)$$

When the predicted masks are equal to the ground truth, $p_i = y_i$, we get $L_i = 1 - 2 \frac{1}{2} = 0$. The dice loss has no dependence on the predictions in other classes as the cross entropy loss in (1.33) and hence handles overlapping classes well. However, another problem arises in multi class segmentation where both large and small objects relative to image size are to be segmented. In these scenarios the labels are said to be *imbalanced*. As the dice score is found to be correlated with the region size of the label [51], the detectors of smaller regions may be neglected. To combat this we utilize the *generalized* dice loss [51, 52],

$$L = \sum_{i=1}^C w_i \left(1 - 2 \frac{y_i \cap p_i}{y_i \cup p_i} \right) \quad (1.35)$$

where we have introduced a class weight w_i to relax or penalize deviations in certain classes more than others. To equally weight all classes the weighting term is chosen as, $w_i = 1 / \left(\sum_{j=1}^N y_{ij} \right)^2$, the inverse of label area. This ensures that the dice loss from smaller regions, e.g. a tumor, are weighted more than the large ones, e.g. the background.

An interesting extension of the generalized dice loss is the generalized Wasserstein dice loss [53]. The Wasserstein dice loss re-introduces an interlabel dependency which penalizes some constellations of misclassifications more than others. The idea behind this is to add some *a priori* information to the network about which misclassifications are meaningful and which that are not. As an example consider classification in an urban environment. A misclassification of car as a mini-van is

a tolerable misclassification whereas the misclassification of a pedestrian as a bus is not. By using the method proposed by Fidon [53] we could penalize such errors differently.

1.3.2 Challenges in Neural Networks

In this section we go through some of the major challenges in training neural networks.

1.3.2.1 Memory requirements

Training a neural network is an extremely parallel process as the gradient from output to input can be calculated concurrently for each neuron in a specific layer. This means that for large neural networks the best way of training the network is to utilize the graphics processing unit (GPU). However, compared to RAM, GPU memory is limited and can quickly become a bottleneck. The memory requirement is especially high during training as not only do the weights have to be in memory, but also the gradients for all weights. As each weight is updated potentially millions of times per epoch, moving data between RAM and GPU memory is unfeasible due to low bandwidth of the PCI bus.

1.3.2.2 Training data

One of the most common questions asked regarding neural networks is; how much data do I need? At the moment there is no concrete answer and finding that answer is an active research topic. The amount needed is dependent on several factors; size of the network, how separable the classes are, how well behaved are the classes and the overall variability in the environment. However, the gun-to-your-head answer is; more than you have.

For successful applications using neural networks there is a need for large training sets. As neural networks are becoming deeper there is an increasing need for larger and larger annotated datasets to train these [54]. The annotation process of large datasets is both expensive and time consuming, and as the data requirement grows other means of procuring them becomes increasingly important. One possibility is to pre-train a network in an unsupervised fashion named unsupervised learning. This is typically done using autoencoders [55], where the encoder part of the network is trained on unlabeled data to produce the best possible compression of the data. The encoder, having learned to identify the key characteristics of the data can then be taken out and fine tuned on smaller labeled dataset.

Another interesting approach is the creation of purely synthetic training data. One way of accomplishing this is the use of *generative adversarial networks* (GAN) where a generator network tries to construct synthetic data to fool a discriminator which is trained to distinguish between synthetic and real examples[56].

In medical imaging, access to large training sets is limited due to several

factors. Labeling is expensive as it often requires evaluation by an expert, data has to protect patient anonymity, multiple imaging modalities with different characteristics etc. For these applications the possibility of creating synthetic training data from simulations or data from other imaging modalities [57, 58] are especially interesting [59].

1.3.2.3 Training deep networks

A typical problem when training deep neural networks is what is called the vanishing gradients problem. As the weights in the neural network are updated using the chain rule to find the gradient of the error to the individual weights, the successive multiplication with potentially small numbers can cause the gradient to become very small or even zero. When this happens the network effectively stops learning as no gradients from the output reach the higher layers and hence do not update the weights. One of the main sources of these problems was the activation functions, where functions such as the tanh and sigmoid nonlinearity was extensively used in the past. These nonlinearities effectively squash the output to be either in the range $(0, 1)$ or $(-1, 1)$, also saturating the gradients. The *Rectified Linear Unit* (ReLU) nonlinearity remedied this problem to an extent as it only saturates in one direction [60]. However, ReLU does saturate the output in the negative direction, meaning that vanishing gradients are still not completely off the table. Another way of combating vanishing gradients is through the use of *batch normalization* [61]. This approach seeks to reduce internal covariate shifts by normalizing the distribution of nonlinearities causing the optimizer to be less likely to operate in the saturated regime.

For very deep neural networks the multiple path networks such as residual networks [44] can also counteract a vanishing gradient. In these networks each layer has a skip connection causing a path through the network which does not go through any layer in the network. This ensures gradient flow to all layers in the network.

1.3.2.4 Overfitting

Given that the NN has sufficient capacity, the easiest solution for (1.24), is to memorize the output for all inputs in the training data. In this case it is unlikely that the performance of the network will generalize to new, previously unseen input. This phenomenon is called overfitting. Overfitting is typically identified during training as the validation loss starts increasing while the training loss continues to fall as shown in Fig. 1.5.

Overfitting is typically attributed to too large capacity of the neural network and / or insufficient training data. Approaches that seek to prevent overfitting hence either try to constrain model capacity or artificially increase the number of training examples. The simplest way of constraining model capacity is to reduce

the size of the neural network until the network no longer is able to overfit on the training data. However, this approach can cause a insufficient model capacity that will not be able to correctly characterize the dynamics of the system. An alternative approach, called dropout [62, 63], is to randomly drop activations within the neural network, causing the output from a random subset of neurons to be set to zero. This approach does in practice limit model capacity dynamically as the network is forced to learn features redundantly to account for the random loss of features. The idea of dropout has also been shown to be beneficial in multi-branch networks [64].

1.4 Thesis Summary

As this work was funded through an industrial PhD grant from the Norwegian Research Council, with Norsvin as the industrial partner. The direction chosen for the research activity was a consequence of leveraging two aspects. What can be done within the confines of the equipment and infrastructure available today, and what equipment and infrastructure is necessary to reduce these limitations in the future.

Norsvin has, through years of data collection, access to a vast amounts of both ultrasound and CT images. Consequently, during the early exploratory stages of the project, we discovered deep learning as a promising route as a data-driven processing method both for ultrasound and CT. With the support of the deep learning group at the Department of Informatics (UIO), we chose to pursue this direction, as a means to improve the performance of tissue characterization using today's equipment. For CT data, the performance using deep learning was impressive, and quickly became area of focus. For ultrasound, deep learning yielded results competitive with state-of-the-art methods, however, a clear ceiling in its performance was observed. This ceiling was also found in literature, indicating a fundamental limitation in the data.

In parallel to exploring the capabilities of the current equipment we wanted to investigate the potential for more advanced imaging methods using dual band ultrasound (SURF). Early on we found that fat, an important tissue to characterize for Norsvin, had a particularly distinct nonlinear acoustic response. Consequently, the nonlinear response of tissue should reflect fat content. Measurement of the nonlinear acoustic response of soft tissue using SURF became a complementary research direction. The horizon for this direction was longer as the equipment available is still an early prototype and the physics are not yet completely understood. As a consequence, development of these methods has been challenging and caused a lot of headaches, but also enlightenment and finally, exultation.

In the frame of tissue characterization the main research questions hence became,

1. Can we improve characterization with deep learning models?
2. At what point does the performance become data limited?
3. How can we improve data quality and avoid these limitations?

Five papers are included as part of this thesis. Paper A and B investigates the performance of deep learning as an image processing method for characterization in CT and ultrasound images respectively. Both highlight the challenges tied to data quality and volume, however the limitations appear greater in ultrasound images. This is discussed in paper B, where it is argued that acoustic noise sources decrease

the meaningfulness of textures in the image and consequently, that other imaging methods are needed. In paper C, the nonlinear bulk elasticity is highlighted as an especially promising acoustic parameter for nonlinear ultrasound imaging. Finally, papers D and E show how this parameter can be detected using SURF imaging, where paper E argues that increased robustness can possibly be achieved by using deep learning.

Paper A: The use of deep learning to automate the segmentation of the skeleton from CT volumes of pigs

This paper applies deep learning to a coarse characterization problem where tissue is characterized based on its locality within the body (ATLAS segmentation). The problem is defined as a semantic segmentation problem which has been one of the main areas of success for deep learning. The article highlights the advanced recognition capabilities and the complexity of learned features, but also highlights its limitations of not being able to draw conclusions beyond the scope of the data set. Overall, the main purpose of the paper is demonstrate the ability of deep learning to learn complex descriptive features of the data.

Development of machine learning architecture, training and tuning of the networks done by the candidate. Collection and annotation of training data in part attributed to Jørgen Kongsro and Lars Erik Gangsei.

This work was published in Translational Animal Science, July 2018

Paper B: In-vivo prediction of intramuscular fat using ultrasound and deep learning

In this paper we investigate whether deep learning is able to extract additional information from ultrasound B-mode images to predict the intramuscular fat content in the *longissimus dorsi* muscle in pigs. Compared with the problem in paper A, the ground truth is not obvious to a human interpreter. The results highlight that there is a ceiling beyond which the textures in the input image is not representative of the ground truth label. The image quality, fundamentally limited by the physics of conventional ultrasound, reduces the meaningfulness of the textures to such a degree that the correspondence between input and output is washed out. In this case, it appears that deep learning is no magical solution, able to extract information beyond what is inherently present in the data set. Consequently, there is a need for other imaging techniques which do not have the same physical limitations to improve performance further.

Development of machine learning architecture, training and tuning of the networks done by the candidate.

This work was published in Computers and Electronics in Agriculture, November 2017

Paper C: Exploiting Ballou's rule for improved tissue classification

The acoustic coefficient of nonlinearity, β_n , has been shown to be highly sensitive to tissue structure and has for a long time been a promising route for tissue characterization using compression waves. However, the parameter has been shown to be correlated with the parameters linked to linear propagation, e.g., the speed of sound. This correlation is called Ballou's rule. In this paper we analyze the correlation between the compressibility of soft tissues and their respective coefficient of nonlinearity using both an empirical approach and a thermodynamic model. We find both empirically and theoretically a strong link between these two parameters and consequently argue that, especially for lipid rich materials, estimation of the observable nonlinear response of the medium, or nonlinear bulk elasticity, $\beta_p = \beta_n \kappa$, is a more sensitive parameter than β_n . Using the dual frequency SURF technique, we are able to measure β_p from the delay development.

Initial idea in part attributed to Sverre Holm and the candidate. Development of theoretical framework and models done by the candidate and Bjørn A.J. Angelsen.

This work has been accepted for publication, subject to minor revision, in the Journal of the Acoustical Society of America, December 2018.

Paper D: Dual frequency transducer design for suppression of multiple scattering

This paper presents the challenges of designing dual frequency transducers. The emphasis is put on reverberation suppression, however the design principle of a homogeneous LF manipulation field is vital to robust estimation of the bulk nonlinear elasticity of tissue. The LF field is an important unknown in estimation of β_p from the nonlinear propagation delay and hence having the field be as homogeneous as possible reduces the required complexity of the LF model and correspondingly estimation errors.

Shared first author with Ola F. Myhre. Design of radiation apertures done by the candidate. Wave propagation simulation tool attributed to the candidate and Stian Solberg. Transducer stack design attributed to Ola Finneng Myhre.

This work was published as a conference paper in IEEE International Ultrasonics Symposium, Tours 2016.

Paper E: SURF Nonlinear Bulk Elasticity Imaging

This paper describes the estimation techniques used to estimate the nonlinear bulk elasticity using dual band frequency transmissions. The paper highlights the challenges of differentiating noisy signals, and proposes a model based approach. The gradient changes can to some extent be identified easily by an experienced operator, but challenging to automate due to complex wave propagation. Consequently, the model based estimation approach presented is a prime candidate for replacement with deep learning based methods. However, the challenge is generation of a sufficient amount of training examples for such an algorithm to be developed.

Development of signal processing techniques done by the candidate. Simulations partly attributed to Stian Solberg. Manufacturing of phantoms done by the candidate and Alfonso Rodriguez-Molares. Measurements partly attributed to Ola Finneng Myhre.

Manuscript is being prepared for submission to IEEE - Ultrasonics, Ferroelectrics, and Frequency Control (UFFC).

1.5 Other contributions

Conference Items

J. Kvam, S. Solberg, O. M. Brende, O. F. Myhre, A. Rodriguez-Molares, J. Kongro, B.A.J. Angelsen. "Tissue characterization with SURF imaging", Poster presentation at *The 8th national PhD Conference in Medical Imaging 2016*, 2016.

P.T. Yemane, B.A.J. Angelsen, **J.Kvam**, M. Afadzi, O. F. Myhre, C. de Lange Davies. "Simulation of ultrasound radiation force: for transport of drugs and non-particles in tumors." Poster presentation at *The 8th national PhD Conference in Medical Imaging 2016*, 2016.

J. Kvam, S. Solberg, O. M. Brende, O. F. Myhre, A. Rodriguez-Molares, J. Kongro, B.A.J. Angelsen. "Nonlinear elasticity imaging with dual frequency ultrasound", Presentation at the *Meeting of the Acoustical Society of America*, Boston, USA, May 2017.

E. Flørenæs, S. Solberg, **J. Kvam**, O. F. Myhre, O. M. Brende, B.A.J. Angelsen. "*In vitro* detection of microcalcifications using dual band ultrasound." In proceedings of *IEEE International Ultrasonics Symposium (IUS) 2017*, Washington DC, USA, 2017.

J. Kvam, S. Holm, B.A.J. Angelsen. "Exploiting Ballou's rule for better tissue classification", Presentation at the *Meeting of the Acoustical Society of America*, Minneapolis, USA, March 2018.

Bibliography

- [1] M. L. Oelze and J. Mamou. “Review of Quantitative Ultrasound: Envelope Statistics and Backscatter Coefficient Imaging and Contributions to Diagnostic Ultrasound.” *IEEE Transactions Ultrasonics, Ferroelectrics, and Frequency Control*, 63(2), **2016**. doi: 10.1109/TUFFC.2015.2513958.
- [2] J. Mamou and M. L. Oelze. *Quantitative ultrasound in soft tissues*. Springer, **2013**.
- [3] V. I. Shah, U. Raju, D. Chitale, et al. “False-negative core needle biopsies of the breast.” *Cancer*, 97(8):1824–1831, **2003**. doi: 10.1002/cncr.11278.
- [4] F. Rabbani, N. Stroumbakis, B. R. Kava, M. S. Cookson, and W. R. Fair. “Incidence and clinical significance of false-negative sextant prostate biopsies.” *The Journal of Urology*, 159(4):1247–1250, **1998**. doi: 10.1016/S0022-5347(01)63574-2.
- [5] D. W. Newcom, T. J. Baas, and J. F. Lampe. “Prediction of intramuscular fat percentage in live swine using real-time ultrasound.” *Journal of Animal Science*, 80(12):3046–3052, **2002**.
- [6] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” *Nature*, 521(7553): 436–444, **2015**.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” *Advances In Neural Information Processing Systems*, pages 1–9, **2012**. doi: <http://dx.doi.org/10.1016/j.protcy.2014.09.007>.
- [8] D. Shen, G. Wu, and H.-I. Suk. “Deep Learning in Medical Image Analysis.” *Annual review of biomedical engineering*, 19:221–248, **2017**. doi: 10.1146/annurev-bioeng-071516-044442.
- [9] G. Litjens, T. Kooi, B. E. Bejnordi, et al. “A survey on deep learning in medical image analysis.” *Medical Image Analysis*, 42:60–88, **2017**. doi: 10.1016/j.media.2017.07.005.
- [10] R. T. Beyer. *Nonlinear Acoustics*. U.S. Naval Sea Systems Command, **1974**.
- [11] M. F. Hamilton and D. T. Blackstock. *Nonlinear Acoustics*. Acoust Soc Am Press, New York, **2008 (originally published in 1998)**.
- [12] M. F. Hamilton and D. T. Blackstock. “On the coefficient of nonlinearity β in nonlinear acoustics.” *The Journal of the Acoustical Society of America*, 83 (1):74–77, **1988**. doi: 10.1121/1.396187.

-
- [13] X. Liu, J. Li, X. Gong, Z. Zhu, and D. Zhang. “Theoretical and experimental study of the third-order nonlinearity parameter C/A for biological media.” *Physica D: Nonlinear Phenomena*, 228(2):172–178, **2007**. doi: 10.1016/j.physd.2007.03.007.
- [14] X.-c. Xu, F. Mao, X.-f. Gong, and D. Zhang. “Theoretical calculation and experimental study on the third-order nonlinearity parameter C/A for organic liquids and biological fluids.” *The Journal of the Acoustical Society of America*, 113(3):1743–1748, **2003**. doi: 10.1121/1.1553460.
- [15] S. Holm, S. P. Näsholm, F. Prieur, and R. Sinkus. “Deriving fractional acoustic wave equations from mechanical and thermal constitutive equations.” *Computers & Mathematics with Applications*, 66(5):621–629, **2013**.
- [16] H. PA, F. Di Gennaro, N. E. Baumgartner C, et al. IT’IS database for thermal and electromagnetic parameters of biological tissues, May 2018.
- [17] J. A. Noble. “Ultrasound image segmentation and tissue characterization.” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):307–316, **2010**. doi: 10.1243/09544119JEIM604. PMID: 20349821.
- [18] H. Kim and T. Varghese. “Hybrid Spectral Domain Method for Attenuation Slope Estimation.” *Ultrasound in Medicine & Biology*, 34(11):1808 – 1819, **2008**. doi: <https://doi.org/10.1016/j.ultrasmedbio.2008.04.011>.
- [19] F. Duck. “Tissue non-linearity.” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):155–170, **2010**. doi: 10.1243/09544119JEIM574.
- [20] J. Zhang, M. S. Kuhlenschmidt, and F. Dunn. “Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A .” *The Journal of the Acoustical Society of America*, 89(1):80–91, **1991**. doi: 10.1121/1.400371.
- [21] D. Zhang and X.-F. Gong. “Experimental investigation of the acoustic nonlinearity parameter tomography for excised pathological biological tissues.” *Ultrasound in Medicine & Biology*, 25(4):593–599, **1999**. doi: 10.1016/S0301-5629(98)00185-9.
- [22] C. M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf. “Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers.” *Ultrasound in Medicine and Biology*, 12(11):865–874, **1986**. doi: 10.1016/0301-5629(86)90004-9.

- [23] X. Gong, D. Zhang, J. Liu, et al. “Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues.” *The Journal of the Acoustical Society of America*, 116(3):1819–1825, **2004**. doi: 10.1121/1.1781709.
- [24] M. Nikoonahad and D. C. Liu. “Pulse-Echo Single Frequency Acoustic Nonlinearity Parameter (B/A) Measurement.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 37(3):127–134, **1990**. doi: 10.1109/58.55301.
- [25] A. Kato and Y. Watanabe. “Measurement Method of Spatial Distribution of Nonlinearity Parameter B/A Using Nonlinear Interaction of Two Sound Waves.” *Japanese Journal of Applied Physics*, 32(Part 1, No. 5B):2274–2278, **1993**. doi: 10.1143/JJAP.32.2274.
- [26] N. Ichida, T. Sato, and M. Linzer. “Imaging the Nonlinear Ultrasonic Parameter of a Medium.” *Ultrasonic Imaging*, 5(4):295–299, **1983**. doi: 10.1016/S0161-7346(83)81002-5.
- [27] N. Ichida, T. Sato, H. Miwa, and K. Murakami. “Real-Time Nonlinear Parameter Tomography Using Impulsive Pumping Waves.” *IEEE Transactions on Sonics and Ultrasonics*, 31(6):635–641, **1984**. doi: 10.1109/T-SU.1984.31548.
- [28] H. Fukukita. “Ultrasound pulse reflection mode measurement of nonlinearity parameter B/A and attenuation coefficient.” *The Journal of the Acoustical Society of America*, 99(5):2775, **1996**. doi: 10.1121/1.414873.
- [29] F. Varray, M. Pasovic, C. Cachard, P. Tortoli, and O. Basset. “Acoustic nonlinearity parameter of tissue on echo mode: review and evaluation of the different approaches for B/A imaging.” In *Ultrasonics Symposium (IUS), 2009 IEEE International*, pages 41–44. IEEE, **2009**.
- [30] R. Hansen, S.-E. Måsoy, T. F. Johansen, and B. a. Angelsen. “Utilizing dual frequency band transmit pulse complexes in medical ultrasound imaging.” *The Journal of the Acoustical Society of America*, 127(1):579–87, **2010**. doi: 10.1121/1.3257584.
- [31] R. Hansen, S.-E. Måsoy, T. A. Tangen, and B. A. Angelsen. “Nonlinear propagation delay and pulse distortion resulting from dual frequency band transmit pulse complexes.” *The Journal of the Acoustical Society of America*, 129(2):1117–27, **2011**. doi: 10.1121/1.3518753.

- [32] T. Rommetveit, T. F. Johansen, J. Deibele, H. Kaupang, and B. Angelsen. “Two way nonlinear manipulation in plane materials using dual frequency pulse complexes.” *Proceedings - IEEE Ultrasonics Symposium*, pages 2380–2383, **2010**. doi: 10.1109/ULTSYM.2010.5935562.
- [33] S.-E. Måsøy, Ø. Standal, J. M. Deibele, et al. “Nonlinear propagation acoustics of dual-frequency wide-band excitation pulses in a focused ultrasound system.” *The Journal of the Acoustical Society of America*, 128(5):2695–703, **2010**. doi: 10.1121/1.3488308.
- [34] B. Angelsen and R. Hansen. “SURF Imaging - A New Method for Ultrasound Contrast Agent Imaging.” In *2007 IEEE Ultrasonics Symposium Proceedings*, pages 531–541. IEEE, **2007**. ISBN 978-1-4244-1383-6. doi: 10.1109/ULTSYM.2007.140.
- [35] S. E. Måsøy, Ø. Standal, P. Näsholm, et al. “SURF imaging: In Vivo demonstration of an ultrasound contrast agent detection technique.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 55(5):1112–1121, **2008**. doi: 10.1109/TUFFC.2008.763.
- [36] J. M. Rau, S.-E. Måsøy, R. Hansen, B. Angelsen, and T. A. Tangen. “Methods for reverberation suppression utilizing dual frequency band imaging.” *The Journal of the Acoustical Society of America*, 134(3):2313–25, **2013**. doi: 10.1121/1.4817900.
- [37] S. P. Näsholm, R. Hansen, S. E. Måsøy, T. F. Johansen, and B. A. J. Angelsen. “Transmit beams adapted to reverberation noise suppression using dual-frequency SURF imaging.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(10):2124–2133, **2009**. doi: 10.1109/TUFFC.2009.1295.
- [38] O. M. Brende and B. Angelsen. “Adaptive reverberation noise delay estimation for reverberation suppression in dual band ultrasound imaging.” *The Journal of the Acoustical Society of America*, 138(5):3341–51, **2015**. doi: 10.1121/1.4935555.
- [39] O. F. Myhre, T. F. Johansen, and B. A. Johan Angelsen. “Analysis of acoustic impedance matching in dual-band ultrasound transducers.” *The Journal of the Acoustical Society of America*, 141(2):1170–1179, **2017**. doi: 10.1121/1.4976096.
- [40] H. Høiland-Kaupang. *Models and methods for investigation of reverberations in nonlinear ultrasound imaging*. PhD thesis, Norwegian University of Science and Technology, The Medical Faculty, **2011**.

- [41] B. E. Treeby and B. T. Cox. “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields.” *Journal of Biomedical Optics*, 15(2):021314, **2010**. doi: 10.1117/1.3360308.
- [42] F. Rosenblatt. “The Perceptron: a probabilistic model for information storage and organization in the brain.” *Psychological Review*, 65(6):19–8, **1958**.
- [43] O. Russakovsky, J. Deng, H. Su, et al. “ImageNet Large Scale Visual Recognition Challenge.” *International Journal of Computer Vision (IJCV)*, 115(3): 211–252, **2015**. doi: 10.1007/s11263-015-0816-y.
- [44] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition.” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, **2016**.
- [45] N. P. Jouppi, C. Young, N. Patil, et al. “In-datacenter performance analysis of a tensor processing unit.” In *Computer Architecture (ISCA), 2017 ACM/IEEE 44th Annual International Symposium on*, pages 1–12. IEEE, **2017**.
- [46] An in-depth look at Google’s first Tensor Processing Unit (TPU). <https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>, 2017. Accessed: 2018-05-28.
- [47] H. Kung and C. E. Leiserson. “Systolic arrays (for VLSI).” In *Sparse Matrix Proceedings 1978*, volume 1, pages 256–282. Society for Industrial and Applied Mathematics, **1979**.
- [48] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization.” *Computing Research Repository (CoRR)*, abs/1412.6980, **2014**.
- [49] J. Duchi, E. Hazan, and Y. Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization.” Technical Report UCB/EECS-2010-24, EECS Department, University of California, Berkeley, **2010**.
- [50] N. Qian. “On the momentum term in gradient descent learning algorithms.” *Neural Networks*, 12(1):145–151, **1999**. doi: 10.1016/S0893-6080(98)00116-6.
- [51] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso. “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations.” In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10553 LNCS, pages 240–248, **2017**. ISBN 9783319675572. doi: 10.1007/978-3-319-67558-9_28.

- [52] W. R. Crum, O. Camara, and D. L. G. Hill. “Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis.” *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, **2006**. doi: 10.1109/TMI.2006.880587.
- [53] L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera, et al. “Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks.” In *International MICCAI Brainlesion Workshop*, pages 64–76. Springer, **2017**.
- [54] J. Deng, W. Dong, R. Socher, et al. “ImageNet: A Large-Scale Hierarchical Image Database.” In *CVPR09*, **2009**.
- [55] Y. Zhang, K. Lee, and H. Lee. “Augmenting supervised neural networks with unsupervised objectives for large-scale image classification.” In *International Conference on Machine Learning*, pages 612–621, **2016**.
- [56] A. Shrivastava, T. Pfister, O. Tuzel, et al. “Learning from Simulated and Unsupervised Images through Adversarial Training.” In *CVPR*, volume 2, page 5, **2017**.
- [57] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, et al. “Deep MR to CT Synthesis Using Unpaired Data.” In S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, editors, *Simulation and Synthesis in Medical Imaging*, pages 14–23, Cham, **2017**. Springer International Publishing. ISBN 978-3-319-68127-6.
- [58] A. Chatsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris. “Adversarial Image Synthesis for Unpaired Multi-modal Cardiac Data.” In S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, editors, *Simulation and Synthesis in Medical Imaging*, pages 3–13, Cham, **2017**. Springer International Publishing. ISBN 978-3-319-68127-6.
- [59] S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince. “Simulation and synthesis in medical imaging: Second international workshop, SASHIMI 2017 held in conjunction with MICCAI 2017 Québec city, QC, Canada, September 10, 2017 proceedings.” In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10557 LNCS, **2017**. ISBN 9783319681269. doi: 10.1007/978-3-319-68127-6.
- [60] X. Glorot, A. Bordes, and Y. Bengio. “Deep Sparse Rectifier Neural Networks.” In G. Gordon, D. Dunson, and M. Dudík, editors, *Proceedings of the*

Fourteenth International Conference on Artificial Intelligence and Statistics, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, **2011**. PMLR.

- [61] S. Ioffe and C. Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” *Computing Research Repository (CoRR)*, abs/1502.03167, **2015**.
- [62] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research*, 15:1929–1958, **2014**.
- [63] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler. “Efficient Object Localization Using Convolutional Networks.” *Computing Research Repository (CoRR)*, abs/1411.4280, **2014**.
- [64] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. “Deep Networks with Stochastic Depth.” *Computing Research Repository (CoRR)*, abs/1603.09382, **2016**.

Part II

Papers

Paper A

The use of deep learning to automate the segmentation of the skeleton from CT volumes of pigs

Johannes Kvam^{1,2}, Lars Erik Gangsei³, Jørgen Kongsro¹ and Anne H Schistad Solberg⁴

¹ Norsvin SA, Hamar Norway

² Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim Norway,

³ Animalia, Oslo, Norway,

⁴ Department of Informatics, University of Oslo, Oslo, Norway

Running title: Automated CT volume segmentation using Deep Learning

Published in Translational Animal Science, Volume 2, Issue 3, Pages 324-335, July 2018

Abstract – Computed tomography (CT) scanning of pigs has been shown to produce detailed phenotypes useful in pig breeding. Due to the large number of individuals scanned and corresponding large data sets, there is a need for automatic tools for analysis of these data sets. In this paper, the feasibility of deep learning for fully automatic segmentation of the skeleton of pigs from CT volumes is explored. To maximize performance, given the training data available, a series of problem simplifications are applied. The deep learning approach can replace our currently used semiautomatic solution, with increased robustness and little or no need for manual control. Accuracy was highly affected by training data, and expanding the training set can further increase performance making this approach especially promising.

A.1 Introduction

Segmentation of bodies using noninvasive imaging methods such as computed tomography (CT), ultrasound (UL), or magnetic resonance imaging (MRI) is of great importance in medicine, biology, and animal science. These methods have

given us the possibility to measure bodies, detect, diagnose, and treat diseases in a noninvasive way. The size and scale of such data sets during the last years have given rise to a demand for more automated interpretation and analysis, compared with interpretation performed by experts such as radiologists, researchers, doctors, or veterinarians. While there is still a gap between advances in medical imaging technologies and computational medical analysis, this gap has recently started narrowing with the help of machine learning techniques [1]. Machine learning provides an effective way to automate the analysis and diagnosis for medical images [2].

In pig breeding and genetics, CT has been used since the early eighties [3]. Single slice images from different locations were used in order to predict the *in vivo* body composition and fatness of the pig. Since then, developments toward whole body helical scanning have improved the accuracy in terms of body composition traits [4]. Helical scanning makes it possible to generate large amount of information, and as a result, large data sets of the object being scanned. In order to automate and sample more detailed phenotypes used in genetics, an atlas segmentation approach was introduced [5]. *In vivo* atlas segmentation of CT images, and associated methodology, has proven to be an effective method for registration of different parts of the pig, i.e., like the composition of meat cuts, meat quality, and diagnostics of diseases. Traits like relative size of and leanness of commercial cuts are heritable [6], thus atlas segmentation results are huge assets in a breeding perspective. Nordbø et al. [7] developed a number of traits describing the morphology of the shoulder blade. Shoulder blades were segmented [8] *in silico* from CT images of test boars. They found moderate-to-high heritability of the morphological traits which were also genetically correlated to shoulder lesions. The atlas segmentation is based on landmarks on the surface of the pig skeleton, which is segmented by applying a simple threshold at 200 Hounsfield units (HU). Landmarks are set for each of the larger skeletal structures. Until now, these structures have been identified by a version of the method specified in Gangsei and Kongso[8]. We refer to this method as the reference method throughout the paper. The reference method fails to do a correct segmentation in a substantial proportion of cases and require manual intervention and quality control in most cases. The need for a more automated method for segmenting bones in the pig skeleton is of great importance in order to apply atlas-based segmentation on a larger scale in a commercial breeding program of pigs. Machine learning, and in particular deep learning, has been shown to be a superior method for segmentation and classification of objects in medical images [9–11]. Cheng et al. [10] even argue that deep learning techniques might potentially change the design paradigm of the computer-aided diagnostic systems.

The aim of this study was to investigate the feasibility of deep learning as a

method for segmentation and classification of different parts of the skeleton in CT volumes of pigs.

A.2 Materials And Methods

Deep learning is a branch of machine learning that has been revitalized in recent years due to its performance in image analysis [12]. The term deep learning heirs from the fact that deep convolutional neural networks (CNNs) are trained to learn rich hierarchical feature sets [13]. These networks are composed of successive convolution operations which are insensitive to the spatial locality of the features meaning that the same features can be identified in multiple locations in images. The depth of the CNN determines the complexity of the features the network is able to recognize given enough training examples are available to tune the network.

To utilize deep learning, a large amount of training examples are needed in order to train all parts of the network. Exactly how much training data are needed is currently an unanswered research question. Elements such as diversity of features, network architecture, and overall problem complexity can vary greatly between both problems and solutions. For supervised learning [13], problems are usually either formulated as classification or regression problems. In classification problems, we categorize the content of images into a discrete set of predefined categories; e.g., determining whether class A or B is in the image. Segmentation is hence typically formulated as a classification problem. For regression problems, the CNN produces a continuous output; e.g., how many instances of class A are present. For 2D images, well-known deep learning methods are available, but for 3D segmentation, it is still a difficult task for machine learning or deep CNNs to segment structures from medical images due to several mutually affected challenges. These challenges include complicated anatomical environments in volumetric images, optimization difficulties of 3D networks, and inadequacy of training samples [14]. In the following sections, we will describe how we perform full-volume segmentation and classification of CT volumes from pigs. By using 2D projections and successively simplifying the images, we are able to utilize proven 2D deep-learning methods to produce 3D segmentations of large full body scans.

A.2.1 3D Through 2D Projections

In contrast to human medicine, the individuals being scanned are not well behaved. Due to the fact that these are farmed animals to be used for food shortly after CT scanning, the animals are not fixated, and sedation is used instead of anesthesia due to a shorter withdrawal time. This can cause limbs to be entangled in a variety of ways which affects image quality, making interpretation of the images a challenge. As we want to investigate the feasibility of a fully automated machine learning-driven atlas segmentation, we attempt to reduce the problem complexity

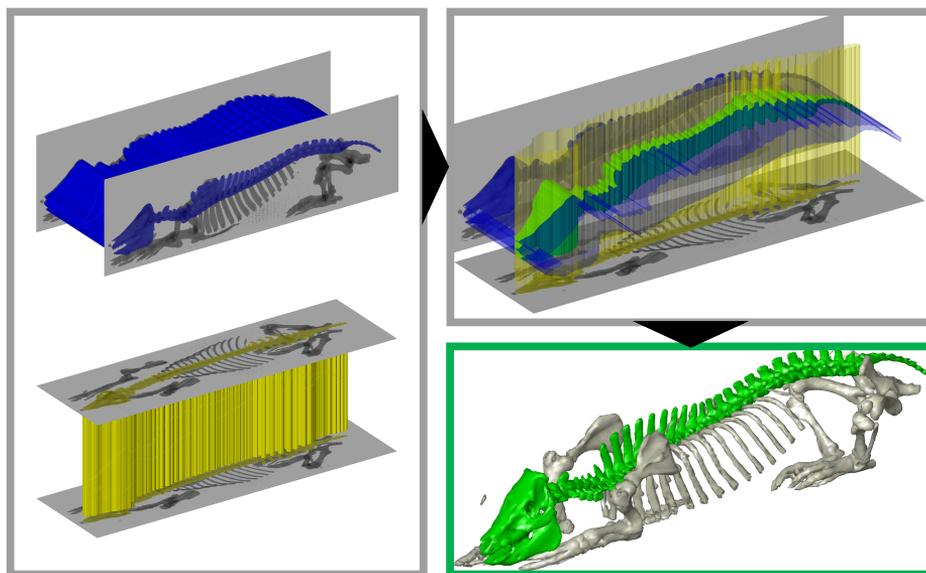


Figure A.1: Combination of sagittal and coronal projections to produce a 3D volume segmentation.

by splitting the overall problem into less confounded subtasks. Prior knowledge of pig anatomy and characteristics of *in vivo* CT scans of pigs is extensively utilized for creating a well adapted and suitable series of subproblems. The overall goal being to maximize the ratio of accuracy to the number of training samples available. As 3D CNNs are still a challenge, particularly in large volumes, we chose to reformulate the problem into a series of 2D problems combined to form a 3D volume. To produce full-volume segmentation, we rely on successive 2D segmentations in the coronal and sagittal images which are then combined to produce a 3D region of interest (ROI) [15]. This approach is illustrated in Fig. A.1. The input coronal and sagittal images are of size 512×1250 . Their pixel values are nonnegative integers reflecting the sum of bone voxels in the associated projection of the binary 3D input mask size $512 \times 512 \times 1250$.

The sagittal and coronal images possess different characteristics. For instance, the forelimbs and hindlimbs are in general recognizable in both sagittal and coronal images, especially if the spine is removed. However, left- and right-side limbs are indistinguishable in the sagittal image, also for the human eye, but distinguishable in the coronal image. As the 2D projections can be cluttered due to overlapping structures, the segmented regions are removed from the CT volume to produce less confounded coronal and sagittal input images for subsequent tasks.

A.2.2 Task Description

The final output of the segmentation algorithm is a skeleton which is segmented in 24 structures. Five of these structures; *cranium*, *cervical vertebrae*, *thoracic vertebrae*, *lumbar vertebrae* and *sacrum+coccyx*, constitutes what we will refer to as the spine, i.e. the major part of the axial skeleton. The remaining part of the axial skeleton contains three major structures; *sternum*, left and right side *costae*. The appendicular skeleton contains 16 major structures, i.e. left and right side of; *pelvis*, *femur*, *tibia+fibula*, *tarsal+metatarsal*, *scapula*, *humerus*, *radius+ulna* and *carpal+metacarpal*. Furthermore, the number of *thoracic vertebrae*, *lumbar vertebrae* and *costae*, which vary between pigs, are predicted.

The full process is divided into five consecutive and interdependent subproblems as shown in Fig. A.2. The first step is to segment and remove the spine from the CT volume. The composition of the spine is further classified by a separate CNN, while the remaining part of the skeleton is the input to a third network that identifies the limbs and *sternum*. The segmented limbs are removed from the volume, and the limbs on the left and right side are individually classified using a fourth network. Finally, the remaining skeleton consisting of the left and right side of the ribcage is classified by the fifth CNN.

In the following paragraphs, a more detailed description of each task is given. For each task, we train a separate CNN, with the same architecture described in section CNN Architecture. An overview of the input, output predictions, and training data for each CNN is given in Tab. A.1.

A.2.2.1 Segmentation Of The Spine

Based on the sagittal and coronal image of the full skeleton, the CNN was trained to segment one single mask containing the spine. By applying the principles of Sec. A.2.1, the spine was segmented from the remaining part of the skeleton. A binary mask separating the spine from the remaining skeleton constitutes the output from the spine-segmentation task. The centerline of the spine is utilized to partition the left- and right-hand side *costae*, see section Segmentation of limbs.

Spine segmentation is the initial task, thus its stability and precision crucial for the performance of subsequent tasks, which is heavily influenced by errors in the spine segmentation.

A.2.2.2 Classification Of The Spine

Using the spine mask from Sec. A.2.2.1 the coronal and sagittal images of isolated spine is used as input for classification of five well defined anatomical classes; *cranium*, *cervical vertebrae*, *thoracic vertebrae*, *lumbar vertebrae*, *sacrum+coccyx*. In addition a regression network is added for prediction of the number of *thoracic* and *lumbar vertebrae*. These can vary substantially between individuals opposed

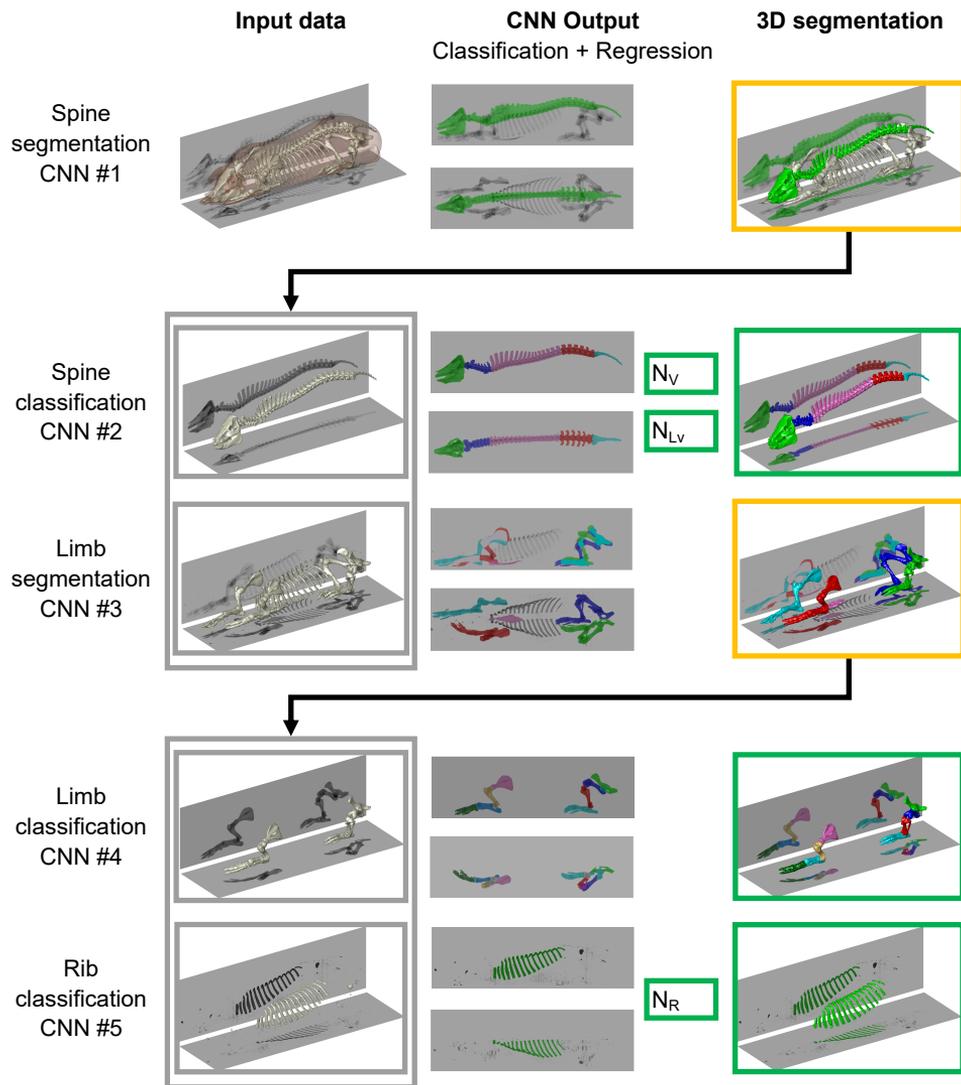


Figure A.2: Segmentations are done in a waterfall fashion. The segmented volume is removed from the CT volume to reduce complexity of input images to the following step. Each row represents a task which is handled by its own CNN. The CNNs operate on 2D projections of the CT volume (left column), and the output from the CNN (middle column) is combined to produce a segmentation in the 3D volume (right column). Green and yellow frames indicate final and intermediate results, respectively.

CNN	Input	Problem	Prediction	Training Data No. Individ.
#1 Spine Segmentation	Whole Skeleton	Classification	Spine	1735
#2 Spine classification	Spine	Classification Regression	<i>Cranium</i> <i>Cervical V.</i> <i>Thoractic V.</i> <i>Lumbar V.</i> <i>Sac. + Coccyx</i> Verterbras <i>Lumbar V.</i>	1000
#3 Limb segmentation	Skeleton without spine	Classification	L. forelimb R. forelimb L. hindlimb R. hindlimb <i>Sternum</i>	1000
#4 Limb classification	Isolated left and right limbs	Classification	<i>Pelvis</i> <i>Femur</i> <i>Tibia+Fibula</i> <i>Tarsal+M.tars</i> <i>Scapula</i> <i>Humerus</i> <i>Radius+Ulna</i> <i>Carpal+M.carp</i>	857
#5 Rib segmentation	Isolated ribcage	Classification Regression	<i>Costae</i> No. <i>costae</i>	500

Table A.1: Description of input, output predictions, and training data for each CNN.

to the number of *cervical vertebrae* which is approximately constant ($N = 7$)[16].

This CNN structure combining five anatomical classes and regression is simpler than incorporating one mask per vertebra, a structure with ≈ 33 classes. The information benefit from this simpler structure is close to equal to the more complex structure, due to the extra information from the regression part. All output, i.e. the five anatomical classes and vertebra numbers, are part of the final result.

A.2.2.3 Segmentation Of Limbs

The animals are not fixated and hence rarely outstretched while scanned, consequently identifying the individual bones in forelimbs and hindlimbs is challenging in the coronal view. Identification of individual bones is easier in the sagittal view. However, determining whether the bone in question belongs to either the left or right limb is not possible relying on the sagittal view alone. This is illustrated in Fig. A.3, where a t-distributed stochastic neighbor embedding approach is utilized to evaluate class separability in the coronal and sagittal views. In order to simplify the classification, we introduce an intermediate step by training an additional CNN which identifies the left and right limbs. This allows us to split the volume in a left and right side for which individual bones are classified independently by the subsequent operation described in section Sec. A.2.2.4.

The input for this task is the skeleton excluding the already segmented spine, see Sec. A.2.2.1. The CNN segments the *sternum*, which is part of the final result, and identifies the four main limbs, i.e., two forelimbs and two hindlimbs with a left and right side, respectively, which are used as input in subsequent tasks. Thus, five classes are implemented in the CNN, reflecting the *sternum* and the four limbs.

Based on the masks from the CNN, a series of operations were used to produce the final output. First, the *sternum* was segmented by the principles described in Sec. A.2.1. After removal of the *sternum*, 3D masks containing the left- and right-side limbs were constructed by the same methodology. As expected *a priori*, the performance of the CNN for distinguishing between the left and right side of the animal in the sagittal view was poor. Thus, the sagittal CNN output masks identifying left and right side were combined. The combined mask in the sagittal view, in addition to the side information from coronal view, was sufficient to segment the limbs and decide which side they belonged to. Finally, after removal of the limbs, the remaining skeleton, where the *costae* constitutes the bulk, was split in a left and right side by utilizing the centerlines of the *sternum* and spine.

Segmentation of limbs is a key task in the process as it yields both the final segmentation of the sternum and masks for the limbs used in the final two tasks.

A.2.2.4 Classification Of Limbs

The isolated left and right limbs, segmented in Sec. A.2.2.3, constituted the input for the limb classification task. The classification of limbs yields the final segment-

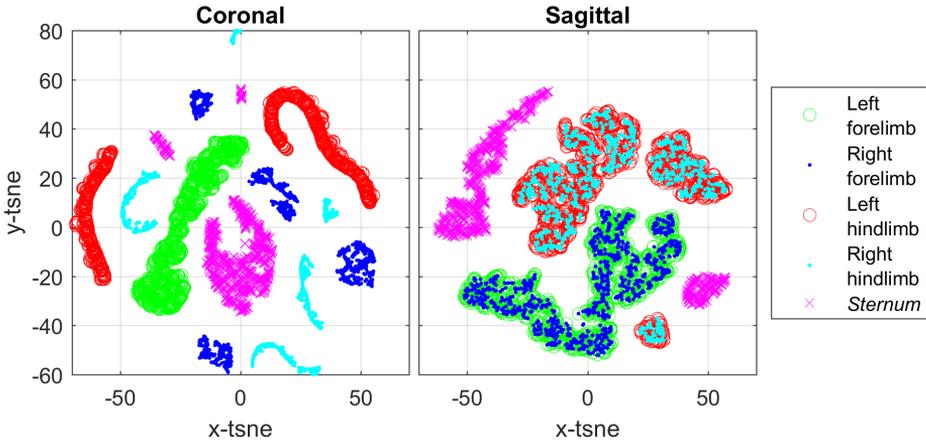


Figure A.3: Visualization of class separation using t-distributed stochastic neighbor embedding. The classes shown are for the limb-segmentation task. Left and right limbs are distinguishable in the coronal view but indistinguishable in the sagittal view.

ation for eight classes: *pelvis*, *femur*, *tibia + fibula*, *tarsal + metatarsal*, *scapula*, *humerus*, *radius + ulna* and *carpal + metacarpal* for the left- and right-hand side of the animal. Thus, making it the CNN with the highest number of output classes. When evaluating the performance of the network emphasis was allocated to the sagittal segmentations, as the eight classes in general have considerably smaller overlap in the sagittal compared with the coronal view.

A.2.2.5 Segmentation Of *Costae* (ribs)

Using the negated mask produced in limb segmentation, Sec. A.2.2.3, the remaining part of the skeleton consisting mainly of the ribcage is produced. The left and right side are represented as individual images and used as input for the *costae* segmentation task. The output from the CNN is a binary mask which segment *costae* from the background and smaller remaining objects with HU intensities in the bone range. Furthermore, as the number of *costae* vary between pigs, a regression network predicting this number is added. All output, i.e., the anatomical class and number of *costae*, are part of the final result.

A.2.3 CNN Architecture

As the images all arise from the same volumetric data set, the images and features are similar. The tasks involve mainly semantic segmentation and regression. For segmentation, an U-Net architecture was chosen, which is a CNN designed for semantic segmentation. The U-Net concatenates feature at different scales through downsampling and upsampling which has been shown to be beneficial in medical images [17]. As our training data are limited, we also implemented dropout at

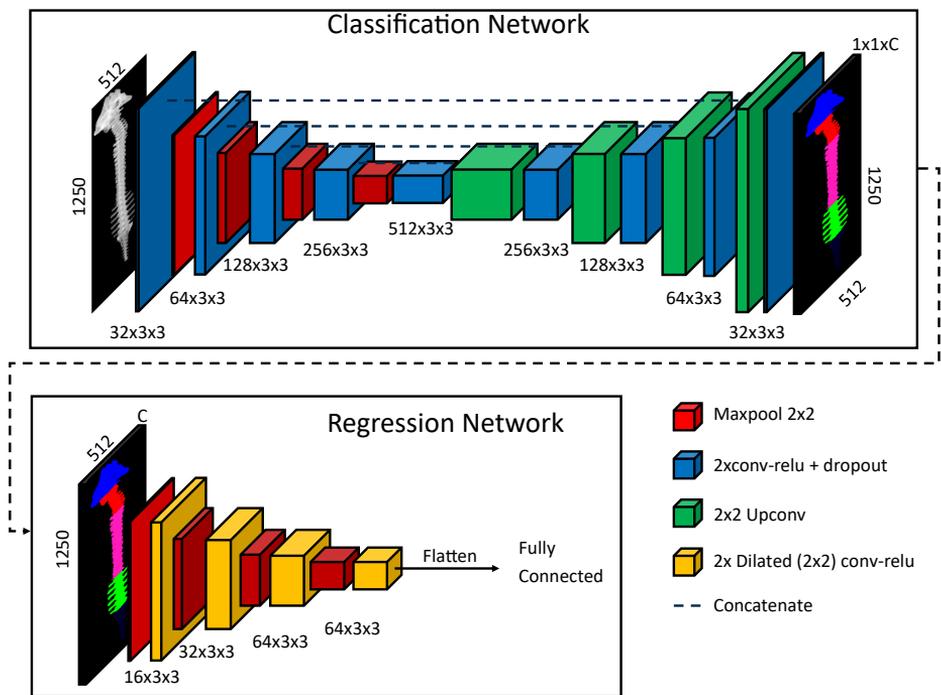


Figure A.4: CNN used for segmentation tasks. The network is based on the U-net [17] with added dropout. In addition, an optional regression head is added for segmentation tasks involving counting of class labels.

s

each of the scale representations to increase robustness and avoid overfitting [18]. In two of the tasks, we require in addition a continuous output describing how many instances of a particular class are present in the image. In the task of classifying the spine, we need to know the total number of vertebrae and the number of lumbar vertebrae as described in section Classification of the spine. We also need to count the number of costae as described in section Segmentation of costae (ribs). To accomplish these tasks, we add an additional regression network at the end of the classification network in these two cases. This regression network is a CNN designed by the authors consisting of a series of dilated convolutions [19] and max-pooling operations to increase field-of-view rapidly. Finally, the features are flattened, and a fully connected layer outputs the estimated number of class instances. An illustration of the architecture of the classification and regression network is shown in Fig. A.4. For each of the tasks described in Sec. A.2.2, we train a separate instance of this CNN with or without the regression network depending on the task.

A.2.3.1 Loss functions

Choosing the right loss function determines the efficiency of the training process. For semantic segmentation, categorical cross-entropy loss or dice losses [20] are typical loss formulations. The categorical cross-entropy loss maximizes the chance of a correct classification by minimizing the probability that it is mistaken for any other class. In the case of overlapping classes, as is present in CT images of entangled limbs, these special cases have to be unique classes to be correctly classified. For instance, if the left and right limb can overlap, there has to be a separate overlap class for categorical cross-entropy to be able to robustly classify it. As we are systematically attempting to reduce the number of classes due to limited training data, a dice loss is chosen as it works well for segmentation tasks and does not need explicit handling of overlap between classes. The dice loss measures ratio between the intersect and union of the predictions and ground truth labels: for multiple labels, the generalized dice loss (GDL) is described as follows:

$$GDL = \frac{\epsilon + 2 \sum_{l=1}^{N_l} w_l \sum_{i=1}^N y_{l,i} \hat{y}_{l,i}}{\epsilon + 2 \sum_{l=1}^{N_l} w_l \sum_{i=1}^N y_{l,i} + \hat{y}_{l,i}} \quad (\text{A.1})$$

where $y_{l,j}$ and $\hat{y}_{l,j}$ are the binary ground truth and predicted label of class l , respectively, and $\epsilon = 1$ is a smoothing term. To avoid class imbalances, each class is weighted by the class weights w_l . Compared with Sudre et al. [20], we have removed the bias term, and the peak classification score is found at $GDL = 1$. The class weights were chosen as $w_l = 1 / \left(\sum_{i=1}^N y_{l,i} \right)^2$ which causes the classes to be weighted equally. For binary classification tasks, e.g., classification of spine and ribs, the term cancels and we get the binary dice loss. An interesting variation of

the dice loss called the Wasserstein loss described by Fidon et al. [21] introduces a label distance matrix that punishes some misclassifications more than others. As there is consistency in the relative positioning of some labels in our data set, the Wasserstein loss would be beneficial to avoid crucial misclassifications such as mistaking left and right or front and back. However, due to the different perspective of the sagittal and coronal views described in Sec. A.2.1, the label distance matrix would have to take on different values in the two views. This would require non-differentiable operations in choosing the correct distance matrix given the particular view. For classification tasks, the generalized dice loss in (A.1) was chosen as the loss function.

For the regression tasks involved in the counting of *vertebrae* and *costae*, another loss is introduced to train this part of the network. A typical regression loss is the mean square error (MSE) given as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^{N_r} (r_i - \hat{r}_i)^2 \quad (\text{A.2})$$

where r_i and \hat{r}_i are the ground truth and predictions, respectively. This loss was chosen due to its simplicity and smooth derivative.

A.3 Experimental Setup

Purebred Duroc and Landrace boars from the boar testing station, Norsvin Delta in Norway, were CT scanned as part of the Topigs Norsvin commercial genetic program. The pigs were CT scanned using a GE Healthcare VCT 32 scanner at 120 kg BW. The protocol and setting used are described by [6]. All animals were cared for according to the laws and regulations for keeping pigs in Norway (Regulation for the keeping of 299 pigs in Norway 2003/02/18/175, 2003; Animal Welfare Act 2009/06/19/97, 300 2009).

A.3.1 Data Annotation

In order to annotate training data, results from the reference method were a major asset. The reference method is, as already mentioned, a heuristic, 3D based, and manually controlled technique. The reference method produces masks for all the 24 major bone structures described in Sec. A.2.2. Input images for CNNs were constructed for all five subtasks based on 3D bone structure masks. The bone structures in question were composed of the associated individual bones from the reference method.

For a CNN with C classes, the corresponding annotated coronal and sagittal masks had a size of $512 \times 1250 \times C$, i.e., C slices of binary masks, each representing one class. The segmentations from the reference method were used to construct these binary masks, where the pixel values in each of the C mask slices

were set to one if the pixel contained the actual projection of class in question. All input images and masks, in both training and test data, were manually corrected.

A.3.2 Training, Validation and Testing

As the annotation process for the different tasks had a varying degree of complexity, the training data for each task were different. For spine segmentation, a total of $3,470 = 1,735$ sagittal + $1,735$ coronal images from 1,735 individuals were used for training. For the other tasks, 2,000 images were used for training the networks. For the tasks of spine classification and limb segmentation, the 2,000 images were based on 1,000 individuals. For limb classification and *costae* segmentation, the number of individuals was approximately 500 as each individual has four associated images as the left and right side are separated.

After the first round of training, we reached performances in line with our requirements, except for the limb classification task (Sec. A.2.2.4). Thus, an additional set of training data, 1,428 images from 357 individuals, was added for a second round of training for this task. In addition, to mimic additional unique individuals, we introduced augmentation of the training data. Randomized deformations weights were applied to a uniform B-spline grid [22] for each input image and its corresponding labels. Out of the available samples used for training, 90% were used for updating the weights of the CNN. The remaining 10% was used as a separate validation set to monitor the training process and prevent overfitting. The final test set was composed of 500 previously unseen CT volumes.

A.3.2.1 Implementation

The networks were implemented in Python 3.5.2 (Python Software Foundation, <https://www.python.org/>) using Keras 2.1.1 [23] with the TensorFlow backend and trained on a NVIDIA GTX 1080 TI GPU using the Adam optimizer. Images were normalized to $\mu = 0$ and $\sigma = 1$. Due to the large size of the images (512×1250), the batch size was kept small due to memory restrictions, typically 3 – 6 images. The small batch size increases training time but with a potential benefit to accuracy due to the stochastic nature of the training process. For tasks involving both classification and regression, the classifier was trained first and kept constant when training the regression network. Training was stopped when the validation loss stopped improving. As an example, the training curves for spine classification task are shown in Fig. A.5.

A.3.3 Performance Evaluation

The different tasks done by the neural networks involved a combination of classification and regression tasks. For classification tasks, the mean dice score was

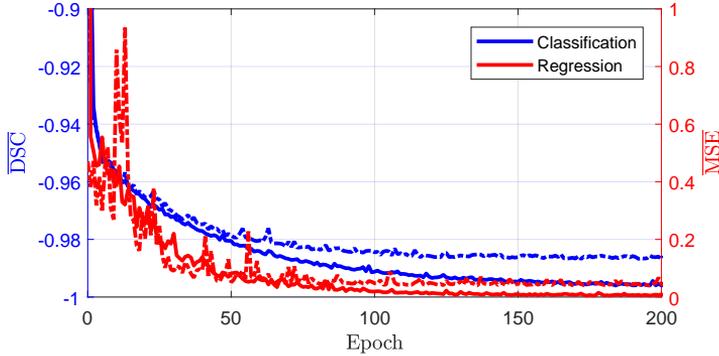


Figure A.5: Training curves for classification and regression for the spine classification task. Whole lines show the training loss, and dashed lines show the validation loss. An epoch is a measure of the number of times all the training examples have been used to update the weights of the network.

used to evaluate performance.

$$\overline{DS} = \frac{2 \sum_{l=1}^{N_l} N_l^{-1} \sum_{i=1}^N y_{l,i} \hat{y}_{l,i}}{\sum_{l=1}^{N_l} N_l^{-1} \sum_{i=1}^N y_{l,i} + \hat{y}_{l,i}} \quad (\text{A.3})$$

As the dice scores (DS) are not necessarily symmetrically distributed around their mean, we included accuracy as an additional measurement for the performance of the CNNs, identifying outliers, i.e., failed classifications. Accuracy was calculated by the number of individuals with $\overline{DS} > D_T$ relative to the population. The threshold was chosen through visual inspection, where $D_T = 0.95$ was chosen as it yielded satisfactory results in all segmentation tasks. For regression tasks, the performance was evaluated using the mean-square-error described in (A.2). For calculation of accuracy, the decimal output was first rounded to the closest integer. The accuracy was calculated as the proportion of individuals where there were no errors in the rounded prediction relative to the test population.

A.4 Results And Discussion

In this section, the performance of each individual task on the $N = 500$ test individuals is presented and discussed. Finally, we discuss some of the implications of the overall results.

Table A.2: Classification and regression results for spine classification.

Class	Sagittal		Coronal	
	\overline{DS}	Acc.	\overline{DS}	Acc.
<i>Cranium</i>	1.000	100.0%	1.000	99.8%
<i>Cervical vertebrae</i>	0.996	99.6%	1.000	99.8%
<i>Thoractic vertebrae</i>	0.994	94.2%	0.996	97.0%
<i>Lumbar vertebrae</i>	0.997	98.2%	0.996	97.4%
<i>Sacrum+Coccyx</i>	0.997	100.0%	0.998	99.8%
N	MSE	Acc.	\overline{MSE}	Acc.
Vertebras	0.045	97.4%	0.052	97.0%
<i>Lumbar</i> vertebras	0.074	93.2%	0.045	95.8%

A.4.1 Segmentation Of The Spine

As described in section Segmentation of the spine, the main purpose of the spine segmentation is to simplify the CT volume in the following steps by effectively dividing the volume into a left and right side. Out of all the tasks, this had the most training data and the lowest number of classes, which made us expect good performance. The mean dice score over the $N = 500$ test individuals was $\overline{DS} > 0.99$ both in the coronal and sagittal view. Manual control deemed all segmentations to be satisfactory.

A.4.2 Classification Of The Spine

In Tab. A.2, the classification and regression results are shown for classification of the spine. With the chosen categorization of vertebrae as described in section Classification of the spine, we achieved a $\geq 97\%$ accuracy on all labels in the coronal view and a $> 95\%$ accuracy in the sagittal view. Similarly, for regression tasks, there was a discrepancy in the performance in the coronal and sagittal view. Visually, it is easier to count the total number of vertebrae in the sagittal view as the transition between vertebrae is clearer. However, the *thoracic* and *lumbar* vertebrae are more distinguishable in the coronal view due to the more characteristic transverse process. As mentioned in section Classification of the spine, the number of *cervical* vertebrae is approximately constant, and the number of *thoracic* vertebrae is given as $N_{t,v} = N_v N_{l,v} 7$.

A.4.3 Segmentation Of Limbs

The performance of the limbs classifier is shown in Tab. A.3. As we can see, there is a drastic difference in the performance in the sagittal and coronal view. As discussed in Sec. A.2.2.3, this is expected as there is simply no information

Table A.3: Classification results for segmentation of limbs.

Class	Sagittal		Coronal	
	\overline{DS}	Acc.	\overline{DS}	Acc.
Left forelimb	0.843	4.0%	0.992	97.8%
Right forelimb	0.850	4.0%	0.993	98.2%
Left hindlimb	0.770	2.4%	0.990	96.8%
Right hindlimb	0.780	2.0%	0.990	95.8%
<i>Sternum</i>	0.967	84.9%	0.983	94.6%

in the sagittal 2D projection of the CT volume that allows distinguishing between left and right, this was further illustrated in Fig. A.3. This shows that the CNN is not able to distinguish what the human eye cannot. However, as also pointed out, producing a 3D segmentation can still be facilitated by combining left and right masks in the sagittal view. Through combination of classes in the sagittal view, the overall accuracy is $\sim 95\%$ which is satisfactory.

A.4.4 Classification Of Limbs

In the generation of training data as described in Sec. A.3.2, the generated inputs and outputs were manually controlled and cleaned of everything but the structures of interest. However, as this task is dependent on the performance of the spine segmentation and limb segmentation, CNN-compounded errors were present in the input data set. These errors manifested themselves in two ways. The first source of error was due to misclassifications or partial failures in the segmentations. Due to the high performance of the spine segmentation CNN, these types of errors mainly originated from the CNN performing limb segmentation. These failures limit the maximum achievable dice scores as small parts of the input data can be missing. This indicates that future efforts to improve performance should be focused on improving the limb-segmentation network. The other source of compounded errors, which proved to be the most challenging, was noise-like artefacts in the input image due to losses in the 3D segmentation and corresponding 2D projections. These types of artefacts were not present in the training data, and it became a major source for misclassifications. These artefacts could have been removed using traditional preprocessing image filtering techniques, but to make the CNN robust to these types of errors, we conducted another round of training with a training set that contained the same types of artefacts. The expanded training set increased the average dice coefficient, \overline{DS} , by between 0.05 and 0.3 which translated into approximately a corresponding increase in accuracy of 34% – 95%. The final classification scores are shown in Tab. A.4.

Table A.4: Classification results for classification of limbs.

Class	Sagittal		Coronal	
	\overline{DS}	Acc.	\overline{DS}	Acc.
<i>Pelvis</i>	0.970	94.4%	0.941	76.4%
<i>Femur</i>	0.981	97.0%	0.976	95.0%
<i>Tibia+Fibula</i>	0.966	92.6%	0.949	81.9%
<i>Tarsal+Metatarsal</i>	0.964	90.4%	0.950	86.4%
<i>Scapula</i>	0.988	98.4%	0.975	94.8%
<i>Humerus</i>	0.972	94.8%	0.967	91.8%
<i>Radius+Ulna</i>	0.957	83.3%	0.936	65.3%
<i>Carpal+Metacarpal</i>	0.948	77.8%	0.950	78.8%

The results confirm that identifying the individual bones is more challenging in the coronal view as discussed in Sec. A.2.2.4. In the sagittal view, the results were satisfactory for most classes; however, some manual correction of the *carpal+metacarpal* and *radius + ulna* might be necessary if deviating atlas segmentation results are encountered.

A.4.5 Segmentation Of Ribs

Segmentation of the ribs achieved acceptable performance with a $\overline{DS} > 0.99$ in both the coronal and sagittal views with an accuracy of $> 96\%$. Some compound errors as described in section Classification of Limbs were present; however, the main source of error was so-called half-ribs [24], underdeveloped ribs that are barely visible in the CT image. As they are hard to detect, we suspect that they may not have been consistently labeled in the training data; however, in the test set, they were consistently included. As a consequence, the performance of the regression network is perhaps underestimated as it is dependent upon the labels from the classifier which often fails to correctly classify these types of ribs. Consequently, the accuracy for the regression was 78.5% in the coronal and 89.4% in the sagittal view. Due to the classifiers problem with half-ribs, the bone was in most cases partially counted, identified as a fraction in the output. As these fractions are rounded to the closest integer, small variations can have a large impact on the accuracy calculation. The accuracy estimate is hence pessimistic, and the performance is deemed acceptable for the application.

A.4.6 Implications

The amount of CT data available from individual animals described in this paper is unique. The authors had access to several thousand animals (> 20.000) CT scanned in the period of 2012–2018 using the same CT-scanning protocol. Devel-

opment of methods in order to analyze and model these data is crucial not only for the animal science community but possibly also for the scientific community in general. The problem complexity reduction that has been applied has both upsides and downsides. A major upside was the relatively short training time and accuracy of each individual network given the training data available. This allowed us to evaluate parts of the segmentation process independently. Verifying annotations were also simpler as each label set was sparse and easily separable through visual inspection. However, there are two main downsides to this approach. Firstly, the splitting into subproblems causes a compounding of error due to classification errors in the preceding networks. This is especially apparent in the task of limb classification, where an additional training round was needed to make the network robust to these types of consequential errors. Luckily, the cost of training this robustness was minimal. The other downside is the inefficiency compared with a single network. Due to the current processing chain, complete CT volume segmentation is a sequentially dependent computation. This is not a critical element but more a question of elegance.

The computational bottleneck in our currently used semiautomatic solution is the need for manual input from an operator. This limits the throughput of the algorithm. The solution proposed here can, in its current state, be deployed for automatic full-volume landmark detection in atlas segmentation. Manual control for a small subset of individuals with deviating atlas segmentation results might still be needed. In addition, the controlled cases can be reintroduced in training which we saw had a significant impact on the performance on the limb classification task described in section Classification of Limbs. This process allows us to efficiently build up a large database of annotated volumes. Once a large collection of annotated data sets has been built up moving toward a single full-volume segmentation network is a natural next step. Especially, interesting is the combination of a convolutional and recurrent neural network (CNN + RNN) as described in [25, 26] for a slice-by-slice full-volume classification.

As CT is known to be robust for segmenting bone from soft tissue and the CNN input data are normalized, we believe that changes in the CT protocol will not have a significant effect on the results in this paper.

A.5 Conclusion

In this paper, the feasibility of fully automatic deep learning-driven segmentation of different parts of the pig skeleton from volumetric CT data was investigated. To maximize performance, given the training samples available, a series of steps were taken to simplify the problem. The final 2D-based solution can replace our currently utilized 3D-based method but more robust and with little or no need for manual intervention. In addition, accuracy was improved by introducing more

training data, confirming the feasibility of the approach.

Bibliography

- [1] D. Shen, G. Wu, and H.-I. Suk. “Deep Learning in Medical Image Analysis.” *Annual review of biomedical engineering*, 19:221–248, **2017**. doi: 10.1146/annurev-bioeng-071516-044442.
- [2] S. Wang and R. M. Summers. “Machine learning and radiology.” *Medical Image Analysis*, 16(5):933–951, **2012**. doi: 10.1016/j.media.2012.02.005.
- [3] H. Skjervold, K. Grønseth, O. Vangen, and A. Evensen. “In vivo estimation of body composition by computerized tomography.” *Z. Tierzüchtgsbiol.*, 98: 77–79, **1981**.
- [4] E. Gjerlaug-Enger, J. Kongsro, J. Odegård, L. Aass, and O. Vangen. “Genetic parameters between slaughter pig efficiency and growth rate of different body tissues estimated by computed tomography in live boars of Landrace and Duroc.” *Animal : an international journal of animal bioscience*, 6(1):9–18, **2012**. doi: 10.1017/S1751731111001455.
- [5] L. E. Gangsei, J. Kongsro, K. Olstad, E. Grindflek, and S. Sæbø. “Building an in vivo anatomical atlas to close the phenomic gap in animal breeding.” *Computers and Electronics in Agriculture*, 127:739–743, **2016**. doi: 10.1016/j.compag.2016.08.003.
- [6] J. Kongsro, L. E. Gangsei, T. M. Karlsson-Drangsholt, and E. Grindflek. “Genetic parameters of in vivo primal cuts and body composition (PigAtlas) in pigs measured by computed tomography (CT)1.” *Translational Animal Science*, 1(4):599–606, **2017**. doi: 10.2527/tas2017.0072.
- [7] Ø. Nordbø, L. E. Gangsei, T. Aasmundstad, E. Grindflek, and J. Kongsro. “The Genetic Correlation Between Scapula Shape and Shoulder Lesions in Sows.” *Journal of Animal Science*, page sky051, **2018**. doi: 10.1093/jas/sky051.
- [8] L. E. Gangsei and J. Kongsro. “Automatic segmentation of Computed Tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra’s algorithm.” *Computers and Electronics in Agriculture*, 121:191 – 194, **2016**. doi: <https://doi.org/10.1016/j.compag.2015.12.002>.
- [9] D. Kumar, A. Wong, and D. A. Clausi. “Lung Nodule Classification Using Deep Features in CT Images.” In *2015 12th Conference on Computer and Robot Vision*, pages 133–138, **2015**. ISBN VO -. doi: 10.1109/CRV.2015.25.

- [10] J.-Z. Cheng, D. Ni, Y.-H. Chou, et al. “Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans.” *Scientific Reports*, 6(1):24454, **2016**. doi: 10.1038/srep24454.
- [11] H. R. Roth, C. T. Lee, H. C. Shin, et al. “Anatomy-specific classification of medical images using deep convolutional nets.” In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 101–104, **2015**. ISBN 1945-7928 VO -. doi: 10.1109/ISBI.2015.7163826.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” *Advances In Neural Information Processing Systems*, pages 1–9, **2012**. doi: <http://dx.doi.org/10.1016/j.protcy.2014.09.007>.
- [13] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” *Nature*, 521(7553): 436–444, **2015**.
- [14] Q. Dou, L. Yu, H. Chen, et al. “3D deeply supervised network for automated segmentation of volumetric medical images.” *Medical Image Analysis*, 41: 40–54, **2017**. doi: 10.1016/j.media.2017.05.001.
- [15] W. N. Martin and J. K. Aggarwal. “Volumetric Descriptions of Objects from Multiple Views.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):150–158, **1983**. doi: 10.1109/TPAMI.1983.4767367.
- [16] J. W. B. King and R. C. Roberts. “Carcass length in the bacon pig; its association with vertebrae numbers and prediction from radiographs of the young pig.” *Animal Science*, 2(1):59–65, **1960**. doi: 10.1017/S0003356100033493.
- [17] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, **2015**. (available on arXiv:1505.04597 [cs.CV]).
- [18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research*, 15:1929–1958, **2014**.
- [19] F. Yu and V. Koltun. “Multi-Scale Context Aggregation by Dilated Convolutions.” *Computing Research Repository (CoRR)*, abs/1511.07122, **2015**.
- [20] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso. “Generalised dice overlap as a deep learning loss function for highly unbalanced

- segmentations.” In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10553 LNCS, pages 240–248, **2017**. ISBN 9783319675572. doi: 10.1007/978-3-319-67558-9_28.
- [21] L. Fidon, W. Li, L. C. Garcia-Peraza-Herrera, et al. “Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks.” In *International MICCAI Brainlesion Workshop*, pages 64–76. Springer, **2017**.
- [22] D. Rueckert, L. I. Sonoda, C. Hayes, et al. “Nonrigid registration using free-form deformations: application to breast MR images.” *IEEE transactions on medical imaging*, 18(8):712–721, **1999**. doi: 10.1109/42.796284.
- [23] F. Chollet. keras. <https://github.com/fchollet/keras>, 2015.
- [24] H. T. Fredeen and J. A. Newman. “Rib and vertebral numbers in swine.: I. Variation observed in a large population.” *Canadian Journal of Animal Science*, 42(2):232–239, **1962**. doi: 10.4141/cjas62-036.
- [25] J. Donahue, L. A. Hendricks, M. Rohrbach, et al. “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):677–691, **2017**. doi: 10.1109/TPAMI.2016.2599174.
- [26] P. Pinheiro and R. Collobert. “Recurrent convolutional neural networks for scene labeling.” In *International Conference on Machine Learning*, pages 82–90, **2014**.

Paper B

***In vivo* Prediction of Intramuscular Fat using Ultrasound and Deep Learning**

Johannes Kvam¹ and Jørgen Kongsro

Norsvin SA
Storhamar gata 44
Hamar, Norway

Running title: Prediction of IMF using Deep Learning

*Published in Computers and Electronics in Agriculture, Volume 142, Part B, Pages
521-523, November 2017*

Abstract – Intramuscular fat (IMF) in pigs determines the succulency and attractiveness of the meat. This paper presents a non-invasive *in vivo* method for estimating IMF using deep convolutional neural networks on ultrasound images. The method performs best on moderate to low IMF images < 6% giving a correlation of $R = 0.82$ and root-mean-square-error $RMSE = 1.2$. At higher IMF content the convolutional neural network fails to generalize due to image quality and lack of training data.

B.1 Introduction

Intramuscular fat (IMF), or marbling, is an important trait for meat due to its association with sensory traits such as juiciness and flavour.[1] Greater levels of IMF have been associated with approved consumer acceptance, and levels below 2.5% have been linked to poorer eating quality [2]. Ultrasound emerged as the tool for developing an objective measure for marbling due to the echo of fatty tissue, its non-invasiveness, ease of use and affordable cost level.[3]. The correlation between the IMF and subcutaneous fat has proven to be very high. Consequently, to correctly measure the trait IMF, we need to avoid the confounding between backfat and IMF in ultrasound images. For pig breeding, it is very important to

¹Also at: Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, P.O. Box 8905, N-7489 Trondheim, Norway.

avoid this confounding in order to differentiate between the subcutaneous fat level and the IMF level. Newcom et al. [2] stated that estimation of the IMF percentage in the live pig using real-time ultrasound is feasible, achieving moderate correlations between chemical IMF (CIMF) and predicted IMF with ultrasound, similar to those found by Ragland (1998). In a study by Lakshmanan [4], the authors found a high correlation and stated that the obtained prediction error of 0.34% IMF can be considered to be close to the possible optimum for predictions using such small regions of interest. However, it is likely that backfat was included in the model. In ultrasound images of beef, Nunes et al. [5] found a root-mean-square error (RMSE) of 1.31 and an R^2 of 0.37 using a cropped region of interest of the muscle. All the above studies used different approaches in terms of image analysis, from texture analysis [2] to acoustic spectral parameters by multivariate prediction [4].

Machine learning with neural networks has until recent years been limited in their application. To apply these techniques, domain expertise and engineering were required to manually convert the raw data to a set of features[6]. The neural network could then find the combination of features that produced the desired output. Brethour [7] showed how manually extracting image features and using simple neural networks could be used to estimate marbling from ultrasound images of cattle. The processing power available has increased drastically in the last ten years with the capability of the graphics processing unit (GPU). Combined with recent progress in representation learning, the features can be extracted automatically instead of being designed manually by a human engineer. Deep learning is a category of representation learning with a deep convolutional neural network (CNN) structure. The deep structure allows the network to learn complex patterns by combining lower level features hierarchically into higher ones. The network can then identify which features are the defining ones of the problem. The aim of this study was to test the predictive power of ultrasound on IMF using deep learning.

B.2 Materials & Methods

B.2.1 Data

Data was obtained from measurements performed in Norway, Spain and France at nucleus farm sites. In total, images from $N = 303$ breeding pigs (Duroc, Landrace and Large White) from Spain, Norway and France, were used. Each animal has on average 10 associated images. The total amount of images is 3037. The IMF range is 0.5 – 14%. In Fig. B.1, histograms of the available data are shown. The images were acquired using an Exago ultrasound device (<http://www.ecmscan.com/en/veterinary/exago-veterinary/>). Three different anatomical scan sites were used: 10th/11th rib called TP₁, last rib called TP₂, and at the last two

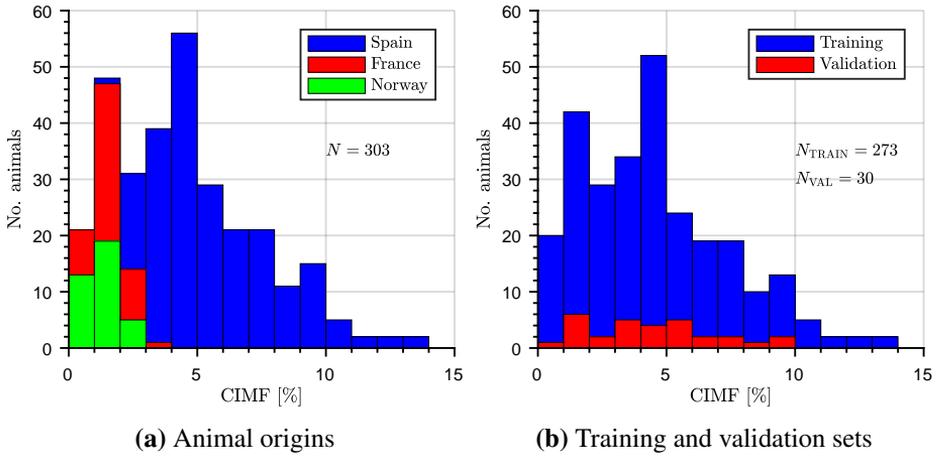


Figure B.1: Distribution of available data. The left panel shows the total distribution with origin of specimens. The right panel shows a typical training and validation set.

ribs called NO. Images were stored as grayscale .png files and imported into MATLAB. To avoid the confounding of IMF from backfat, the images were cropped to only contain a 100x300 pixel crop of the central part of the *M. Longissimus dorsi* muscle. The cropping was done automatically with a fixed position depending on the ultrasound measurement point for that image. All images were normalized to zero mean with unit standard deviation. CIMF content was measured using NIR spectroscopy (FOSS NIR- Systems, Hillerød, Denmark) in Norway or chemical extraction in France and Spain. The sample was taken from either the head of the loin (France and Spain) or at the last rib (Norway)[8].

B.2.2 Convolutional Neural Network

The features in pre-trained deep CNNs are usually complex and follow a hierarchical structure. The defining features of IMF in ultrasound images are in comparison simple structures and textures. It was therefore decided to try to build a custom CNN instead of employing transfer learning. However, training deep CNNs from scratch requires a large amount of data. Consequently, due to the limited amount of data available, the depth and width of our network was constrained to avoid overfitting. The final network architecture is shown in Fig. B.2. The network is composed of two large convolutions of size 11x11 and 5x5 for a rapid increase in field of view. Local response normalization (LRN) was used to normalize outputs. After the convolutions two inception modules[9] were used. A fully connected (FC) layer with dropout(DROP)[10] was used to further decrease the chance of overfitting before the final fully connected layer produced the IMF estimate. All the convolution layers used a padding to maintain spatial conformity and rectified

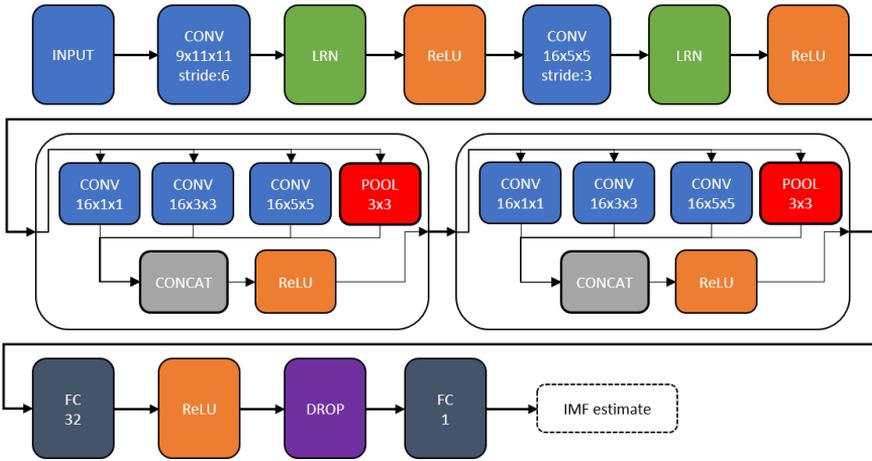


Figure B.2: Architecture of the CNN. The network is in total 16 layers deep with two inception blocks.

linear units (ReLU) were used as nonlinearities.

The network was implemented using Tensorflow ([tensorflow.org](https://www.tensorflow.org)) in Python and trained on an NVIDIA GTX 1080Ti graphics processing unit. A stochastic gradient descent optimizer (SGD) with momentum was used. To avoid overfitting, L2 regularization and random horizontal and vertical flipping of the input images were employed. In addition, dropout with $p = 0.5$ was used. The RMSE was used in the loss evaluation. To measure the generalization of the trained network, K-fold validation was used where the correlation coefficient and RMSE was averaged over the $K = 8$ different folds. An example of a training and validation set is shown in Fig. B.1(b). The network was trained until there was no further improvement in the loss function or the training loss diverged from the validation loss, typically around 100 epochs. The process was repeated for the 8 different folds in the K-fold validation scheme.

B.3 Results

In Fig. B.3 the evaluation on one of the validation sets in the K-fold validation scheme with $K = 8$ is shown. Estimates of IMF from the input images were averaged for each animal to obtain the final estimate. Each fold in the K-fold validation scheme had a mean of $\mu \approx 4.25$ and sample standard deviation of $\sigma \approx 2.4$. Averaged across the folds the correlation was found to be $\bar{R} = 0.74$ with $\overline{\text{RMSE}} = 1.8$. For the moderate to low IMF animals ($< 6\%$) the correlation was $\bar{R} = 0.82$ with $\overline{\text{RMSE}} = 1.2$.

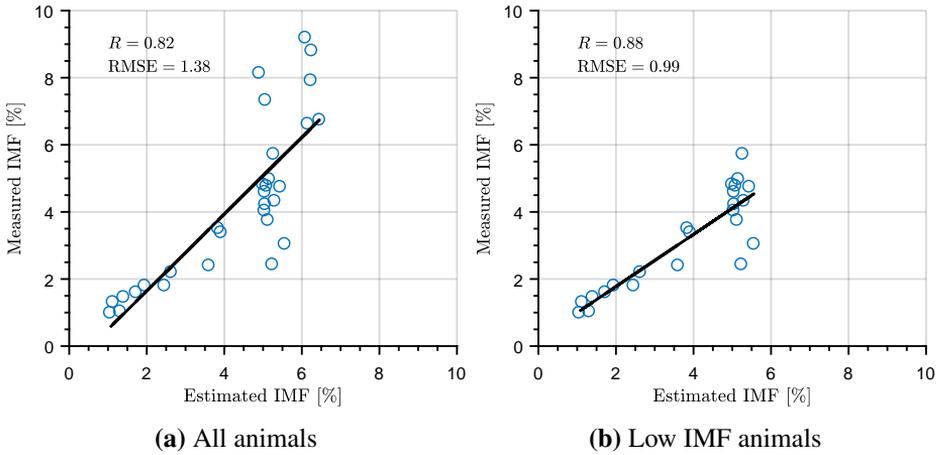


Figure B.3: Generalization of deep CNN on independent test set. The left panel shows the performance over the entire test set ($n = 30$ animals). The right panel shows the performance on the animals with a CIMF $< 6\%$, ($n = 23$ animals).

B.4 Discussion

The validation results show that the network performs well at mid to low IMF ($< 6\%$). However, for IMF levels $> 6\%$ the network fails to generalize producing a poor correlation. This indicates that there is insufficient or poor training data at higher IMF content.

Overall the results confirm that *in vivo* ultrasound estimation of intramuscular fat using deep learning is feasible compared to previous studies using ultrasound data or images from the real world. Comparing these results with other methods is challenging as the authors have not found a study with a similar IMF range. Lakshmanan[4] investigated ranges from 0 – 4% but only classifies them as either $< 1\%$, 1 – 2% or $> 2\%$. Nunes et al. [5] shows a performance similar to ours in cattle, but with an IMF range of 2 – 8%. Moreover, we see that also they get a clear deterioration in performance around 6% IMF, indicating a fundamental problem of image quality.

The images were obtained in a farm environment where optimization of image quality is very difficult. Image quality deteriorates with increased IMF content due to increased heterogeneity. Effects such as phase front aberrations [11, chapter 11], speckle[11, chapter 8.7] and multiple scattering[11, chapter 11] are a function of sample heterogeneity and affects the image quality of samples with high IMF more than low. Phase front aberrations cause a deterioration of imaging resolution, which can potentially become non-uniform. Speckle is due to the interference of two or more neighboring scatterers where their scattered signals are summed co-

herently or in-coherently. This effect produces random structures in the image that appear physical but is not a correct representation of the underlying anatomy. The effect of speckle on IMF estimation is hard to predict due to its random nature. Multiple scattering noise occurs as there multiple paths that produce the same time of flight and is hence mapped to the same region. Multiple scattering noise produces a cloud of noise that originate from scatterers at shallower depths in the image. All of these effects increase as the IMF content increases, deteriorating the image quality.

The CIMF reference is also a limiting factor due to the animal's physiology and the precision of the estimation method itself. The CIMF is measured at a single position (loin sample at the head of the loin or at last rib), whereas the ultrasound images are taken at multiple locations. The IMF in the different parts can be assumed to be highly correlated, but not uniform as some variance is to be expected. Imaging the exact same region is unrealistic. The method for estimation of CIMF also plays a role. In the data two main methods for estimating CIMF were used, NIRS and chemical extraction. However, the two methods have proven to be highly correlated and unbiased.

By expanding the training data and using better equipment for higher quality images, the network can be tuned to possibly give better results also at high IMF content.

B.5 Conclusion

We have shown the feasibility of using deep convolutional neural networks to estimate the intramuscular fat *in vivo* from ultrasound images. The method works best for the animals with a moderate or low amount of IMF (< 6%) giving a correlation of $\bar{R} = 0.82$ and $\overline{\text{RMSE}} = 1.2$. At higher IMF content the network fails to generalize due to decreased image quality and lack of training data.

Acknowledgments

The authors acknowledges support from the Norwegian Research Council, project number 254633. The authors would also like to thank Sigmund L.J. Rolfsjord for sharing his knowledge on machine learning.

Bibliography

- [1] N. R. Lambe, K. A. McLean, J. Gordon, et al. "Prediction of intramuscular fat content using CT scanning of packaged lamb cuts and relationships with meat eating quality." *Meat Science*, 123:112–119, **2017**.
- [2] D. W. Newcom, T. J. Baas, and J. F. Lampe. "Prediction of intramuscular fat percentage in live swine using real-time ultrasound." *Journal of Animal Science*, 80(12):3046–3052, **2002**.

-
- [3] K. D. Ragland. *Assessment of intramuscular fat, lean growth, and lean composition using real-time ultrasound*. Doctor of philosophy, Iowa State University, **1998**.
- [4] S. Lakshmanan, T. Koch, S. Brand, et al. “Prediction of the intramuscular fat content in loin muscle of pig carcasses by quantitative time-resolved ultrasound.” *Meat Science*, 90(1):216–225, **2012**.
- [5] J. L. Nunes, M. Piquerez, L. Pujadas, et al. “Beef quality parameters estimation using ultrasound and color images.” *BMC Bioinformatics*, 16(4):1–12, **2015**.
- [6] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” *Nature*, 521(7553):436–444, **2015**.
- [7] J. R. Brethour. “Estimating Marbling Score in Live Cattle From Ultrasound Images Using Pattern-Recognition and Neural-Network Procedures.” *Journal of Animal Science*, 72(6):1425–1432, **1994**.
- [8] J. Kongsro and E. Gjerlaug-Enger. “In vivo prediction of intramuscular fat in pigs using Computed Tomography.” *Open Journal of Animal Sciences (OJAS)*, 3(4):321–325, **2013**.
- [9] C. Szegedy, W. Liu, Y. Jia, et al. “Going Deeper With Convolutions.” In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, **2015**.
- [10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research*, 15:1929–1958, **2014**.
- [11] B. A. J. Angelsen. *Ultrasound Imaging - Waves, Signals and Signal Processing*. Emantec AS, Bugges veg 4B, 7051 Trondheim, Norway, **2000**.

Paper C

Exploiting Ballou's Rule for Improved Tissue Classification

Johannes Kvam^{1,2}, Sverre Holm³ and Bjørn A.J. Angelsen¹

¹ Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim Norway

² Norsvin SA, Hamar Norway

³ Department of Informatics, University of Oslo, Oslo, Norway

Running title: Relation between coefficient of nonlinearity and compressibility

Accepted for publication subject to minor revision in the Journal of Acoustical Society of America, December 2018

Abstract – Ultrasound tissue characterization based on the coefficient of nonlinearity, $\beta_n = 1 + B/2A$, has been demonstrated to produce added diagnostic value due to its large variation and sensitivity to tissue structure. However, the parameter has been observed to be significantly correlated to the speed of sound and density. These relationships are analyzed empirically as well as theoretically by developing a pressure-density relation based on a thermodynamic model and the Mie intermolecular potential. The results indicate that for many soft tissues the coefficient of nonlinearity is largely determined by the compressibility, κ_0 . Consequently, for tissue characterization, estimating the nonlinear response of the medium, given by $\beta_p = \beta_n \kappa_0$, appears to be beneficial due to correlated quantities.

C.1 Introduction

Healthy and pathological tissue has different mechanical characteristics which are observable in the acoustic parameters. Due to the low cost of ultrasound, tissue characterization based on this modality has the potential to provide substantial and affordable diagnostic value. Consequently, achieving robust tissue classification

techniques using ultrasound has been, and will likely continue to be, a key research topic.

There are multiple proposed methods for tissue characterization. Shear waves using acoustic radiation force have demonstrated a high characterization potential but is challenging in non-shallow applications as producing sufficient acoustic power to generate a shear wave deep in the body has proven difficult. For compression waves, methods include estimation of the sound speed and backscattering coefficients [1], acoustic absorption coefficients[2], and the nonlinearity parameter B/A [3]. Characterization based on the nonlinearity parameter has been especially promising as it has been shown to be highly sensitive to structural changes [4]. The variation in B/A between media also appears to be greater than the parameters associated with linear propagation such as speed of sound [5].

The parameter B/A describes the nonlinear wave propagation in relation to the linear wave propagation through a Taylor development of the pressure density relation. Consequently, the parameter of nonlinearity is dependent on the compressibility and density of the material. There have been indications that this dependence is significant as already in the 1960s a significant correlation between the nonlinearity parameter and the reciprocal sound speed was found in liquid metals [6]. Hamilton and Blackstock refer to this result as Ballou's rule after its discoverer [7]. Hartmann gave a more theoretical foundation for Ballou's result by calculating the sound speed in liquids using the Mie intermolecular potential function and relating it to the nonlinearity parameter[8]. Mast analyzed empirically the correlation between acoustic parameters in soft tissue and found a similar relationship between the reciprocal sound speed and B/A [9].

In this paper we analyze the correlation between the nonlinearity parameter and the compressibility. We will use both empirical modeling, and develop a thermodynamic model using the Mie intermolecular potential function, similarly to the work by Hartmann[8]. In contrast to Hartmann's analysis, we will be focusing on soft tissues, which can be assumed to be composed of basic constituents, water, protein and lipids[10]. We will analyze the potential specific parameters of the intermolecular potential and produce a model for the coefficient of nonlinearity in terms of the material's compressibility. We found a good agreement between empirical observation and our model, and discuss the potential consequences with respect to classification of soft tissues based on the nonlinear response of soft tissues.

C.2 Background

C.2.1 Nonlinearity parameter B/A

Soft tissues have nonlinear bulk elasticity, i.e. when the volume is compressed with positive pressure the material gets stiffer, and when the volume is expanded

with negative pressure the material gets softer. This phenomenon is described to the 2nd order by a Taylor expansion of the pressure-density relation,

$$p = p_0 + A \left(\frac{\rho - \rho_0}{\rho_0} \right) + \frac{B}{2} \left(\frac{\rho - \rho_0}{\rho_0} \right)^2. \quad (\text{C.1})$$

Here, p and p_0 are the instantaneous and hydrostatic pressures and, ρ and ρ_0 are the corresponding instantaneous and equilibrium densities. By developing the equation to higher orders, we can develop expressions for higher orders of nonlinearity. However, for typical ultrasonic pressures nonlinearity beyond second order are negligible and not considered here. The parameters B and A are given by,

$$B = \rho_0^2 \left(\frac{\partial^2 p}{\partial \rho^2} \right)_{0,s} \quad (\text{C.2})$$

$$A = \rho_0 \left(\frac{\partial p}{\partial \rho} \right)_{0,s} = \rho_0 c_0^2 = 1/\kappa_0, \quad (\text{C.3})$$

where the subscripts 0 and s describes at equilibrium density and constant entropy respectively. The compressibility is given by κ_0 , and is related to the speed of sound through,

$$c_0 = (\rho_0 \kappa_0)^{-1/2}. \quad (\text{C.4})$$

The adiabatic parameter of nonlinearity, B/A , can hence be formulated as,

$$\frac{B}{A} = B\kappa_0 = \rho_0^2 \left(\frac{\partial^2 p}{\partial \rho^2} \right)_{0,s} \kappa_0. \quad (\text{C.5})$$

Under isothermal conditions the parameter B/A is expressed slightly differently, however, for typical ultrasound applications, isentropic conditions apply. Therefore, in the analysis in this paper, we will be analyzing the adiabatic parameter of nonlinearity. In this case, the parameters, κ_0 , ρ_0 and c_0 are denoted κ_s , ρ_s and c_s to explicitly state isentropic conditions.

For an acoustic wave, the pressure-density relation in eq. (C.1), is often described in terms of the particle displacement, $\underline{\psi}$, [11]. In this case, under isentropic conditions, the expression takes the form,

$$p = p_0 - \frac{1}{\kappa_s} \nabla \underline{\psi} + \frac{\beta_n}{\kappa_s} (\nabla \underline{\psi})^2. \quad (\text{C.6})$$

See appendix C.6 for details. The coefficient of nonlinearity, β_n is related to the parameter of nonlinearity as,

$$\beta_n = 1 + \frac{B}{2A}, \quad (\text{C.7})$$

which is the typical way the parameter B/A appears in the wave equation. The unit term in eq. (C.7) arises from the convective acceleration in the Euler description or the continuity equation in the Lagrangian description[7].

As seen in eq. (C.5) the nonlinearity parameter, B/A , and consequently coefficient of nonlinearity β_n , is directly affected density, ρ_s , and compressibility, κ_s . Consequently, if the variation in B is low the nonlinearity parameter is largely determined by the variation in A . In these cases the parameter B/A will have a significant correlation with either parameter in $1/A = \kappa_s = 1/\rho_s c_s^2$, as was observed by Ballou, where a significant correlation was found between B/A and the reciprocal sound speed.

C.2.2 Tissue characterization based on β_n

Fundamentally, nonlinearity in acoustic wave propagation is due to a pressure dependent wave velocity[11], i.e.,

$$c(p) \approx c_0(1 + \beta_p p), \quad (\text{C.8})$$

where p is the acoustic pressure and β_p can be viewed as the observable nonlinear response of the medium, given as,

$$\beta_p = \beta_n \kappa_0 = (1 + B\kappa_0/2)\kappa_0, \quad (\text{C.9})$$

where β_n is the coefficient of nonlinearity related to the nonlinearity parameter through eq. (C.7). The pressure dependence in eq. (C.8) causes an accumulative distortion of the acoustic wave, as the peaks of the wave propagate faster than its troughs. The most well-known distortion is the generation of harmonics.

Tissue characterization methods based on nonlinear propagation, hence, attempts to quantify some manifestation of the nonlinear distortion and estimate the coefficient of nonlinearity, β_n , relating it to B/A through eq. (C.7). A key advantage in characterizing media based on the nonlinearity coefficient is its pressure dependence. In a back-scattered acoustic wave, the pressure has typically dropped significantly causing distortion to only be accumulated in the forward propagating wave. Consequently, characterization of the nonlinearity parameter does not require computed tomography setups and is theoretically more suited for pulse echo estimation. In practice however, producing robust echo modes has proven difficult. [3].

There are multiple proposed methods for estimating the coefficient of nonlinearity. Examples include the use of parametric arrays [12], estimation based on the second harmonic [13, 14] or using dual frequency techniques with widely separated bands [15–19].

In Varray[20], three typical β_n echo mode estimation approaches are discussed.

Common for these approaches is that they all solve expressions of the form

$$\beta_n = \rho_0 c_0^3 D_{\text{nl}}(p) = \frac{c_0}{\kappa_0} D_{\text{nl}}(p). \quad (\text{C.10})$$

where D_{nl} is the nonlinear pressure dependent distortion. As a consequence, these methods all require assumptions on the parameters c_0 and κ_0 .

C.3 Modeling of β_n in soft tissue

In the following sections we will present an empirical and a theoretical analysis, determining the variability of B and A in the coefficient of nonlinearity β_n . Emphasis will be given to soft tissues, but a similar analysis is applicable to other materials.

C.3.1 Regression model

To analyze how B and A influence the coefficient of nonlinearity β_n we will use empirical data for several soft tissues and use the relationships in eqs. (C.2) and (C.3) to approximate them. Using linear regression fitting we can analyze their individual effect on B/A .

In table C.1 a list of materials and their acoustic parameters is given. The values are taken from Mast [9] where all materials with valid B/A values are included. The table gives the values in terms of the coefficient of nonlinearity $\beta_n = 1 + B/2A$. We assume that the parameters are determined under isentropic conditions, i.e. $(\rho_0, c_0, \beta_n) = (\rho_s, c_s, \beta_{n,s})$.

With access to the density, speed of sound and B/A we can calculate,

$$\kappa_s = \frac{1}{\rho_s c_s^2} = \frac{1}{A}, \quad (\text{C.11})$$

and based on the expression in eq. (C.2) we obtain,

$$B = \left(\frac{B}{A} \right) \frac{1}{\kappa_s}. \quad (\text{C.12})$$

In fig. C.1 the variation in the parameter β_n , κ_s , β_p and B , with respect to their individual local minima, is shown for the materials in table C.1. We observe that the measurable nonlinear response, β_p , has a higher variation than κ_s and β_n individually for several soft tissues. This indicates that the parameters B and κ_s are either correlated or that the variation in $\beta_n = 1 + B\kappa_s/2$ is to a larger degree determined by κ_s than B .

To analyze how much B and A vary between tissue types, we formulate two different linear models of $\beta_n = 1 + B/2A$. In one model we allow $A = 1/\kappa_s$

Table C.1: Materials used for evaluation, sorted by increasing coefficient of nonlinearity, β_n . Data taken from [9] where materials with a valid entry for β_n is included.

	Name	ρ kg/m ³	c_0 m/s ²	κ_s^1 GPa ⁻¹	β_n –	B^2 GPa
1	Blood	1060	1584	0.376	4.05	16.2
2	Liver	1060	1595	0.371	4.3	17.8
3	Muscle, skeletal	1050	1580	0.382	4.3	17.3
4	Non-fatty	1055	1575	0.382	4.5	18.3
5	Brain	1040	1560	0.395	4.55	18
6	Muscle, cardiac	1060	1576	0.38	4.55	18.7
7	Kidney	1050	1560	0.391	4.7	18.9
8	Spleen	1054	1567	0.386	4.9	20.2
9	Skin	1090	1615	0.352	4.95	22.5
10	Fatty	985	1465	0.473	5.25	18
11	Adipose	950	1450	0.501	6	20

¹ Calculated from eq. (C.11).

² Calculated from eq. (C.12).

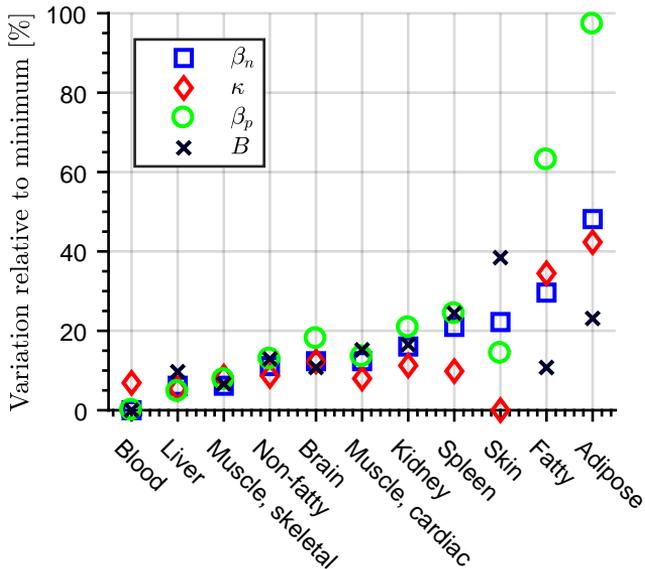


Figure C.1: Variation of parameters β_n , κ_s , β_p and B relative to their respective minima. Sorted by increasing β_n .

to vary and in the other we allow $B = \rho_s^2 (\partial^2 p / \partial \rho^2)_{0,s}$ to vary. This can be expressed as,

$$\hat{\beta}_n(\kappa) = a_\kappa \kappa_s + b_\kappa, \quad (\text{C.13a})$$

$$\hat{\beta}_n(B) = a_B B + b_B, \quad (\text{C.13b})$$

where the coefficients a_κ, b_κ and a_B, b_B are calculated as the least squares fit. Note that we have $2a_B \sim \bar{\kappa}_s$ and $2a_\kappa \sim \bar{B}$ from eq. (C.5), where $(\bar{\kappa}_s, \bar{B})$ are the averages of the dataset. Similarly, for the measurable nonlinear response of the medium, β_p , we formulate corresponding models as,

$$\hat{\beta}_p(\kappa) = a_{\kappa,2} \kappa_s^2 + a_{\kappa,1} \kappa_s + b_\kappa, \quad (\text{C.14a})$$

$$\hat{\beta}_p(B) = 2a_B(a_B B + 1) + b_B, \quad (\text{C.14b})$$

Note that the model in κ_s is of second order, while the model in B is of first order with $2a_B \sim \kappa$.

The performance of each model can then be calculated through the absolute relative error,

$$e_{\kappa,B} = \frac{|y - \hat{y}_{\kappa,B}|}{y}, \quad (\text{C.15})$$

where y is either β_n or β_p and $\hat{y}_{\kappa,B}$ are the corresponding models in eqs. (C.13a), (C.13b), (C.14a) and (C.14b). Taking the mean value of eq. (C.15) and multiplying by 100 yields the mean-absolute-percentage-error (MAPE). The interpretation of this is that if $e_\kappa < e_B$ the nonlinearity of the material is more determined by the compressibility of the material than the nonlinear factor B . On the other hand, if the inverse is true, we have $e_B < e_\kappa$ and the nonlinearity is to a larger degree determined by B .

C.3.2 Thermodynamic Model

Inspired by Hartmann's analysis[8], we will develop a thermodynamic model for the acoustic compression using intermolecular potentials. Emphasis is given to soft tissues, focusing on the relation between the compressibility and the coefficient of nonlinearity, β_n .

From the fundamental thermodynamic relations, Huang [21] showed that the general state equation for pressure, P , as a function of volume, V , and temperature, T , can be written as,

$$P(V, T) = - \left(\frac{\partial U}{\partial V} \right)_T + T \left(\frac{\partial S}{\partial V} \right)_T, \quad (\text{C.16})$$

where $U = U(V, T)$ is the internal energy of the material and S is the entropy. The change in entropy with respect to volume is given by,

$$\left(\frac{\partial S}{\partial V} \right)_T = \frac{\alpha}{\kappa_T}, \quad (\text{C.17})$$

where α is the isobaric thermal expansion coefficient and κ_T is the isothermal compressibility. We assume that the internal energy term $U(V, T)$ can be split into separate terms for its variation in V and T as,

$$U(V, T) = U_p(V) + U_k(T), \quad (\text{C.18})$$

where the subscript p and k denote the potential and kinetic internal energy respectively. This allows us to approximate,

$$\left(\frac{\partial U}{\partial V} \right)_T \approx \frac{\partial U_p}{\partial V}. \quad (\text{C.19})$$

At rest soft tissues are in a state where the temperature corresponds to body temperature, $T_0 = 310$ K. Correspondingly, the operational volume is given by V_T . As acoustic compression yields a small perturbation in volume, we perform a Taylor expansion of eq. (C.16), both in temperature and volume around this operating point (T_0, V_T) . As we are interested in the pressure up to second order we will expand up to second order in volume. For temperature, a first order expansion will suffice. The general state equation thus becomes,

$$\begin{aligned} P(V, T) &= P(V_T, T_0) + P^V(V_T, T_0)(V - V_T) \\ &+ \frac{1}{2} P^{2V}(V_T, T_0)(V - V_T)^2 \\ &+ P^T(V_T, T_0)(T - T_0) \end{aligned} \quad (\text{C.20})$$

where the superscript V and $2V$ indicate first and second order differentiation with respect to volume. The superscript T indicates differentiation with respect to temperature. The relative volume change is related to the particle displacement, $\underline{\psi}$, as,

$$\frac{V - V_T}{V_T} = \frac{\partial V}{\Delta V} = \nabla \underline{\psi}. \quad (\text{C.21})$$

In addition, the pressure can be expressed in terms of an environmental pressure and the acoustic pressure as,

$$P = p_a + p_e. \quad (\text{C.22})$$

The environmental pressure, $p_e \approx 101$ kPa, is a bias term independent of the volumetric and thermal changes. As the thermal pressure, $\alpha T / \kappa_T > 200$ MPa $\gg p_e$, neglecting the environmental pressure is reasonable. Utilizing these relations and

inserting for the potential in eq. (C.21) allows us to define,

$$p_e = -U_p^V(V_T) + S^V(V_T)T_0 \approx 0 \quad (\text{C.23a})$$

$$\begin{aligned} p_a &= V_T [-U_p^{2V}(V_T) + S^{2V}(V_T)T_0] \nabla \underline{\psi} \\ &+ \frac{V_T^2}{2} [-U_p^{3V}(V_T) + S^{3V}(V_T)T_0] (\nabla \underline{\psi})^2 \\ &+ \frac{\alpha}{\kappa_T} (T - T_0). \end{aligned} \quad (\text{C.23b})$$

Identifying with eq. (C.6) we get,

$$p_a = -\frac{1}{\kappa_T} \nabla \underline{\psi} + \frac{\beta_n}{\kappa_s} (\nabla \underline{\psi})^2 + \frac{\alpha}{\kappa_T} (T - T_0), \quad (\text{C.24})$$

note that the linear compression is given in terms of the isothermal compressibility due to the isothermal condition in eq. (C.16). Going from isothermal to isentropic is introduced through the last term in eq. (C.24), describing the thermal effects due to the acoustic compression. The term can be split into,

$$T - T_0 = \delta T_s - (\delta T_s - \delta T), \quad (\text{C.25})$$

where the first term, δT_s , corresponds to isentropic compression, i.e. $dS = 0$, where no energy is lost from translational degrees of freedom to the internal degrees of freedom. Huang [21] showed that this could be expressed as,

$$\delta T_s = -\frac{\alpha T}{C_V \kappa_T} \nabla \underline{\psi}, \quad (\text{C.26})$$

where C_V is the isochoric heat capacity per volume.

The second term in eq. (C.25) represents energy lost to the internal degrees of freedom, introducing acoustic absorption. Modeling absorption is a separate topic. It is typically described by a set of differential equations, with a solution that can be described as a convolution with a loss operator, h_{ab} , as,

$$\frac{\alpha}{\kappa_T} (\delta T_s - \delta T) = h_{ab} *_t \nabla \underline{\psi}, \quad (\text{C.27})$$

where the convolution in time with h_{ab} represents frequency dependent acoustic absorption. Inserting into eq. (C.24) we obtain,

$$p_a = -\frac{1}{\kappa_s} \nabla \underline{\psi} + \frac{\beta_n}{\kappa_s} (\nabla \underline{\psi})^2 - h_{ab} *_t \nabla \underline{\psi}. \quad (\text{C.28})$$

Where, through the relation in eq. (C.26), we have,

$$\frac{1}{\kappa_s} = \frac{1}{\kappa_T} \left(1 + \frac{\alpha^2 T}{C_V \kappa_T} \right) = \frac{\gamma}{\kappa_T}. \quad (\text{C.29})$$

Equation (C.28) is now the same as eq. (C.6) with the exception of the acoustic absorption term. Finding the final model, we hence need to model the derivatives of U_p and S .

The expression for the entropy in eq. (C.17) has no explicit volume dependence although both compressibility and thermal expansion coefficient are volume dependent. To introduce an explicit volume dependence, we rely on kinetic theory which gives us,

$$\left(\frac{\partial S}{\partial V}\right)_T = \frac{\alpha}{\kappa_T} = \frac{N_m R}{V}, \quad (\text{C.30})$$

where N_m is the number of moles in the volume and R is the ideal gas constant. This leads to,

$$\begin{aligned} V \left(\frac{\partial^2 S}{\partial V^2}\right)_T &= -\frac{\alpha}{\kappa_T}, \\ V^2 \left(\frac{\partial^3 S}{\partial V^3}\right)_T &= 2\frac{\alpha}{\kappa_T}. \end{aligned} \quad (\text{C.31})$$

Hartmann[8] showed that the the potential energy $U_p(V)$ can be modeled using the Mie intermolecular potential as,

$$U_p(V) = \frac{U_0}{n-m} \left[m \left(\frac{V}{V_0}\right)^{n/3} - n \left(\frac{V}{V_0}\right)^{m/3} \right], \quad (\text{C.32})$$

where U_0 is the depth of the potential energy well and V_0 is the equilibrium volume at zero temperature and pressure. The exponents m and n describe the characteristics of the long range and short range forces respectively. Consequently, these are dependent upon the material type. The attractive term is typically modeled as van der Waals force or dispersion force yielding a value of $m = 6$. The repulsive term describes Pauli repulsion, however, the exponent has no clear theoretical justification and is chosen empirically, where $n = 12$ is a typical result. For $m = 6$ and $n = 12$ the potential reduces to the Lennard-Jones potential. In the analysis by Hartmann, $m = 6$ and $n = 9$ was chosen as fixed for a large range of materials with a variation in speed of sound of 700 – 2500 m/s.

By differentiating eqs. (C.32) and (C.30), inserting into eqs. (C.23b) and (C.23a), identifying with eq. (C.24), and further utilizing the relationship,

$$\frac{\alpha T_0}{\kappa_T} \approx \frac{nm}{3(n-m)} \frac{U_0}{V_0} \frac{\eta^{(n-m)/3} - 1}{\eta^{n/3+1}}, \quad (\text{C.33})$$

we obtain the following relations,

$$\kappa_T = \gamma\kappa_s = 9\kappa_p \frac{n-m}{nm} \frac{\eta^{n+1}}{n-m\eta^{n-m}} \quad (\text{C.34a})$$

$$\beta_n\gamma = \frac{3 + n/3 + m/3 + nm\alpha T_0/9}{2} \quad (\text{C.34b})$$

where $\kappa_p = V_0/U_0$ is a potential specific compressibility and $\eta = V_T/V_0$ is the relative position of the operating volume at body temperature, related to the thermal expansion coefficient, through eq.(C.33), as,

$$\eta = \left(\frac{3 + n\alpha T_0}{3 + m\alpha T_0} \right)^{3/(n-m)}. \quad (\text{C.35})$$

Taking fat at body temperature, $T_0 = 310$ K, as an example, we have $\alpha \approx 0.001 \text{ K}^{-1}$ [22], setting $m = 6$ and $n = 12$ we obtain $\beta_n = 5.74$ which is close to that of adipose tissue in table C.1.

Equation (C.34b) corresponds to the result by Hartmann[8] where he noted that the coefficient of nonlinearity is given by the thermal expansion coefficient for a given m and n through its effect on the volume. At constant volume, $dV = 0$, the thermal expansion coefficient, α , is inherently tied to the compressibility as,

$$\begin{aligned} \alpha &= \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_p = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T \left(\frac{\partial p}{\partial T} \right)_V \\ &= \kappa_T \frac{N_m}{V} R, \end{aligned} \quad (\text{C.36})$$

using the kinetic theory in eq. (C.30). As soft tissues can be modeled as a composition of the basic constituents, water, protein and lipids, a reasonable assumption is that the number of moles in the volume, N_m/V , is close to constant. Most soft tissues are composed of $\sim 70\%$ water with added molecules composed of carbon, nitrogen and oxygen which have similar atomic numbers. Consequently, we expect the thermal expansion coefficient and compressibility to be highly correlated and have an approximately linear relationships in soft tissues.

In soft tissues we can approximate $\gamma \approx 1$ [8]. In this case eq.(C.34a) describes the compressibility under isentropic conditions. To investigate the correlation between κ_s and α we need to find a combination of parameters (n, m, κ_p, α) that satisfies eqs. (C.34a) and (C.34b). We only have two equations, making the system under-determined, however, by making two assumptions we can reduce the degrees of freedom.

1. The exponent, $m = 6$, is fixed. In the Mie potential, m , characterizes the long-range attractive force which can be approximated by a van der Waals

force. The exponent, n , describes the short-range repulsive force, does not have a clear justification and is often empirically determined.

2. The variation in κ_p is small. As previously mentioned, soft tissues are essentially compositions of the same basic constituents water ($\sim 70\%$), protein and lipids, composed of atoms of similar atomic numbers. Therefore, a reasonable approximation is a constant ratio U_0/V_0 is constant.

Using these assumptions, we have a system of equations where the variation in β_n and κ_s is described by variations in α and n . However, it is important to note that the parameters α, κ_p, n are not orthogonal as they have similar effects on the potential. Hence, incorrect assumptions will lead to compensation in the the free parameters α and n .

To find values of α and n for the materials in table C.1, we formulate the optimization problem,

$$\begin{aligned} \min_{\alpha, n, \kappa_p} \quad & \left(\frac{\hat{\kappa}_s - \kappa_s}{\kappa_s} \right)^2 + \left(\frac{\hat{\beta}_n - \beta_n}{\beta_n} \right)^2 + \lambda K_p^2 \quad (C.37) \\ \text{subject to:} \quad & \eta > 1, n > m, \kappa_p > 0, \end{aligned}$$

where $\hat{\kappa}_s$ and $\hat{\beta}_n$ is the modeled compressibility and coefficient of nonlinearity given in eqs. (C.34a) and (C.34b) respectively. The last term, $K_p = (\kappa_p - \bar{\kappa}_p)/\bar{\kappa}_p$, is due to our second assumption, where we favor solutions where κ_p varies little from its mean value $\bar{\kappa}_p$.

For β_n to be largely given by κ_s the consequence is that both α and n must be highly correlated with κ_s . This means that a model for β_n based on the thermodynamic model can take the form,

$$\beta_n = \frac{5 + n(\kappa_s)/3 [1 + 2\alpha(\kappa_s)T_0/3]}{2}, \quad (C.38)$$

where the functions $n(\kappa_s)$ and $\alpha(\kappa_s)$ can be determined by empirical regression.

C.4 Results

In fig. C.2 (A) the model errors for the least-squares fit to the models in eqs. (C.13a) and (C.13b) are shown. An overview of the corresponding correlation coefficients and mean average percentage errors are shown in table C.2. In fig. C.2 (B) the model errors for the measurable nonlinear response of the medium, β_p , is shown for the models in eqs. (C.14a) and (C.14b). From table C.2 we observe a broadening of the gap between the models in κ_s and B . The correlation coefficient between κ_s and B is found to be $R = -0.0083$, meaning that the two quantities are uncorrelated on the data set.

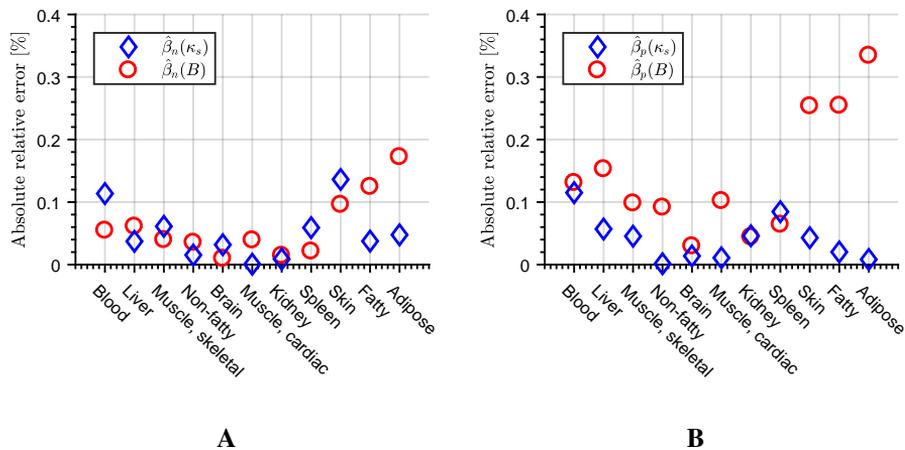


Figure C.2: (A) Relative errors of models for the coefficient of nonlinearity, β_n , given in eqs. (C.13a) and (C.13b). (B) Relative errors of models of the measurable nonlinear response of the medium, β_p , given in eqs. (C.14a) and (C.14b).

Table C.2: Correlation coefficient and mean average percentage error (MAPE) for empirical regression.

	Empirical Regression			
	$\beta_n(\kappa)$	$\beta_p(\kappa)$	$\beta_n(B)$	$\beta_p(B)$
R	0.812	0.978	0.527	0.279
MAPE	5.0%	4.0%	6.1%	14.4%

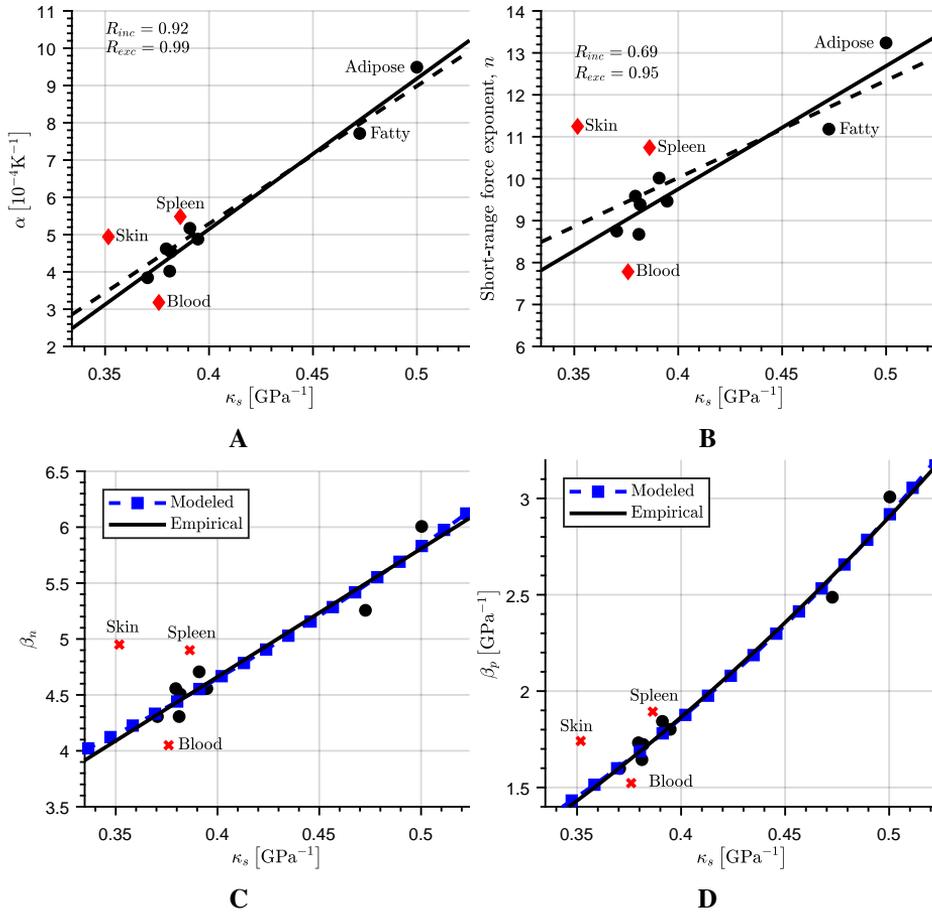


Figure C.3: Relationships based on thermodynamic model. The solution to optimization problem in eq. (C.37), yielding the isobaric thermal expansion coefficient in (A) and the short-range force exponent n in (B). Based on the linear relationship between n and κ_s , and α and κ_s the modeled regression in eq. (C.39) is shown in (C) and the corresponding model for the measurable nonlinear response of the medium, β_p , in (D).

In fig. C.3 the results of the thermodynamic model is summarized. The potential specific compressibility is $\bar{\kappa}_p \approx 1.22 \text{ GPa}^{-1}$ with a $\pm 0.2\%$ variation about the mean. Similarly, the mean operating point is found to be $\bar{\eta} \approx 1.12$ with a $\pm 3.7\%$ variation. In (A) the isobaric thermal expansion coefficient is given. In (B) the short-range force exponent, n , from the optimization problem in eq. (C.37) is shown. With the exception of skin, blood and spleen, approximating the variation in $n(\kappa_s)$ and $\alpha(\kappa_s)$ as linear appears as a reasonable model. By calculating linear regressions for n and α separately, excluding the outliers, and inserting into eq. (C.38) we obtain,

$$\begin{aligned}\beta_n &= \frac{5 + (a_n \kappa_s + b_n) [1 + 2(a_\alpha \kappa_s + b_\alpha) T_0]}{2}, \\ &= 1.23 \times 10^{19} \kappa_s^2 + 7.18 \times 10^8 \kappa_s + 2.39.\end{aligned}\quad (\text{C.39})$$

In (C) this modeled regression is shown compared to the empirical regression excluding the same outliers. Note that the modeled regression is quadratic in κ_s , due to κ_s dependence in both n and α , while the empirical regression is linear in κ_s . In (D) the corresponding modeled and empirically determined regression is shown for the measurable nonlinear response of the medium, β_p .

C.5 Discussion

The effect of nonlinear bulk elasticity on ultrasound wave propagation is produced by the variation of the propagation velocity with pressure, as described in eq. (C.8). The propagation velocity depends on the pressure through the parameter $\beta_p = \beta_n \kappa_s$ in eq. (C.9). We note that even if $B = 0$, we have a pressure dependency of the propagation velocity through the first, unit term in $\beta_n = 1 + B/2A$. In the Lagrange description that we use, this term enters from the continuity equation, as shown in appendix C.6. Using the Euler description to develop the wave equation, this unit term arises from the convective acceleration, as shown by Hamilton & Blackstock [7]. Hence, eq. (C.9) shows that:

1. The wave propagation is nonlinear even when $B = 0$, but $B > 0$ greatly enhances the nonlinear effect.
2. The nonlinear wave propagation is determined by β_p which is highly influenced by the linear compressibility κ_s .

Using the empirical data in table C.1, we found no correlation between κ_s and B ($R = -0.0083$). Consequently, for β_n to be correlated with κ_s , the variation in κ_s must be larger than the variation in B . By modeling the coefficient of nonlinearity, β_n , to be either a function of B , eq. (C.13b), or a function of κ_s , eq. (C.13a), we observed in fig. C.2, that for several materials, the variation is described by

the compressibility to a large degree. Fatty and adipose tissues appear as extremes where β_n appears close to fully defined by the compressibility. Skin and blood however, appear as outliers, having the highest and lowest value of B respectively, shown in table C.1. For the measurable nonlinear response of the medium, β_p , the outliers become less distinguished and close to the model $\propto \kappa_s^2$.

Through the thermodynamic model we obtained an equation for β_n as a function of the isobaric thermal expansion coefficient and n , describing the characteristics of the short-range intermolecular force. The initial system of equations in eqs. (C.34a) and (C.34b) is under-determined, but the parameters, κ_p and n all have interconnected effects on the characteristics of β_n and κ_s . As we chose a fixed long-range force exponent, $m = 6$, we could reduce the number of free parameters by assuming κ_p to be close to constant, thus limiting the variation to the short-range force exponent, n , and the thermal expansion coefficient, α , both of which are important in the determination of β_n in eq. (C.34b). Note, however, that erroneous assumptions of m and κ_p will cause a compensation in n and α . Consequently, the parameters are to a degree confounded and not precise. However, both α and n appears within reasonable ranges and we hence assume that our assumptions are to a certain degree valid for this simple model.

The consequence of eq. (C.34b) is that for a given material with a given n and m the coefficient of nonlinearity, β_n , is defined by α through its effect on the volume. As discussed in section C.3.2, α and κ_T are inherently linked as soft tissues are composed of the same basic constituents, where for the majority of materials, water constitutes the bulk. Consequently, from the kinetic model, the ratio α/κ_T in eq. (C.36) should be approximately constant as the number and size of the molecules in the volume is similar. The close to linear relationship in fig. C.3 (A) seems to support this. In addition, there is a reasonable agreement between the estimated values and those found in [22]. However, as different amino acids can have vastly different effects on the thermal expansion coefficient in mixtures [23] and α is possibly a confounded parameter in the model, a perfect correspondence is not expected.

The second parameter in the thermodynamic model, n , describes the characteristics of the short-range intermolecular force. From fig. C.3 (B) we see that a large portion of the materials is in the range $n = 8.5 - 10$, in line with the selection of $n = 9$ by Hartmann[8]. Adipose tissue has the strongest short-range force, $n \approx 13$, while "average" fatty tissue[9], which can be viewed as soft tissues with higher lipid content, reasonably appears in between at $n \approx 11$. Skin and spleen is classified in the same range as fatty tissue, while blood had the lowest value of $n \approx 7.8$, further supporting them to be viewed as outliers. Interestingly, with the exception of these outliers there appears to be a strong linear relationship between n and κ_s indicated by the high correlation coefficient ($R = 0.95$). Consequently,

modeling $n(\kappa_s)$ and $\alpha(\kappa_s)$ in eq. (C.38) as the linear regressions for n and α excluding these outliers, we obtained a modeled relationship between β_n and κ_s given in eq. (C.39). This model is entirely given by κ_s and we observed in fig. C.3 (C) and (D) an excellent agreement between the modeled and empirical regression.

Both our empirical and thermodynamic model indicate that for several soft tissues the observed nonlinearity is, to a large degree, defined by the compressibility of the material. This warrants a discussion on ultrasound tissue characterization based on the nonlinear response of soft tissues.

First, the magnitude of the nonlinear distortion is determined by the change in speed of sound given by β_p in eq. (C.8). It is this parameter we are able to measure from nonlinear wave propagation. Consequently, unless the speed of sound is also measured, methods estimating the coefficient of nonlinearity typically infer assumptions on the compressibility and speed of sound as discussed in section C.2.2. However, due to the correlation between β_n and κ_s the product of these has a higher variability compared to the individual parameters alone. This is seen in fig. C.1, where the variation span in β_p is higher for the bulk of materials with the exception of skin which has been consistently identified as an outlier. Fatty and adipose tissues appear as extremes, indicating that estimation of β_p can yield increased sensitivity in detection of fat content. In, for instance, diagnosis of alcoholic or non-alcoholic fatty liver disease[5, 24, 25], or detection of lipid rich vulnerable plaque[26], this relationship could be exploited for improved detection.

Second, another benefit of characterizing based on the nonlinear response of the medium, β_p , is that we reduce the number of assumptions in the estimation. The expression in eq. (C.10) can be reformed to only require an assumption on the speed of sound. For the soft tissues in table C.1, the variation in density is $\sim \pm 7\%$ and the variation in compressibility is $\sim \pm 20\%$. However, due to the negative correlation between density and compressibility[9], and the square root relationship, the variation in the speed of sound is only $\sim \pm 5\%$. Consequently, errors due to invalid assumptions on the ambient conditions can be reduced accordingly. Alternatively, using the model in eq.(C.13a), we can produce an estimate on the values for compressibility and speed of sound. As seen in fig. C.3 (D) our model appears to be a good fit to the measurable nonlinear response of the medium, β_p . Consequently, the compressibility can be estimated as,

$$\hat{\kappa}_s = \frac{\sqrt{4a_\kappa\beta_p + b_\kappa^2} - b_\kappa}{2a_\kappa}. \quad (\text{C.40})$$

Moreover, Mast reports, although negative, a strong correlation between the density and compressibility[9]. By approximating the density as a linear function of compressibility,

$$\hat{\rho}_s = a_{\rho,\kappa}\kappa_s + b_{\rho,\kappa} \quad (\text{C.41})$$

we can estimate the speed of sound as,

$$\hat{c}_s = (\hat{\rho}_s \hat{\kappa}_s)^{-1/2}. \quad (\text{C.42})$$

Third, a natural question that arises due to the observed relation between β_n and κ_s in several soft tissues is what additional information is included in β_n . If indeed the coefficient of nonlinearity is largely determined by the compressibility, the nonlinear response of the medium β_p becomes an estimate of the compressibility squared as shown in eq. (C.9). In this case, estimates of β_n roughly becomes a more sensitive estimate of the speed of sound. However, the estimation approach can be simpler due to its pressure dependence as discussed in section C.2.2. Clearly the outliers in our analysis are the exceptions where the information in κ_s and β_n is distinctly dissimilar. The question therefore arises whether the thermodynamic relations in eqs. (C.34a) and (C.34b) could provide more information. The materials appearing as outliers are therefore of special importance and having a larger collection of materials, is needed for to provide further insight. Especially interesting is where tissues with pathology will align within this framework.

Finally, the motivation behind the derived relationships in eqs. (C.34a) and (C.34b) was finding a relationship between the compressibility and coefficient of nonlinearity. However, as a consequence, we obtained an equation linking the coefficient of nonlinearity to thermal properties. This indicates that eq. (C.34b) could express temperature, potentially allowing thermal monitoring in tissue ablation and high intensity focused ultrasound (HIFU) applications. In addition, as swapping out the intermolecular potential in the thermodynamic model is trivial, exploring other intermolecular potentials is grounds for future work. A special emphasis should be put on potentials with more intuitive and more orthogonal parameters.

C.6 Conclusion

As the nonlinearity parameter B/A is tied to the compressibility through $A = 1/\kappa_0$, the strength of its correlation to the sound speed, density and compressibility is determined by the variation in B . Both empirically and theoretically by a thermodynamic model using intermolecular potentials, we found that for several materials the variation in B is similar or small compared to the variation in A , meaning that the nonlinearity of those materials is partially or strongly determined by its compressibility. Consequentially, in ultrasound tissue characterization applications where materials with a high variation in κ_0 is involved, especially in detection of lipid rich tissue, estimating the product $\beta_p = \beta_n \kappa_0$ as it arises naturally in the wave equation provides a higher sensitivity due to correlated variables.

Acknowledgments

The authors acknowledges support from the Norwegian Research Council, project number 254633. The authors would also like to thank Mark Hamilton for introducing us to Ballou's rule and Douglas Mast for discussion on the empirical correlations.

Appendix: Derivation of EQ.(6)

The continuity equation in Lagrangian coordinates takes the form,

$$\rho(\Delta V + \delta V) = \rho_0 \Delta V, \quad (\text{C.43})$$

where ΔV is the undisturbed volume element and δV is a small perturbation due to the compression. This can be written as,

$$\rho(\Delta V + \delta V) = (\rho - \Delta\rho)\Delta V, \quad (\text{C.44})$$

where $\Delta\rho = \rho - \rho_0$ is the change in density due to the acoustic compression and ρ_0 is the density in the relaxed material. The relation to the particle displacement then becomes,

$$-\nabla\underline{\psi} = -\frac{\delta V}{\Delta V} = \frac{\rho - \rho_0}{\rho}. \quad (\text{C.45})$$

Rearranging and assuming small volumetric changes we obtain,

$$\frac{\rho - \rho_0}{\rho_0} = -\frac{\nabla\underline{\psi}}{1 + \nabla\underline{\psi}} \approx -\nabla\underline{\psi}(1 - \nabla\underline{\psi}). \quad (\text{C.46})$$

Inserting into eq. (C.1) and only including terms up to second order yields,

$$p - p_0 = A\nabla\underline{\psi}(1 - \nabla\underline{\psi}) + \frac{B}{2} [-\nabla\underline{\psi}(1 - \nabla\underline{\psi})]^2 \quad (\text{C.47})$$

$$= -\frac{1}{\kappa_s} \nabla\underline{\psi} + \frac{\beta_n}{\kappa_s} (\nabla\underline{\psi})^2, \quad (\text{C.48})$$

where $\beta_n = 1 + B\kappa_s/2$. This corresponds to the isentropic state equation.

Bibliography

- [1] M. L. Oelze and J. Mamou. "Review of Quantitative Ultrasound: Envelope Statistics and Backscatter Coefficient Imaging and Contributions to Diagnostic Ultrasound." *IEEE Transactions Ultrasonics, Ferroelectrics, and Frequency Control*, 63(2), **2016**. doi: 10.1109/TUFFFC.2015.2513958.
- [2] H. Kim and T. Varghese. "Hybrid Spectral Domain Method for Attenuation Slope Estimation." *Ultrasound in Medicine & Biology*, 34(11):1808 – 1819, **2008**. doi: <https://doi.org/10.1016/j.ultrasmedbio.2008.04.011>.

- [3] F. Duck. “Tissue non-linearity.” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):155–170, **2010**. doi: 10.1243/09544119JEIM574.
- [4] J. Zhang, M. S. Kuhlenschmidt, and F. Dunn. “Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A.” *The Journal of the Acoustical Society of America*, 89(1):80–91, **1991**. doi: 10.1121/1.400371.
- [5] D. Zhang and X.-F. Gong. “Experimental investigation of the acoustic nonlinearity parameter tomography for excised pathological biological tissues.” *Ultrasound in Medicine & Biology*, 25(4):593–599, **1999**. doi: 10.1016/S0301-5629(98)00185-9.
- [6] A. B. Coppens, R. T. Beyer, and J. Ballou. “Parameter of Nonlinearity in Fluids. III. Values of Sound Velocity in Liquid Metals.” *The Journal of the Acoustical Society of America*, 41(6):1443–1448, **1967**. doi: 10.1121/1.1910504.
- [7] M. F. Hamilton and D. T. Blackstock. *Nonlinear Acoustics*. Acoust Soc Am Press, New York, **2008 (originally published in 1998)**.
- [8] B. Hartmann. “Potential energy effects on the sound speed in liquids.” *The Journal of the Acoustical Society of America*, 65(6):1392–1396, **1979**. doi: 10.1121/1.382924.
- [9] T. D. Mast. “Empirical relationships between acoustic parameters in human soft tissues.” *Acoustics Research Letters Online*, 1(2):37–42, **2000**. doi: 10.1121/1.1336896.
- [10] R. E. Apfel. “Prediction of tissue composition from ultrasonic measurements and mixture rules.” *Journal of the Acoustical Society of America*, 79(1):148–152, **1986**. doi: 10.1121/1.393638.
- [11] R. Hansen, S.-E. Måsøy, T. A. Tangen, and B. A. Angelsen. “Nonlinear propagation delay and pulse distortion resulting from dual frequency band transmit pulse complexes.” *The Journal of the Acoustical Society of America*, 129(2):1117–27, **2011**. doi: 10.1121/1.3518753.
- [12] X. Gong, D. Zhang, J. Liu, et al. “Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues.” *The Journal of the Acoustical Society of America*, 116(3):1819–1825, **2004**. doi: 10.1121/1.1781709.

-
- [13] M. Nikoonahad and D. C. Liu. “Pulse-Echo Single Frequency Acoustic Nonlinearity Parameter (B/A) Measurement.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 37(3):127–134, **1990**. doi: 10.1109/58.55301.
- [14] M. Fatemi and J. F. Greenleaf. “Real-time assessment of the parameter of nonlinearity in tissue using ‘nonlinear shadowing’.” *Ultrasound in Medicine and Biology*, 22(9):1215–1228, **1996**. doi: 10.1016/S0301-5629(96)00140-8.
- [15] A. Kato and Y. Watanabe. “Measurement Method of Spatial Distribution of Nonlinearity Parameter B/A Using Nonlinear Interaction of Two Sound Waves.” *Japanese Journal of Applied Physics*, 32(Part 1, No. 5B):2274–2278, **1993**. doi: 10.1143/JJAP.32.2274.
- [16] N. Ichida, T. Sato, and M. Linzer. “Imaging the Nonlinear Ultrasonic Parameter of a Medium.” *Ultrasonic Imaging*, 5(4):295–299, **1983**. doi: 10.1016/S0161-7346(83)81002-5.
- [17] N. Ichida, T. Sato, H. Miwa, and K. Murakami. “Real-Time Nonlinear Parameter Tomography Using Impulsive Pumping Waves.” *IEEE Transactions on Sonics and Ultrasonics*, 31(6):635–641, **1984**. doi: 10.1109/T-SU.1984.31548.
- [18] H. Fukukita. “Ultrasound pulse reflection mode measurement of nonlinearity parameter B/A and attenuation coefficient.” *The Journal of the Acoustical Society of America*, 99(5):2775, **1996**. doi: 10.1121/1.414873.
- [19] R. Hansen, S.-E. Måsoy, T. F. Johansen, and B. a. Angelsen. “Utilizing dual frequency band transmit pulse complexes in medical ultrasound imaging.” *The Journal of the Acoustical Society of America*, 127(1):579–87, **2010**. doi: 10.1121/1.3257584.
- [20] F. Varray, M. Pasovic, C. Cachard, P. Tortoli, and O. Basset. “Acoustic nonlinearity parameter of tissue on echo mode: review and evaluation of the different approaches for B/A imaging.” In *Ultrasonics Symposium (IUS), 2009 IEEE International*, pages 41–44. IEEE, **2009**.
- [21] K. Huang. *Statistical Mechanics*. John Wiley & Sons, New York, 2 edition, **1987**. ISBN 9780471815181.
- [22] F. A. Duck. *Physical properties of tissues: a comprehensive reference book*. IPEM, United Kingdom, **2013**. ISBN 9781903613504.

- [23] L.-N. Lin, J. F. Brandts, J. Brandts, and V. Plotnikov. “Determination of the Volumetric Properties of Proteins and Other Solutes Using Pressure Perturbation Calorimetry.” *Analytical Biochemistry*, 302(1):144–160, **2002**. doi: 10.1006/ABIO.2001.5524.
- [24] M. Bayard, J. Holt, and E. Boroughs. “Nonalcoholic fatty liver disease.” *American Family Physician*, 73(11):1961–1969, **2006**. doi: 10.1056/NEJMra011775.
- [25] U. R. Acharya, O. Faust, F. Molinari, et al. “Ultrasound-based tissue characterization and classification of fatty liver disease: A screening and diagnostic paradigm.” *Knowledge-Based Systems*, 75:66 – 77, **2015**. doi: <https://doi.org/10.1016/j.knosys.2014.11.021>.
- [26] J. L. Fleg, G. W. Stone, Z. A. Fayad, et al. “Detection of high-risk atherosclerotic plaque: report of the NHLBI Working Group on current status and future directions.” *JACC. Cardiovascular imaging*, 5(9):941–55, **2012**. doi: 10.1016/j.jcmg.2012.07.007.

Paper D

Dual frequency transducer design for suppression of multiple scattering

Ola F. Myhre, Johannes Kvam¹ and Bjørn A.J. Angelsen

Department of Circulation and Medical Imaging, Norwegian University of Science and
Technology
P.O. Box 8905, N-7489
Trondheim, Norway.

Running title: Dual Frequency Transducer Design

Published in 2016 IEEE International Ultrasonics Symposium (IUS), Tours, 2016, pp. 1-4.

Abstract – Reverberation noise reduces the contrast resolution in ultrasound images. In some cases the signal from the anatomy can be completely masked out, making diagnosis challenging. SURF (Second-order Ultrasound Field) imaging is a dual band imaging technique that has shown the capability to suppress reverberation noise and enhance the anatomical signal, providing a higher contrast resolution. SURF transducers radiate pulse complexes comprised of two widely separated frequencies through a partially common radiation surface. In order to achieve optimal noise suppression, careful design of the acoustic stack and radiation apertures is needed. This paper presents the optimization criteria for SURF probes, and describes design solutions using a 9/0.5 MHz linear array for carotid imaging as an example. Simulated transfer functions are compared to those of a manufactured probe.

D.1 Introduction

SURF is a dual band imaging technique that utilizes a low frequency (LF) pulse to manipulate the non-linear elasticity of the medium observed by a co-propagating high frequency (HF) imaging pulse. The technique has proven capabilities for

¹Also at: Norsvin SA, Hamar, Norway

suppression of reverberation noise[1]. The LF pressure, p_m , alters the observed propagation velocity of the HF as,

$$c = c_0(1 + \beta_n \kappa p_m), \quad (\text{D.1})$$

where $\beta_n = 1 + B/2A$ is the coefficient of non-linearity, and κ is the bulk compressibility of the medium. This manipulation of propagation velocity produces an accumulative time delay which is proportional to the LF pressure. A backscattered HF pulse with a co-propagating LF pulse will hence arrive at time,

$$t_{\text{recv}}(z) = 2z/c_0 + \tau(z), \quad (\text{D.2})$$

where $\tau(z)$ is called the non-linear propagation delay (NPD). The NPD development is given by manipulation of the non-linear elasticity due to the LF pressure along propagation path s to depth z given by [2],

$$\tau(z) = - \int_0^z \frac{\beta_n \kappa}{c}(s) p_m(s) ds. \quad (\text{D.3})$$

Reverberation noise can be suppressed by using the measured NPD between received signals which are acquired from transmissions with different LF polarities.

The HF also experiences an accumulative pulse form distortion (PFD) due to the variation of p_m over the HF pulse. This variation is related to the waveform of the LF pulse, and the position of the HF pulse on the LF pulse, shown in Fig. D.1.

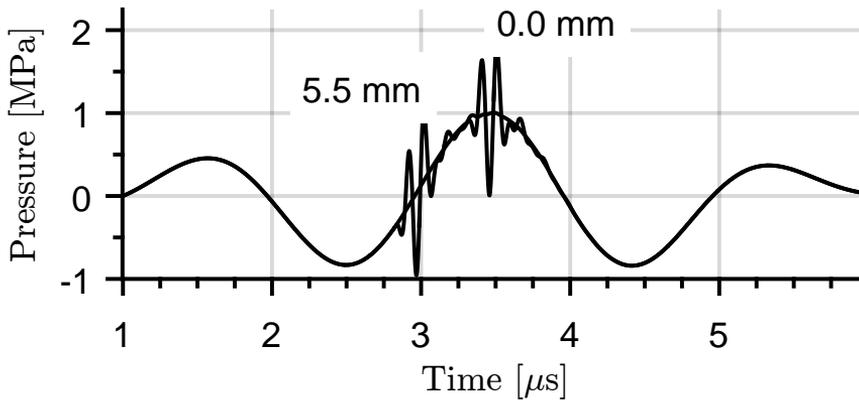


Figure D.1: SURF pulse complex with a 18:1 HF-to-LF frequency ratio. The two HF pulses show the positioning of the HF pulses at the center (0.0 mm) and edge (5.5 mm) of the transmit aperture. The F-number is 2, and the focus is at 22 mm.

High suppression of reverberation noise is favoured by minimum PFD and a linear NPD [1]. This is achieved by designing transducers with high LF-to-HF ratio, and optimal LF aperture size.

The HF transducer is placed in front of the LF transducer in an acoustic stack. The LF wave propagates through the HF transducer through a partially common aperture, as shown in Fig. D.3. Optimization of the LF aperture is discussed in Sec. D.2. The transducer stack configuration is shown in Fig. D.4, and discussed in Sec. D.3.

D.2 Radiation Apertures

Designing the radiation apertures of a SURF transducer involves geometrical optimization of a LF radiation surface to achieve the most uniform manipulation field for a co-propagating HF pulse.

D.2.1 Optimization metrics

In SURF processing we want the NPD development in (D.3) to be defined by the material parameters, meaning that the manipulation LF pressure, $p_m(s)$, should be constant over the imaging region. For a homogeneous material, where $\beta_n \kappa / c$ is constant, this corresponds to a linear development of NPD.

In the practical case, as shown in Fig. D.1, $p_m(s)$ will not be constant. It varies across the HF pulse length due to a limited HF-to-LF frequency ratio. Diffraction and different focusing of the HF and LF beams affect the two pulses differently causing a transverse varying HF-to-LF phase relation. Non-uniformity of the manipulation due to these effects produces an accumulative PFD.

Based on non-linear simulations using [3] we simulate the propagation of an HF pulse y_0 , a SURF complex with a positive LF y_+ and a pure LF y_{LF} . We obtain the HF component of the SURF complex as,

$$y_{+,HF}(z) = y_+(z) - y_{LF}(z). \quad (D.4)$$

We can then use the following two metrics for benchmarking performance of radiation apertures for SURF processing. The first metric measures the linearity of the NPD development,

$$\sigma_{\text{NPD}}(z) = \left| \frac{\tau(z) - \alpha_\tau z}{\alpha_\tau z} \right|, \quad (D.5)$$

where the measured NPD between a dual band pulse and single band pulse is given by $\tau(z)$. The NPD can be calculated through e.g. a cross correlation method. The term $\alpha_\tau z$ is a linear regression of the measured NPD.

The second metric measures the accumulated PFD of the SURF pulse,

$$\sigma_{\text{PFD}}(z) = \frac{\int_{\Gamma} [y_{+,HF}(t - \tau(z)) - y_0(t)]^2 dt}{\int_{\Gamma} y_0^2(t) dt}. \quad (D.6)$$

The SURF pulse complex is delay corrected to compensate for the NPD. The energy in the difference pulse is calculated over an interval $\Gamma = [z/c - T_p/2, z/c +$

$T_p/2]$, a small area around z corresponding to the pulse length, T_p . We then normalize by the energy in the single band pulse, giving us a measurement of the change in pulse shape.

D.2.2 High Frequency Field Considerations

The HF field is received at the transducer and beamformed to generate the image. The HF apertures are designed based on geometrical considerations for a given practical application. The example we are using here is a focused linear array for imaging of the carotid artery. In this application the object (carotid artery) resides approximately at 20 mm with a maximum imaging depth of approximately 50 mm.

Figure D.3 shows the geometrical layout of the HF and LF radiation surfaces and Fig. D.4 shows a cross-sectional view of the acoustic stack. In azimuth the array is focused electronically, while in the elevation direction the HF beam is focused by an acoustic lens. As the LF wave propagates through a partially common aperture with the HF transducer, part of the LF wave will experience focusing through the acoustic HF lens. This will deteriorate LF field uniformity. Choice of HF elevation aperture hence becomes a balance between adequate imaging performance and minimal impact on LF wave uniformity. In Tab. D.1 the parameters of the HF array is shown.

D.2.3 Low Frequency Field Considerations

A good starting point to achieve lateral uniformity is to transmit a LF plane wave in the azimuthal direction. Due to the low frequency the LF aperture has to be larger than the HF aperture in order to avoid diffraction limitations. In order to achieve axial uniformity we can set the aperture size so that it transitions into the the far-field as described in [4, pp. 5.5], toward the end of the imaging region,

$$z_{\text{transition}} = \frac{L_x^2}{4\lambda_{\text{LF}}} \sim z_{\text{max}}. \quad (\text{D.7})$$

This can counteract loss with depth due to absorption and maintain axial uniformity.

The main challenge with LF field uniformity however is in the elevation direction due to size limitations as we want the probe itself to not be larger than necessary. This limits the height of the LF elements, generally causing it to be diffraction limited. In addition the HF acoustic lens causes a slight focusing of the LF wave propagating through the HF transducer, reducing field uniformity.

In order to find the optimal aperture providing the least PFD and the most linear NPD we perform non-linear simulations of the SURF pulse complexes. We then find the NPD linearity through (D.5) and the accumulated PFD through (D.6). This is then done for multiple element heights to find the optimal geometry.

Table D.1: High Frequency (HF) array parameters.

	Parameter	Symbol	Value
Object	Region of interest	ROI	10–30 mm
Frequency	Center frequency	f_c	9.0 MHz
	Wavelength	λ	$\sim 170.0 \mu\text{m}$
	Maximum imaging depth	z_{max}	$\sim 50.0 \text{ mm}$
Azimuth	Element pitch	p	$270.0 \mu\text{m}$
	Number of elements	N_x	128.0
	Maximum aperture	$L_{x,\text{max}}$	34.6 mm
Elevation	Element height	L_y	5.1 mm
	Lens focus	F_y	22.0 mm
	Far-field transition	$L_y^2/4\lambda$	38.0 mm

Table D.2: Low Frequency (LF) array parameters.

	Parameter	Symbol	Value
Frequency	Center frequency	f_c	550.0 kHz
	Wavelength	λ	3.1 mm
Azimuth	Element pitch	p	$810.0 \mu\text{m}$
	Number of elements	N_x	48.0
	Maximum aperture	$L_{x,\text{max}}$	38.9 mm
Elevation	Element height	L_y	11.0 mm
	Lens focus	F_y	N/A
	Far-field transition	$L_y^2/4\lambda$	9.8 mm

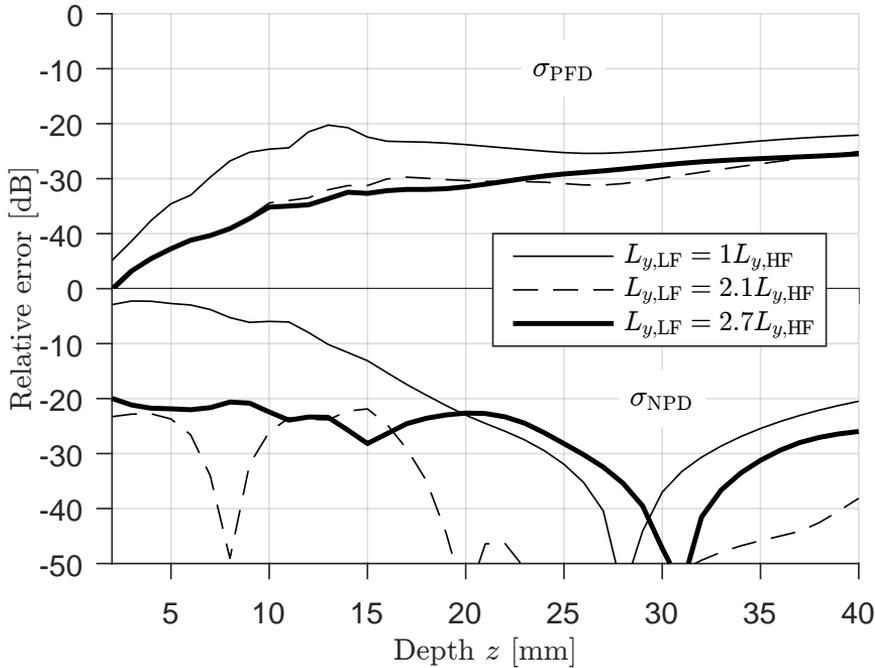


Figure D.2: The accumulated PFD (top) and NPD (bottom) for LF element heights $L_{y,\text{LF}} = \{5.1 \text{ mm}, 10.7 \text{ mm}, 13.77 \text{ mm}\}$ corresponding to $\{100\%, 210\%, 270\%\}$ of the HF element height respectively.

Figure D.2 shows the result from such a simulation based on the HF design presented above for carotid imaging. We observe that for a element height of 10.7 mm we achieve the least accumulation of PFD and deviation from linear NPD development. In Tab. D.2 the parameters for the LF is shown for SURF imaging of the carotid artery.

D.3 Stack Design

In the acoustic stack, the LF transducer is placed behind the HF transducer, as shown by Fig. D.4. The challenge is to find a design which produces HF pulses appropriate for imaging, and simultaneously produces LF pulses that are suitable for manipulating the HF propagation. The HF-to-LF ratio should be large, the HF and LF pulses must be sufficiently short, and the LF pulses must have sufficient amplitude.

The stack design is analyzed using a one-dimensional wave propagation model [4, pp. 3.40]. We define three sections which are a number of layers of material with thicknesses that are much smaller than their lateral extent: the HF, isolation,

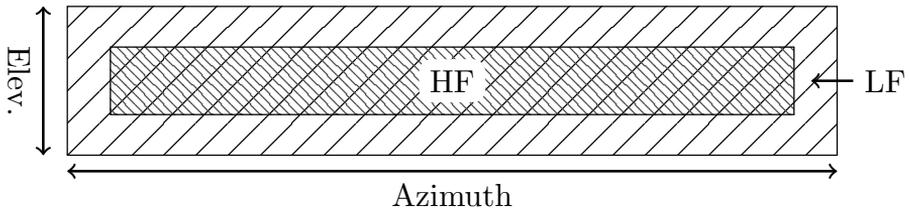


Figure D.3: Front view of the probe radiation apertures. The LF aperture is larger than the HF aperture, and located behind it. In practice, the area in front of the part of the LF aperture which is outside the HF aperture is covered with inactive HF elements.

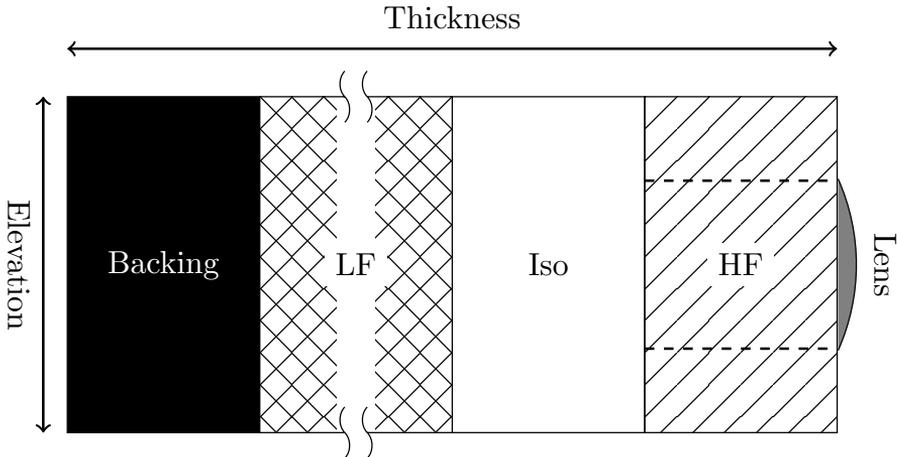


Figure D.4: Cross-sectional overview of the stack structure, showing the different sections. Each section may contain multiple layers of material. The dashed lines show the active part of the HF section, which coincides with the elevation width of the lens.

and LF sections. The HF transducer and the matching layers in front of it are termed the HF section of the stack. The layers of material which are in between the HF and LF transducers are collectively called the isolation section, and are crucial for performance. Behind the isolation section is the LF section, which, in this case, consists only of the LF transducer. The isolation section has two functions; preventing propagation of HF waves from the HF section into the LF section, and serving as an impedance matching structure for the LF section to the load.

The HF transducer should have a very low backing impedance so that little energy is transmitted backwards into the LF transducer and backing of the stack. This prevents unwanted ringing due to multiple reflections in the layers behind the HF transducer. The HF transducer will therefore be more efficient, but less

broad band than a transducer with an absorbent backing. Unlike tissue harmonic imaging (THI), SURF imaging utilizes the fundamental HF band to suppress reverberations, and a lower HF bandwidth is therefore acceptable. This is a trade-off between range resolution and contrast resolution.

D.3.1 Single layer isolation section

A low backing impedance can be achieved by using a single layer of material in between the HF and LF sections. Since the LF transducer has a high impedance, an isolation layer with a thickness of a quarter HF wave length and low impedance will transform the impedance at the back of the HF section to a low impedance.

Due to the low thickness of the HF section relative to the LF wavelength, and the low impedance of the loading material, the HF section is essentially a heavy mass at LF. Conversely, due to its low impedance and small thickness, the isolation layer can be considered to be a spring at LF. This mass-spring interaction between the isolation layer and the HF section at the LF determines the loading impedance of the LF transducer. However, the thicknesses and impedances of the layers creating the interaction are optimized for HF performance. Therefore, the frequency band in which the LF transducer is matched to the load is set by the HF optimization when a single isolation layer is used.

Figure D.5 shows the normalized HF backing impedance, and the normalised LF loading impedance, as a function of normalized frequency when a single isolation layer is used. The impedance magnitude is normalized to the characteristic impedance of the HF transducer, and the frequency is normalized to the HF. In this case, the isolation layer has a thickness of a quarter wavelength at HF, and a low impedance relative to the HF and LF transducers. The impedance resonance seen at $\omega_r = 1/6$ provides a suitable matching for the LF, since the impedance at resonance is also purely real valued. If the LF is selected so that the HF-to-LF frequency ratio is 6 to 1, the LF transducer efficiency is maximized.

Also shown in Fig. D.5 is the backing impedance of the HF transducer when using a single isolation layer. The backing impedance exhibits an oscillation which is given by the thickness of the LF transducer, indicating reflections from the back of the LF transducer. Simulations of HF pulse transmission from a lossless stack model show that the amplitude of these reflections is -25 dB relative to the main transmit pulse amplitude. These reflections deteriorate image quality and should therefore be suppressed.

D.3.2 Triple layer isolation section

Reflections from the back of the LF transducer can be suppressed by using multiple layers of material in the isolation section. These layers are enumerated from front to back, so that the isolation layer which is directly behind the HF transducer is the first isolation layer.

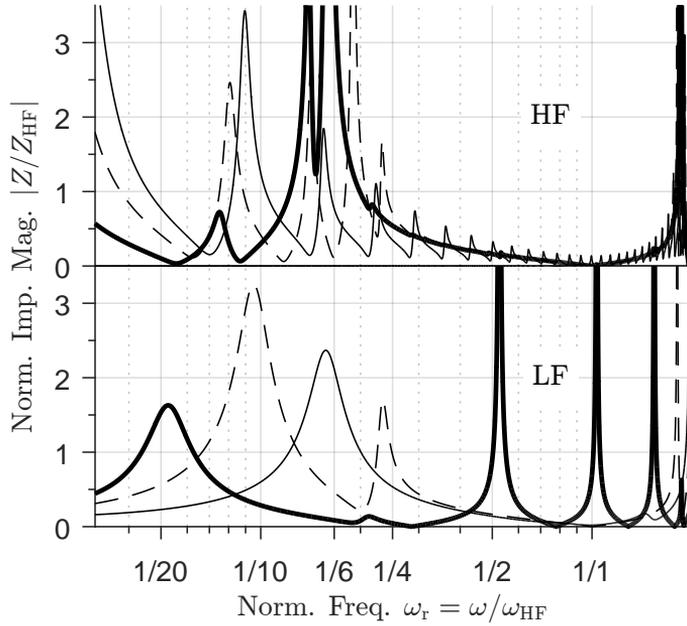


Figure D.5: Normalized HF backing impedance (top) and normalized LF loading impedance (bottom) when using a single isolation layer (thin solid), when using three layers with alternating low and high impedance (dashed), and when using three layers with the thickness of the third layer tuned to achieve an impedance resonance at $\omega_r = 1/18$ (thick solid).

If all the layers have a thickness of a quarter wave length at HF, the HF backing impedance at HF is, for n matching layers

$$Z_{\text{HF}}^{\text{B}n}(\omega_{\text{HF}}) = \begin{cases} \frac{Z_{\text{LF}} \prod_{i=1}^{n/2} Z_{\text{I}(2i-1)}^2}{\prod_{i=1}^{n/2} Z_{\text{I}(2i)}^2} & n \text{ even} \\ \frac{\prod_{i=1}^{(n+1)/2} Z_{\text{I}(2i-1)}^2}{Z_{\text{LF}} \prod_{i=1}^{(n-1)/2} Z_{\text{I}(2i)}^2} & n \text{ odd} \end{cases} \quad (\text{D.8})$$

when the LF section is regarded as infinitely thick. $Z = \rho c$ is characteristic impedance, and $Z_{\text{I}n}$ is the impedance of the n 'th matching layer. Equation (D.8) shows that selecting an odd number of isolation layers is advantageous for achieving a low HF backing impedance. The result also shows that odd-numbered isolation layers should have a low characteristic impedance, whereas even-numbered isolation layers should have a high characteristic impedance. Such a structure is, interestingly, found in flexible printed circuits.

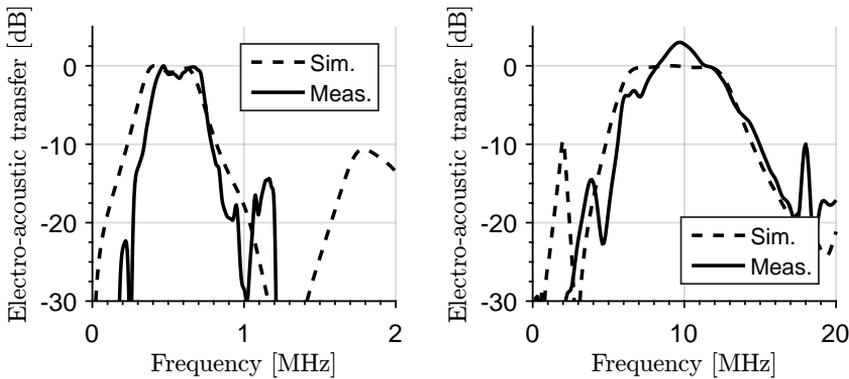


Figure D.6: Simulated (dashed) and measured (solid) transfer functions of the LF (left) and HF (right) arrays of the Vora-T probe.

Figure D.5 shows how the additional isolation layers affect the backing impedance of the HF transducer, and the LF loading impedance when $n = 3$ (dashed). Also shown is the effect of increasing the thickness of the third isolation layer to one HF wavelength (thick solid). In the HF pass band, the oscillations in the HF backing impedance disappear when using three isolation layers, *i.e.* there are less reflections from the rear of the LF section. This occurs irrespective of the thickness of the third isolation layer. Simulations of pulse excitations in a lossless model of the stack show that the amplitude of these reflections are smaller than -60 dB compared to the main pulse. The addition of isolation layers also changes the LF performance by moving the impedance resonance further down in frequency.

D.4 Results

A linear array SURF probe named Vora-T, operating at 9/0.5 MHz was manufactured by Vermon SA. The simulated and measured transfer functions are shown in Fig. D.6.

D.5 Conclusion

It is a challenge to design dual-band ultrasound probes that are optimized for SURF imaging. The LF manipulation should be as uniform as possible across the HF pulse. Lower LF gives more uniform manipulation, but requires a larger aperture. It is possible to obtain a large HF-to-LF ratio in a dual band transducer stack without hampering the performance of the LF and HF transducers in the stack. In particular, Fig. D.5 shows that the thickness of the third isolation layer can be tuned so that the LF loading impedance resonance occurs at a specified frequency. This tuning of layer thickness does not hamper HF backing impedance and performance. By optimizing the HF and LF radiation apertures one can find geometries

that produce minimum pulse form distortion and a linear development of nonlinear propagation delay. This allows for the design of dual band probes that improve the reverberation suppression capabilities of the SURF method. These probes seem well suited for quasi-simultaneous ultrasound imaging and therapy, due to their high acoustic efficiency, and their ability to operate at a wide set of frequencies. Their efficiency and power emission should therefore be studied in detail.

Bibliography

- [1] O. M. Brende and B. Angelsen. “Adaptive reverberation noise delay estimation for reverberation suppression in dual band ultrasound imaging.” *The Journal of the Acoustical Society of America*, 138(5):3341–51, **2015**. doi: 10.1121/1.4935555.
- [2] R. Hansen, S.-E. Måsøy, T. A. Tangen, and B. A. Angelsen. “Nonlinear propagation delay and pulse distortion resulting from dual frequency band transmit pulse complexes.” *The Journal of the Acoustical Society of America*, 129(2):1117–27, **2011**. doi: 10.1121/1.3518753.
- [3] J. Kvam, B. A. Angelsen, and A. C. Elster. “GPU simulation of nonlinear propagation of dual band ultrasound pulse complexes.” In *AIP Conference Proceedings*, volume 1685, page 070003. AIP Publishing, **2015**.
- [4] B. A. J. Angelsen. *Ultrasound Imaging - Waves, Signals and Signal Processing*. Emantec AS, Bugges veg 4B, 7051 Trondheim, Norway, **2000**.

Paper E

SURF Nonlinear Bulk Elasticity Imaging

Johannes Kvam^{1,2}, Stian Solberg³, Ola F. Myhre¹, Alfonso Rodriguez-Molares¹ and Bjørn A.J. Angelsen¹

¹ Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim Norway

² Norsvin SA, Hamar Norway

³ SURF Technology AS, Trondheim, Norway

Running title: SURF Nonlinear Bulk Elasticity Imaging

Manuscript in preparation for submission to IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control.

Abstract – Tissue characterization based on the coefficient of nonlinearity, β_n , has shown promise due to its sensitivity to tissue structure. However, a lack of robust echo modes has limited its clinical success. In this paper we present a pulse-echo estimation technique for measurement of the nonlinear bulk elasticity, given as $\beta_p = \beta_n \kappa$, using a dual frequency approach. The technique is tested in both simulations and *in vitro*, with a prototype scanner, using scanned plane wave transmissions. Resulting estimates are in good agreement with previously published results in literature. The approach is considered mainly qualitative due to uncertainty of the magnitude of the acoustic pressure, but shows quantitative potential with multiple angle acquisitions and pressure modeling.

E.1 Introduction

Healthy and pathological tissue have different acoustic characteristics and hence being able to distinguish between such differences has substantial diagnostic value. Due to the portability and relative inexpensiveness of ultrasound systems, performing tissue characterization with this modality has been an important research topic in the last few decades. Initial attempts at quantitative ultrasound (QUS) were

made to distinguish materials based on the linear parameters e.g. sound speed and backscatter coefficient [1].

Characterizing media based on the coefficient of nonlinearity, $\beta_n = 1 + B/2A$, where B/A is the parameter of nonlinearity, has been argued to be especially promising [2]. The parameter has been shown to be highly affected by structural changes [3, 4] and sensitive to pathological tissue [5].

Since the 80s there have been multiple attempts to estimate the nonlinearity parameter B/A . Most of these methods are either based on parametric arrays [6], estimating the second harmonic [7, 8] or using a probe and pump wave with a high frequency separation [9–11]. A review of approaches is found in [2]. Although some of these methods have shown good results in tomographical or multi-transducer setups, Duck argues that main reason for little clinical success of B/A characterization is the lack of robust pulse echo modes. Varray *et al.* attempted to extend some of the methods to echo mode and highlighted the estimation challenges, mainly modeling of diffraction [12].

In 2007 a dual frequency band imaging method called *second order ultrasound field* (SURF) imaging was introduced for imaging of contrast agents [13]. Over the last decade much work has been dedicated to the exploration of the use of SURF in diagnostic imaging [14]. In SURF imaging, pulse complexes composed of a high frequency (HF or probe wave) superimposed on a low frequency (LF or pump wave), as shown in Fig. E.1, are formed. These pulse complexes are transmitted through a partially common aperture. During the propagation the low frequency manipulates the medium compressibility, which in turn affects the propagation speed of the HF pulse. This causes an accumulative delay or advancement proportional to the nonlinear bulk elasticity of the material, $\beta_p = \beta_n \kappa$. By transmitting multiple pulse complexes of varying LF polarity, the nonlinear properties of the medium can be investigated, however, until now, there have been no attempts at directly estimating tissue nonlinearity. Fukukita showed a similar setup using two circular transducers, but placed the HF at the zero crossing of the LF [15]. This causes an accumulative frequency shift of the HF which is determined by the nonlinear properties of the medium. Using this approach, Fukukita produced B/A estimates in layered medium along a single line.

We present a nonlinear imaging method using a linear, dual-frequency array. The method can be implemented on conventional clinical scanners using pulse-echo acquisition. We describe and evaluate the method in simulations and *in vitro*. In both cases a heterogeneous domain is studied with two regions with different nonlinear properties.

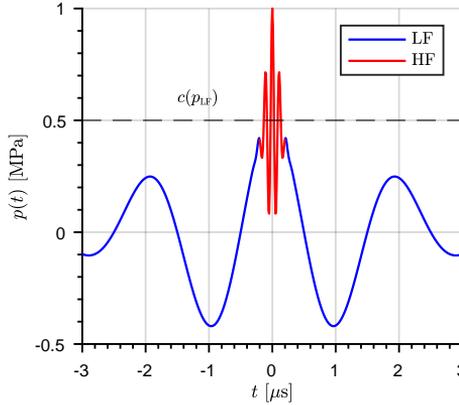


Figure E.1: SURF pulse complex with 10MHz HF and 1MHz LF.

E.2 Theory

In SURF dual band imaging a dual frequency pulse is transmitted through a partially common aperture. The frequency separation between the HF and LF is large, typically 1:10 or 1:20. [16] In Fig. E.1 an example of a SURF pulse complex is shown. The LF causes the HF to observe a manipulated speed of sound, [14]

$$c(p_{LF}) \approx c_0(1 + \beta_p p_{LF}) \quad (\text{E.1})$$

where c_0 is the ambient speed of sound and p_{LF} is the LF pressure. The parameter,

$$\beta_p = \beta_n \kappa_0, \quad (\text{E.2})$$

describes the nonlinear bulk elasticity of the medium, where κ_0 is the compressibility or inverse bulk modulus and $\beta_n = 1 + B/2A$ is the coefficient of nonlinearity. The observed speed of sound is hence increased or decreased based on the polarity of the LF. By transmitting two pulse complexes with different manipulation, e.g. one with a positive LF and one with a negative LF, one observes an accumulative delay or advancement in the backscattered HF component given as,

$$\tau_x(z) = - \int_{\Gamma(z)} \frac{\beta_p(s)}{c_0(s)} p_{LF}(s) ds. \quad (\text{E.3})$$

This is called the nonlinear propagation delay (NPD). The observed delay at a depth z is given by the accumulative manipulation of the nonlinear bulk elasticity $\beta_p = \beta_n(s)\kappa(s)$ along orthogonal trajectories, $\Gamma(z)$, of the HF wavefront. For plane waves we have $ds = dz$ and $\Gamma(z) = [0, z]$ [17]. The delay only continues to accumulate as long as there is a sufficient LF pressure. For diffuse scatterers

the magnitude of the scattered wave is decreased so much that the LF pressure is negligible in backpropagation. Consequently, the NPD is proportional to the forward propagation length z and not by roundtrip distance $2z$. In the case of specular reflectors with a high impedance mismatch the magnitude of the reflected LF wave may not be negligible causing an over- or underestimation of the NPD based on the phase-shift in the reflected wave. This effect is particularly strong in solid materials as shown by [18], but can also be observed, albeit to a lesser extent, in soft tissues.

By exploiting this nonlinear interaction effect between HF and LF, SURF has demonstrated applicability to detection of ultrasound contrast agents [14, 17], suppression of reverberation noise [19, 20] and detection of microcalcification [21].

For a heterogeneous material, different media yield a change in the gradient of the measured delay. By differentiating with respect to receive time, t , we then obtain

$$\beta_p(\underline{r}_\perp, t) = \frac{1}{p_{\text{LF}}(\underline{r}_\perp, t)} \frac{\partial \tau_x(\underline{r}_\perp, t)}{\partial t}. \quad (\text{E.4})$$

To obtain quantitative estimates of the nonlinear bulk elasticity, β_p , the experienced LF pressure must be known. A completely homogeneous LF field is the best case as the field is close to constant and the delay gradient is described by variations in β_p . LF field homogeneity is hence important. This is done by optimizing LF and HF frequency ratio of the transducer stack [16] and careful design of the radiation apertures[22]. Due to high frequency separation of the HF and LF, the effect of absorption and aberration of the LF is small as the propagation distance is on the order of $\sim 10^1 \lambda_{\text{LF}}$. Qualitative results can be obtained by simple LF modeling, where the simplest model is approximating it as constant over the entire propagation. For quantitative results a more complex LF model is needed.

Up to this point it has been assumed that the measured delay, which we denote τ_y , is a direct observation of the first order delay, $\tau_x(\underline{r}_\perp, z)$, *i.e.*, the delay between two signals that originate from a single scatterer at (\underline{r}_\perp, z) . For a heterogeneous medium with background scattering however, the received signal also consists of signals from acoustic sources outside this region. Multiple scattering noise, random interference, as well as refraction effects causes the observed delay to be a weighted sum of contributions from both inside and outside the investigation region, based on a first-order-signal-to-noise ratio (FSNR). The measured delay between two SURF pulse complexes, mapped to a depth z , with different LF polarities can be calculated as,

$$\tau_y(z) = \frac{1}{N_B} \sum_B \frac{1}{\omega} \angle Y_+(\omega, z) Y_-(\omega, z)^*. \quad (\text{E.5})$$

Here, Y_+ and Y_- are the Fourier transforms in a small interval around receive

depth, z , for SURF complexes with a positive and negative co-propagating LF respectively. The measured delay τ_y is the average linear phase component across the bandwidth B related to the HF band. For a heterogeneous scattering medium the delay is composed of a first order delay, corresponding to signal originating from the investigated region, z , and a noise delay from interfering signals originated from high side lobe levels or reverberation noise [19, 20]. The measured delay can be described as a weighted average as,

$$\tau_y(z) = \alpha(z)\tau_x(z) + (1 - \alpha(z))\tau_n(z). \quad (\text{E.6})$$

The variables τ_x and τ_n describe the first order and noise delay respectively and the parameter α is the first-order-signal-to-noise-ratio (FSNR). In situations where the dominating noise is due to multiple scattering, this parameter is called the signal-to-reverberation noise ratio [20]. The noise delay, τ_n , is more random than the first order delay, τ_x , as it a combination of first order delays from other acoustic scatterers outside the observation depth z . Its effect on the measured delay is challenging to predict as it depends both on the FSNR, α , and the combination of acoustic scatterers producing the noise. As the material parameters are given by the first order delay development, τ_x , we need to identify segments of the measured delay, τ_y , which have a high FSNR, *i.e.*, where $\alpha \approx 1$.

E.3 Estimation Approach

In a heterogeneous material estimating the first order delay development, τ_x , is challenging. Random interference as well as multiple scattering affect the measurement of the NPD and modeling becomes necessary to find the delay, τ_x , in (E.6). In this section we use a model based fitting approach used to estimate the first order delay development. As we are only characterizing the nonlinear response of the medium we will present the analysis in terms of the nonlinear bulk elasticity parameter, β_p , given in (E.2). However, with a known material compressibility, $\kappa_0 = 1/\rho_0 c_0^2$, the coefficient of nonlinearity can be calculated.

E.3.1 Delay estimation

To calculate the delay between two SURF complexes of different LF polarities, *e.g.*, $y_+(t)$ and $y_-(t)$, we utilize a cross-correlation based approach. A sliding window of a few wavelengths is used to extract segments from the received RF data from two different LF polarities. As the delay is typically on the order of nanoseconds the RF signals are interpolated to allow sufficient precision. The delay is then calculated based on the maximum correlation between y_+ and y_- .

The normalized cross-correlation for segment i of the sampled RF signals is

given by

$$R_i(l) = \frac{\sum_{n=0}^{N-1} y_+(iN+n)y_-(iN+n+l)}{\sqrt{\sum_{n=0}^{N-1} y_+(iN+n)^2 \sum_{n=0}^{N-1} y_-(iN+n+l)^2}} \quad (\text{E.7})$$

$$l \in [-l_{\max}, l_{\max}],$$

where N is the window length in samples. Let \hat{l}_i be the lag giving maximum correlation for segment i ,

$$\hat{l}_i = \arg \max_l R_i(l). \quad (\text{E.8})$$

By modeling the the cross-correlation as a second order polynomial close to its peak position, the delay giving maximum correlation can be interpolated from the sampled cross-correlation as,

$$\tau_{y,i} = \Delta t \left(\hat{l}_i + \frac{R_i(\hat{l}_i - 1) - R_i(\hat{l}_i + 1)}{2R_i(\hat{l}_i - 1) - 4R_i(\hat{l}_i) + 2R_i(\hat{l}_i + 1)} \right). \quad (\text{E.9})$$

Finally, linear interpolation of $\tau_{y,i}$ gives an estimate of the delay for all depths between each segment, $\tau_y(z)$.

E.3.2 Estimation of β_p

Due to the influence of interfering signals, mentioned in Sec. E.3.1, differentiation cannot be applied to estimate β_p . Instead, we utilize a model-based approach to obtain a more robust estimate.

Regarding the first order delay development a few assumptions can be made. Firstly, given satisfactory transducer design, the delay development is monotonic within the imaging region. This essentially means that the HF does not slide from a peak of the LF to a trough or vice-versa. Such an effect will be due to diffraction of the LF as the effect of the phase shift compared to the HF is large. This can be circumvented by ensuring a minimal phase shift over the imaging region, *i.e.*, ensuring that the LF is a plane wave. Secondly, the nonlinear bulk elasticity, β_p , has a lower and upper bound in soft tissues, typically $\sim 1 - 4 \text{ GPa}^{-1}$, allowing us to limit the feasible solution space. Finally, the measured delay is composed of signal and noise, meaning that some segments of τ_y are more trustworthy than others.

Based on these assumptions and in addition approximating the propagation as plane, *i.e.*, $ds = dz$, we can formulate the optimization problem, as a function of

receive time, t , as,

$$\begin{aligned}
 \min_{\beta_p} \quad & W_x(\underline{r}_\perp, t) \left| \tau_y(\underline{r}_\perp, t) - \frac{1}{2} \int_0^t \beta_p(\underline{r}_\perp, u) p_{\text{LF}}^m(\underline{r}_\perp, u) du \right| \\
 & + \underline{\eta}_v W_n(\underline{r}_\perp, t) |\nabla \beta_p(\underline{r}_\perp, t)| \\
 \text{s.t:} \quad & \beta_p(\underline{r}_\perp, t) > \beta_{p,\min} \\
 & \beta_p(\underline{r}_\perp, t) < \beta_{p,\max}.
 \end{aligned} \tag{E.10}$$

Here, τ_y is the measured delay, u is the integration variable in time, p_{LF}^m is the modeled LF pressure, and $|\nabla \beta_p(\underline{r}_\perp, t)|$ is a smoothness term that penalizes rapid variations in $\beta_p(\underline{r}_\perp, t)$. The term $W_x(\underline{r}_\perp, z)$ is a spatial weighting term which can either increase or decrease the penalty for deviation from the measured delay. Correspondingly, $W_n(\underline{r}_\perp, z)$ can relax the smoothness criteria in regions where a gradient shift is expected. For instance, at strong acoustic interfaces, the FSNR is expected to be high, due to a coherent backscattered signal, and we hence assume that the measured delay will be close to the first order delay, *i.e.*, $\tau_y \approx \tau_x$. Consequently, at these locations we want $W_x = 1$. Moreover, the smoothness criteria should also be relaxed, $W_n \sim 0$, as it is likely a transition into a new material with a new characteristic β_p . In contrast, if a region has highly incoherent signals, we expect the FSNR to be lower and can consequently relax the penalty for deviation, *i.e.* $W_x = 0$ and $W_n = 1$. In these regions the smoothing terms $\underline{\eta}_v |\nabla \beta_p(\underline{r}_\perp, t)|$ become dominant, suppressing rapid variations. These weights can be applied in both the lateral and axial direction by $\underline{\eta}_v = (\eta_{v,\underline{r}_\perp}, \eta_{v,t})$.

In this work the weight factor was chosen as the ratio,

$$W_x(\underline{r}_\perp, t) = \frac{h_{T_\lambda} \otimes_t e_0(\underline{r}_\perp, t)}{h_{T_{3\lambda}} \otimes_t e_0(\underline{r}_\perp, t)} \tag{E.11}$$

where h_{T_λ} and $h_{T_{3\lambda}}$ are low pass filters of one and three HF wavelengths respectively and $e_0(\underline{r}_\perp, t)$ is the envelope of the received RF data for a 0 LF transmission. This simple weighting factor detects axially short coherent signals, enforcing the model to fit the measured delay at interfaces. The weighting factor is normalized for each transmission, $W_x(\underline{r}_\perp, t) \in (0, 1)$, based on the maximum value along each scan line, \underline{r}_\perp . Correspondingly, the weighting of the variation penalty is selected as,

$$W_n(\underline{r}_\perp, t) = 1 - W_x(\underline{r}_\perp, t), \tag{E.12}$$

meaning that at these strong interfaces we allow a rapid variation in β_p .

To reduce the number of parameters in the optimization routine, the depth interval at which β_p is sampled can be increased to a few wavelengths. The estimates at the original sampling frequency can then be found through interpolation. Averaging the result of multiple sampling schemes can also be done to combine the smoothness of coarser sampling with the improved local resolution of a finer sampling. In this work we chose a sub-sampling corresponding to 8 HF wavelengths, and introduced a smoothness penalty only in the depth direction, *i.e.*, $(\eta_{v,r_\perp}, \eta_{v,t}) = (0, 1)$.

The transducer utilized is a linear 2D array and the images are formed as a function of azimuth and time. Consequently, we have, $r_\perp = \underline{x}$, where \underline{x} describes the azimuth position of the elements.

The optimization problem is solved using *lsqnonlin* in Matlab, which minimizes the sum of square errors using a trust-region method. The optimization is run until convergence, and the nonlinear bulk elasticity is directly extracted from the results.

E.4 Materials & Methods

The proposed method for estimating the nonlinear bulk elasticity, β_p , using SURF pulse complexes has been tested both in simulations and *in vitro* using a phantom. The phantom was constructed using materials with linear- and nonlinear bulk properties previously characterized in literature. Simulations were setup to mimic the experimental situation but allows full control over the acoustic parameters. The experiments are conducted using a dual-band imaging system from SURF Technology A/S.

The simulations are set up to mimic the experimental setup, consequently we will first describe the tissue mimicking phantom and experimental scanner before the simulations are presented.

E.4.1 Experimental setup

In this section an overview of the experimental setup is given.

E.4.1.1 Phantom

An agar-based tissue mimicking phantom was made following the instructions in Annex II of IEC60601-2-37 [23]. The phantom was cut into a 5cm x 5cm x 5cm block with a cylindrical hole of 7 mm diameter that was filled with a mixture of corn-oil and corn starch in a concentration of 4 g/dL. Corn starch was used to increase the backscattering intensity of the fluidic phase, making it possible to track the nonlinear delay and reducing the contrast between target and background. At such low concentration it is assumed to have a negligible impact of the nonlinear properties of corn oil.

The linear and nonlinear properties of both media have been previously studied

by [24, 25], and are reproduced in Tab. E.1 for convenience. The absorption is given at 8 MHz.

E.4.1.2 Dual band system

The dual-band imaging system from SURF Technology is based on an Ultrasonix SonixMDP (Ultrasonix, Richmond, BC, Canada), which transmits and receives the HF signals. The LF pulses are transmitted by an Aurotech Manus (Aurotech, Tydal, Norway), which is controlled via the SonixMDP.

The SURF transducer used in the experiments transmits and receives HF at 8 MHz, and LF at 0.8 MHz, through a partially common radiation aperture. The HF-to-LF ratio is 10:1. The HF array consists of 128 elements with a $300\mu\text{m}$ pitch. The system has been set up to transmit scanned plane waves for both the HF and the LF. The transmit aperture of the LF is set to $D_{\text{LF}} = 46.8$ mm and the transmit aperture for the HF is set to $D_{\text{HF}} = 9$ mm. The reasoning behind this is that a plane LF wave gives the most homogeneous LF field. When the LF is plane, the HF should also be plane to avoid a lateral variation in LF pressure as experienced by the HF wave. Images are formed by transmitting 128 plane waves and beamforming using a delay and sum beamformer. For each scanline, two SURF pulse complexes are transmitted, one with a positive LF pressure and one with a negative LF pressure.

E.4.1.3 Low frequency field characterization

The transmitted LF field is characterized by hydrophone measurements in water, in order to accurately solve (E.4). The equivalent manipulation pressure is found from the water tank measurements by averaging the LF pressure over the part of the LF pulse that is common to the HF and LF pulses in a SURF complex, *i.e.*,

$$p_{\text{LF}}(\underline{r}_{\perp}, z) = \frac{1}{t_2(\underline{r}) - t_1(\underline{r})} \int_{t_1(\underline{r})}^{t_2(\underline{r})} p_{\text{LF}}(\underline{r}, t) dt, \quad (\text{E.13})$$

where $\underline{r} = (\underline{r}_{\perp}, z)$, and $t_1(\underline{r})$ and $t_2(\underline{r})$ mark the beginning and end of the HF pulse at \underline{r} , respectively. The LF pressure, $p_{\text{LF}}(\underline{r}, t)$ is found by lowpass filtering the measured SURF complex, and the integration limits, $t_1(\underline{r})$ and $t_2(\underline{r})$ are defined by the 6 dB length of the envelope of the HF pulse, which is found by high pass filtering the measured SURF complex.

The measurements are made for a plane wave setup for both the LF and HF, transmitting using the entire array, yielding an LF aperture of 46.8 mm and HF aperture of 38.4 mm.

The measurements are conducted with an Onda AIMS-III measurement system, an HGL-0200 hydrophone and an AH-2020 pre-amplifier. All measurements are conducted in compliance with established measurement procedures[26]. The

hydrophone is only calibrated down to 1 MHz however, introducing an uncertainty in the measurement of the LF field (0.8 MHz).

E.4.2 Simulations

Simulations were performed in 2D using k-wave[27]. The simulations were set up to mimic the experimental phantom and imaging system. The material parameters used are listed in Tab. E.1. The acoustic parameters for the phantom and oil-filled inclusion is set to mimic the acoustic properties of agar and corn oil respectively[24, 25]. However, as k-Wave does not support a spatial variation in the exponent of the absorption power law, the exponent was set equal to that of the phantom for the whole medium. The value for the acoustic absorption coefficient in the inclusion, α , was therefore set to the absorption of corn oil at ~ 8 MHz. This leads to a slightly higher attenuation of the LF field.

A small spatial variation in the acoustic parameters were introduced to ensure heterogeneity and scattering with realistic speckle patterns. This was done by adding random gaussian noise to each parameter, with -50 dB noise power relative the parameter mean.

As the transducer surface can also act as a near perfect reflector, multiple scattering noise with the second scatterer being the transducer surface, often produces the strongest reverberation noise [20]. To include this in the simulation, a high impedance is set at the simulated acoustic sources to produce a high reflection coefficient equal to 0.5. Consequently, the simulated transducer surface acts as a strong reflector, giving us a realistic signal-to-noise ratio.

Equivalent to the experiments, the simulations uses a plane pressure source with 9 mm diameter for the HF, that is scanned across the medium, to form 128 scanlines. In order to speed up the simulations, only a part of the medium surrounding the active HF source is simulated for each transmit. The diameter of the active simulation area is set to two times the diameter of the HF aperture. This limits the LF aperture to 18 mm, compared to 46.8 mm in the experiments. However, the diffraction focus for this limited aperture is more than two times deeper than the simulation depth, so the difference in LF pressure compared to the larger aperture should be negligible within the HF beamwidth.

The backscattered signals are beamformed using a simple delay and sum beamformer, using $F_{\#} = 2$.

E.5 Results

In this section results from simulations and the *in vitro* experiment is summarized.

E.5.1 Simulations

In the top panel of Fig. E.2 the beamformed image of the simulated phantom using 128 scanned plane waves, described in Sec. E.4.2, is shown. The annotation

Table E.1: Material parameters used in simulation mimicking experiment.

Region	c_0 m/s ²	ρ_0 kg/m ³	α dB/cm/MHz	β_n –	β_p GPa ⁻¹
Phantom (Agar)	1538	1030	0.49	3.7	1.52
Inclusion (Corn oil)	1468	920	0.33	6.25	3.16

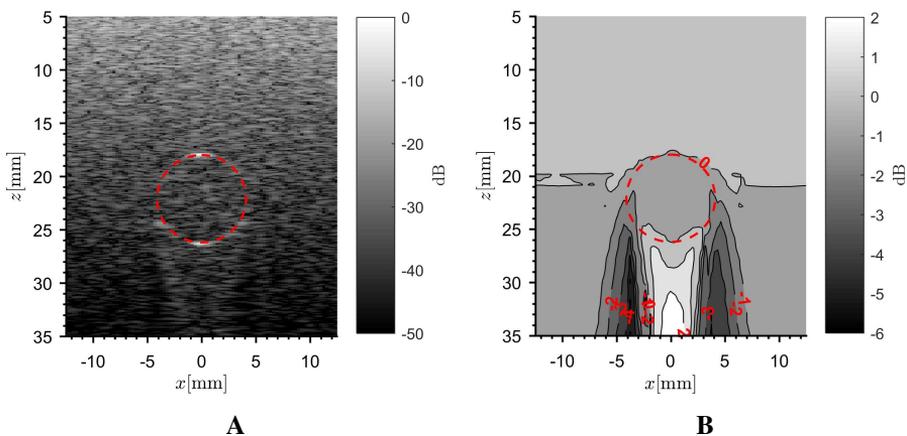


Figure E.2: Simulated phantom. In (A) B-mode image of phantom beamformed from 128 plane waves with $D_{\text{HF}} = 9\text{mm}$ transmit aperture. In (B) the measured experienced LF pressure calculated through (E.13) on the forward propagating wave. Dashed red circle indicate the location of the inclusion.

included is to highlight the outline of the inclusion. In the bottom panel, the LF pressure experienced by the HF is shown, calculated according to (E.13) on the forward propagating wave. We observe that the circular region acts as a lens which produces an in-homogeneity in the LF field behind the inclusion with two regions with a lower LF pressure along the edge of the inclusion and a focal region behind.

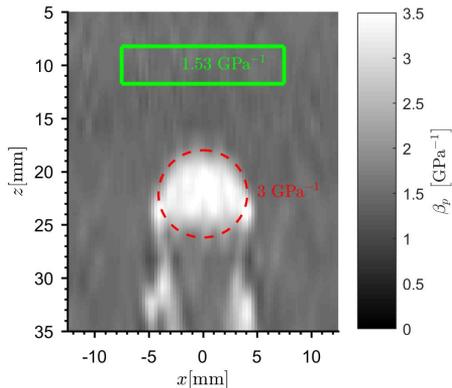


Figure E.3: Resulting β_p map for simulated phantom, calculated from (E.10), for the simulations of 128 scanned plane waves.

In Fig. E.3 the estimated nonlinear bulk elasticity map is shown, calculated from the optimization problem in (E.10). The inclusion is clearly distinguished, with values close to those in Tab. E.1. However, we do observe some artefacts behind the inclusion, especially along the edge of the inclusion.

The artefacts following the edge of the inclusion are due to two effects. Firstly, due to the wide transmit beam, the delay-and-sum beamforming causes the beamformed signal to contain echoes from parts of the beam that has propagated through the oil. Consequently, the observed signal delay becomes a weighted average of the delays corresponding to propagation through pure agar and side lobes with delays corresponding to propagation through the oil. This is seen in Fig. E.4 (A) where beyond the inclusion the observed delay along the edge becomes close to the mean value of the delays in the oil and the agar, which is shown as the dashed line. This is analogous to (E.6), where α describes the main-lobe to side-lobe level. Further, the acoustic lens effect of the oil as seen in Fig. E.2 (B) also affects the HF, which is observable in the B-mode image as a bright cone behind the inclusion. This further decreases the SNR, as along the edge the signal strength of the side lobe is high compared to the main lobe which has a local minima. Consequently, the influence of the side lobes become high causing an observed delay significantly higher than delay development corresponding to propagation through pure agar.

Secondly, the effect is further amplified as the experienced LF pressure in this region is lower, causing the wrong delay to be divided by the wrong LF pressure. These two effects combined cause the significant overestimation of β_p .

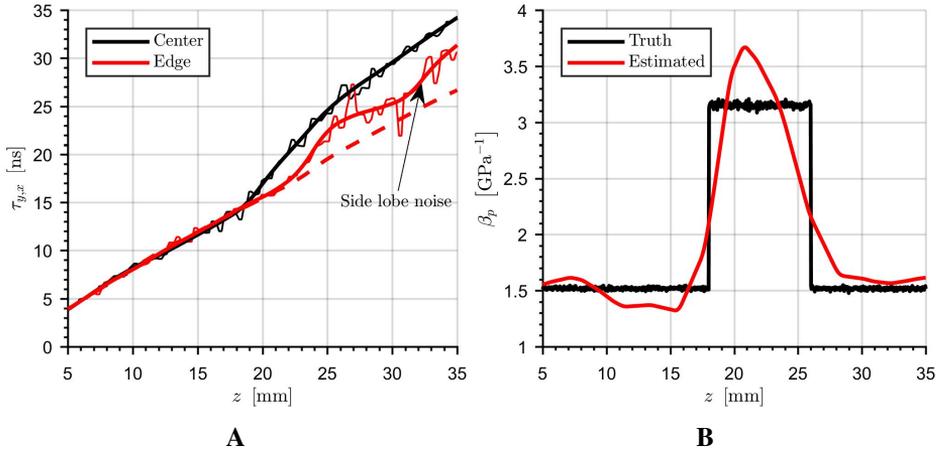


Figure E.4: Simulation results. In (A) the measured delay, τ_y , (thin) and estimated first order delay, τ_x , (thick), along the center and edge of the circular inclusion. The dashed red line show the expected delay along the edge of the inclusion. In (B) the estimated β_p along the center channel.

E.5.2 Experimental

In this section the characterization of the LF field and the experimental results on the agar phantom described in Sec. E.4.1.1 is summarized.

E.5.2.1 Low frequency field characterization

In Fig. E.5 the results of the LF field characterization in the water tank is shown. The figure shows the experienced LF pressure observed by the HF calculated through (E.13). The field is shown in terms of its deviation, in dB, from the mean value of the LF pressure in the region, $\bar{p}_{\text{LF}} = 280\text{kPa}$. Ideally, a completely homogeneous LF field is wanted, *i.e.*, $\sim 0\text{dB}$ deviation from its mean. In such a case the LF field can be approximated as a constant. However, compared with the simulated LF field in Fig. E.2 (B), the field is less homogeneous. This is mainly due to a limited elevation aperture, whereas in the simulations the field is infinitely plane due to only being a 2D simulation. In addition, an uneven electrical response of the piezoelectric elements cause a deviation from an ideal transmission, as this is an early prototype transducer.

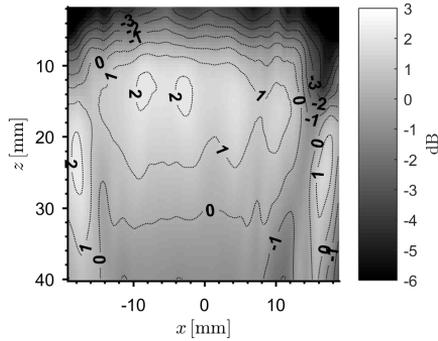


Figure E.5: The measured experienced LF pressure, calculated from (E.13). The figure shows the deviation in decibels from the average experienced manipulation pressure over the entire region, $\bar{p}_{LF} = 280\text{kPa}$.

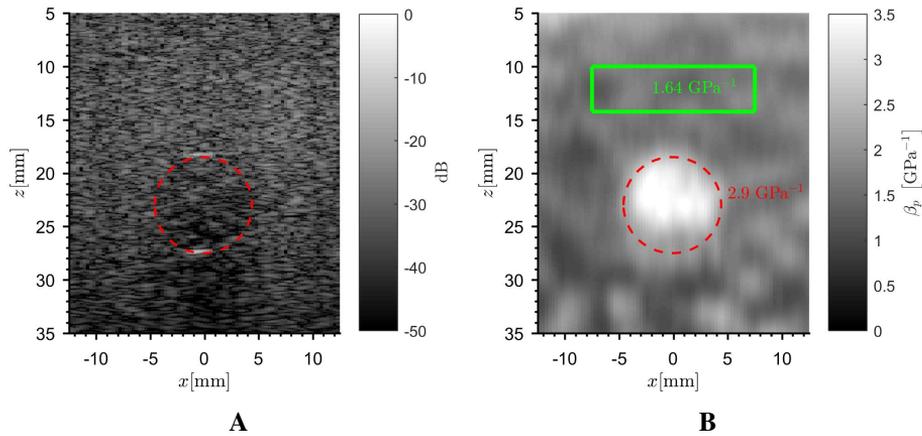


Figure E.6: *In-vitro* results with scanned plane waves. In (A), B-mode image with $D_{HF} = 9\text{mm}$ transmit aperture with corresponding β_p map in (B). Smoothed with a $3\text{mm} \times 3\text{mm}$ window.

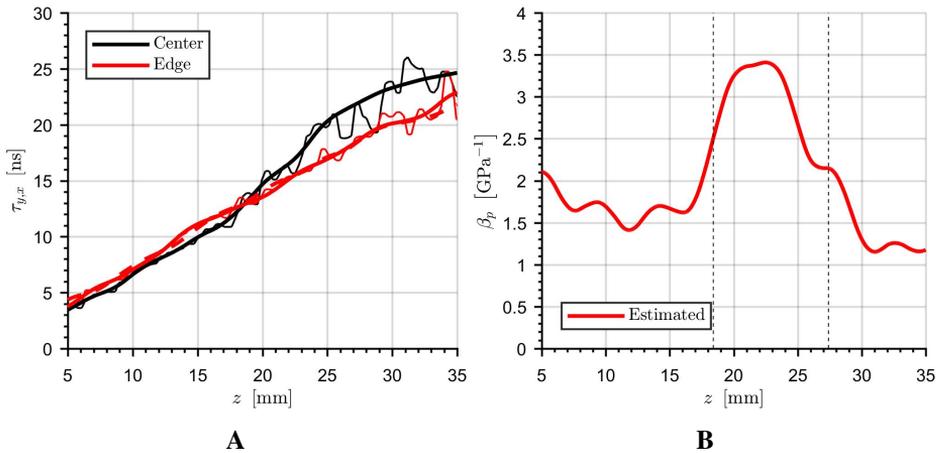


Figure E.7: *In vitro* results. In (A) the measured delay, τ_y , (thin) and estimated first order delay, τ_x , (thick), along the center and edge of the circular inclusion. In (B) the estimated β_p in a $3\text{mm} \times 3\text{mm}$ smoothed area around the center channel.

E.5.2.2 *In-vitro* results

In Fig. E.6 the B-mode image and resulting β_p map is shown of the agar-corn oil phantom described in Sec. E.4.1.1. The transmit aperture is $D_{\text{HF}} = 9\text{mm}$ for the HF and $D_{\text{LF}} = 38.4\text{mm}$ for the LF. Comparing with the simulation results in Fig. E.3 we observe a slightly higher variation in the estimate of β_p , but with less artefacts behind the inclusion. Comparing the B-mode images we observe that there is no apparent increase in the signal level behind the inclusion as we saw in the simulations. Consequently, the lens effect in the simulations is less apparent in the *in vitro* experiment.

In Fig. E.7 (A) the delay developments along the edge and center of the inclusion is shown. Compared with the development in simulations, shown in Fig. E.4 (A), the variation is substantially larger. Due to the increased inhomogeneity of the LF manipulation pressure there is a larger variation in the delay development which consequently, depending on the strength of scatterers within the main and side lobe, causes a higher variation of the observed delay. However, the clear side lobe noise seen in Fig. E.4 is not apparent, further supporting that there is little or no lens effect due to the oil. Comparing the estimates of β_p in Fig. E.4 (B) and Fig. E.7 (B) we also see a higher variation in the *in vitro* results. As we are investigating the changes in the gradients of the delays, the differentiation nature of the problem causes small errors in the delay to produce a potentially large variation in the gradient, hence amplifying the noise.

E.6 Discussion

For SURF pulse complexes there are two main effects that occur due to the interaction of HF and LF. The first effect is the development of a nonlinear propagation delay (NPD) due to an average LF manipulation pressure, $|p_{LF}| > 0$, over the HF pulse length. The other effect is pulse form distortion (PFD) due to a varying LF pressure across the HF pulse length, altering the frequency content of the HF [17, 22]. Fukukita *et al.* [15] used PFD to estimate β_n by measuring the accumulative frequency shift due to this distortion. However, this approach is more sensitive to absorption as the uncertainty of the estimate is tied to the absorption of the HF pulse which experiences a significantly higher absorption than the LF pulse. By measuring the NPD, the uncertainty due to absorption is tied to the magnitude of the LF pulse which consequently becomes lower due to the low frequency. To ensure that the main interaction effect between LF and HF is the development of a NPD, plane wave transmissions were used for both HF and LF. This ensures the most homogeneous experienced LF pressure across the HF wave front.

In this paper we have worked under the assumption that the density, speed of sound and compressibility of the material is unknown. In pulse echo imaging this is the typical case unless these are estimated. Consequently, we are unable to measure the coefficient of nonlinearity, β_n , directly. What we can measure is the nonlinear bulk elasticity, $\beta_p = \beta_n \kappa$, as the manipulated speed of sound is given as $c(p) \approx c_0(1 + \beta_p p)$. As a consequence, the results are presented in terms of the nonlinear bulk elasticity rather than the coefficient of nonlinearity.

The estimates obtained for β_p in simulations, shown in Figs. E.3 and E.4, are in good agreement with the ground truth. Correspondingly, *in vitro*, the results, shown in Fig. E.6, are close to those previously reported in literature [24]. The main artefacts in the estimates are due to variations in the NPD across the main lobe and side lobes due to in-homogeneities in the experienced LF pressure. This was apparent in simulations where the inclusion acted as a lens, producing an acoustic shadow in the main lobe along the edge of the inclusion with high side lobe levels. *In vitro* the same was observed as a larger variation in the spatial β_p map due to a less homogeneous manipulation field and consequently larger variation in NPD. Note that the results are acquired with only one plane wave angle and is hence expected to improve by introducing additional transmission angles and compounding the results. Unfortunately, due to hardware limitations, this was not possible using the current setup as it is necessary to steer LF in addition to the HF. If the LF is not steered, the observed LF manipulation pressure will be different for different angles and hence making compounding of results challenging.

The resolution and accuracy of the produced β_p maps is tied to the minimum window length that can be used in (E.10), where gradients can still be identified without being dominated by noise. This is again related to the transmit beam

and receive beamforming, and the homogeneity of the LF manipulation field. An alternative to using plane waves would be the use of focused beams which would produce a narrower transmit-receive beam, however, as we want the LF field to be plane for homogeneity, the focusing curvature of the HF would produce a lateral variation in the experienced LF pressure. This, in turn produces a variation in the delay development across the HF beamwidth, which can be detrimental to β_p estimation. Transmit apodization could also be applied to decrease side lobes, but was not possible due to limitations in the prototype scanner. Single plane wave transmissions is possible, but produces a slightly higher variation in the observed delays likely due to the current inhomogeneity in the LF field from the prototype transducer. However, the potential for high frame rates imaging makes the route of plane wave imaging interesting.

The *in vitro* results are expected to be mostly qualitative. In the simple experiment presented here the approach shows a quantitative potential, but in more heterogeneous structures, *in vivo*, the uncertainty of the LF field increases. The transducer and scanner used are still early prototypes and to produce robust quantitative estimation, optimal transducer design for homogeneous manipulation, multiple angle acquisitions and LF modeling becomes increasingly important. In both simulations and experiments a simple delay-and-sum beamforming strategy was employed, hence the potential for utilizing alternative beamforming strategies should be investigated.

Estimation of β_p from the nonlinear propagation delay in (E.9) is challenging as it involves calculating the derivative of a noisy signal. The optimization definition in (E.10) is a beneficial way of formulating the problem as constraints and likelihood estimates can be included easily. However, the processing is quite heavy, especially for fine resolution. This could possibly be circumvented by choosing the regions, z_i , in an adaptive fashion, allowing the algorithm to choose image lines which require higher or lower resolution. In addition, a weakness is the two step process of first estimating τ_y as described in Sec. E.3.1 and then finding β_p through (E.10). The success of the fitting is dependent on the quality of the input and hence seeking ways of combining the two into a single operation could be beneficial.

The loss weighting in (E.11) acts as a trustworthiness estimate for the delay and was in this paper simply chosen as an interface detector. However, more advanced weighting could be included to further suppress the influence of noise on the estimates. An example of such a weight could be a reverberation noise detector as in [20].

Although the observed delay is noisy, for an experienced human operator, identifying the most likely gradients is easy. Formulating these mathematically however, is hard, due to the complicated physics involved. Consequently, it is the

authors believe that the optimization problem is an ideal application for supervised machine learning or deep learning approaches[28]. However, producing *in vivo* or *in vitro* training data would be the main challenge. Perhaps such data can be generated synthetically using k-Wave[27]. Nonetheless, such an approach is a promising avenue of future work.

E.7 Conclusion

In this paper a dual frequency approach for pulse echo determination of the non-linear acoustic properties of tissue using scanned plane waves, transmitted by a single linear array transducer, was presented. Only a single plane wave angle was utilized and a spatial nonlinearity map produced. The method shows good qualitative results with a potential for quantitative estimates, however challenges due to side lobes and uncertainty of the manipulation pressure was observed.

Acknowledgements

The authors acknowledge support from the Norwegian Research Council, project number 254633.

Bibliography

- [1] M. L. Oelze and J. Mamou. “Review of Quantitative Ultrasound: Envelope Statistics and Backscatter Coefficient Imaging and Contributions to Diagnostic Ultrasound.” *IEEE Transactions Ultrasonics, Ferroelectrics, and Frequency Control*, 63(2), **2016**. doi: 10.1109/TUFFC.2015.2513958.
- [2] F. Duck. “Tissue non-linearity.” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):155–170, **2010**. doi: 10.1243/09544119JEIM574.
- [3] J. Zhang, M. S. Kuhlenschmidt, and F. Dunn. “Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A.” *The Journal of the Acoustical Society of America*, 89(1):80–91, **1991**. doi: 10.1121/1.400371.
- [4] M. J. Choi, S. R. Guntur, J. M. Lee, et al. “Changes in ultrasonic properties of liver tissue *in vitro* during heating-cooling cycle concomitant with thermal coagulation.” *Ultrasound in Medicine and Biology*, 37(12):2000–2012, **2011**. doi: 10.1016/j.ultrasmedbio.2011.06.015.
- [5] D. Zhang and X.-F. Gong. “Experimental investigation of the acoustic nonlinearity parameter tomography for excised pathological biological tissues.” *Ultrasound in Medicine & Biology*, 25(4):593–599, **1999**. doi: 10.1016/S0301-5629(98)00185-9.

-
- [6] X. Gong, D. Zhang, J. Liu, et al. “Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues.” *The Journal of the Acoustical Society of America*, 116(3):1819–1825, **2004**. doi: 10.1121/1.1781709.
- [7] M. Nikoonahad and D. C. Liu. “Pulse-Echo Single Frequency Acoustic Nonlinearity Parameter (B/A) Measurement.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 37(3):127–134, **1990**. doi: 10.1109/58.55301.
- [8] M. Fatemi and J. F. Greenleaf. “Real-time assessment of the parameter of nonlinearity in tissue using ‘nonlinear shadowing’.” *Ultrasound in Medicine and Biology*, 22(9):1215–1228, **1996**. doi: 10.1016/S0301-5629(96)00140-8.
- [9] A. Kato and Y. Watanabe. “Measurement Method of Spatial Distribution of Nonlinearity Parameter B/A Using Nonlinear Interaction of Two Sound Waves.” *Japanese Journal of Applied Physics*, 32(Part 1, No. 5B):2274–2278, **1993**. doi: 10.1143/JJAP.32.2274.
- [10] N. Ichida, T. Sato, and M. Linzer. “Imaging the Nonlinear Ultrasonic Parameter of a Medium.” *Ultrasonic Imaging*, 5(4):295–299, **1983**. doi: 10.1016/S0161-7346(83)81002-5.
- [11] N. Ichida, T. Sato, H. Miwa, and K. Murakami. “Real-Time Nonlinear Parameter Tomography Using Impulsive Pumping Waves.” *IEEE Transactions on Sonics and Ultrasonics*, 31(6):635–641, **1984**. doi: 10.1109/T-SU.1984.31548.
- [12] F. Varray, O. Basset, P. Tortoli, and C. Cachard. “Extensions of nonlinear B/A parameter imaging methods for echo mode.” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 58(6):1232–1244, **2011**. doi: 10.1109/TUFFC.2011.1933.
- [13] B. Angelsen and R. Hansen. “SURF Imaging - A New Method for Ultrasound Contrast Agent Imaging.” In *2007 IEEE Ultrasonics Symposium Proceedings*, pages 531–541. IEEE, **2007**. ISBN 978-1-4244-1383-6. doi: 10.1109/ULTSYM.2007.140.
- [14] R. Hansen, S.-E. Måsoy, T. F. Johansen, and B. a. Angelsen. “Utilizing dual frequency band transmit pulse complexes in medical ultrasound imaging.” *The Journal of the Acoustical Society of America*, 127(1):579–87, **2010**. doi: 10.1121/1.3257584.

- [15] H. Fukukita. “Ultrasound pulse reflection mode measurement of nonlinearity parameter B/A and attenuation coefficient.” *The Journal of the Acoustical Society of America*, 99(5):2775, **1996**. doi: 10.1121/1.414873.
- [16] O. F. Myhre, T. F. Johansen, and B. A. Johan Angelsen. “Analysis of acoustic impedance matching in dual-band ultrasound transducers.” *The Journal of the Acoustical Society of America*, 141(2):1170–1179, **2017**. doi: 10.1121/1.4976096.
- [17] R. Hansen, S.-E. Måsøy, T. A. Tangen, and B. A. Angelsen. “Nonlinear propagation delay and pulse distortion resulting from dual frequency band transmit pulse complexes.” *The Journal of the Acoustical Society of America*, 129(2):1117–27, **2011**. doi: 10.1121/1.3518753.
- [18] T. Rommetveit, T. F. Johansen, J. Deibele, H. Kaupang, and B. Angelsen. “Two way nonlinear manipulation in plane materials using dual frequency pulse complexes.” *Proceedings - IEEE Ultrasonics Symposium*, pages 2380–2383, **2010**. doi: 10.1109/ULTSYM.2010.5935562.
- [19] J. M. Rau, S.-E. Måsøy, R. Hansen, B. Angelsen, and T. A. Tangen. “Methods for reverberation suppression utilizing dual frequency band imaging.” *The Journal of the Acoustical Society of America*, 134(3):2313–25, **2013**. doi: 10.1121/1.4817900.
- [20] O. M. Brende and B. Angelsen. “Adaptive reverberation noise delay estimation for reverberation suppression in dual band ultrasound imaging.” *The Journal of the Acoustical Society of America*, 138(5):3341–51, **2015**. doi: 10.1121/1.4935555.
- [21] E. Flørenæs, S. Solberg, J. Kvam, et al. “In vitro detection of microcalcifications using dual band ultrasound.” In *IEEE International Ultrasonics Symposium, IUS, 2017*. ISBN 9781538633830. doi: 10.1109/ULTSYM.2017.8092857.
- [22] O. F. Myhre, J. Kvam, and B. A. J. Angelsen. “Dual Frequency Transducer Design for Suppression of Multiple Scattering.” In *Ultrasonics Symposium (IUS), 2016 IEEE International*, pages 8–11, **2016**. ISBN 9781467398978.
- [23] IEC 60601-2-37:2007. “Medical electrical equipment - Part 2-37: Particular requirements for the basic safety and essential performance of ultrasonic medical diagnostic and monitoring equipment.” International standard, International Electrotechnical Commission, Geneva, CH, **2007**.

- [24] F. Dong, E. L. Madsen, M. C. MacDonald, and J. A. Zagzebski. “Nonlinearity parameter for tissue-mimicking materials.” *Ultrasound in Medicine and Biology*, 25(5):831–838, **1999**. doi: 10.1016/S0301-5629(99)00016-2.
- [25] R. Chanamai and D. J. McClements. “Ultrasonic attenuation of edible oils.” *Journal of the American Oil Chemists’ Society*, 75(10):1447–1448, **1998**. doi: 10.1007/s11746-998-0198-1.
- [26] IEC 62127-1:2007. “Ultrasonics - Hydrophones - Part 1: Measurement and characterization of medical ultrasonic fields up to 40MHz.” Standard, International Electrotechnical Commission, Geneva, CH, **2007**.
- [27] B. E. Treeby and B. T. Cox. “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields.” *Journal of Biomedical Optics*, 15(2):021314, **2010**. doi: 10.1117/1.3360308.
- [28] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” *Nature*, 521(7553): 436–444, **2015**.

Part III

Complete Index

List of Tables

A.1	Description of input, output predictions, and training data for each CNN.	47
A.2	Classification and regression results for spine classification.	55
A.3	Classification results for segmentation of limbs.	56
A.4	Classification results for classification of limbs.	57
C.1	Materials used for evaluation, sorted by increasing coefficient of nonlinearity, β_n . Data taken from [9] where materials with a valid entry for β_n is included.	76
C.2	Correlation coefficient and mean average percentage error (MAPE) for empirical regression.	83
D.1	High Frequency (HF) array parameters.	97
D.2	Low Frequency (LF) array parameters.	97
E.1	Material parameters used in simulation mimicking experiment.	115

List of Figures

1.1	Variation of acoustic parameters in soft tissue. Gaussian fit to data in [16].	7
1.2	A positive and negative SURF pulse composed of a 8 MHz high frequency (HF) imaging pulse and a 0.8 MHz low frequency (LF) manipulation pulse.	10
1.3	Example of development of first-order delay, τ_x , noise delay, τ_n and the measured signal delay τ_y	14
1.4	Simple neural network	19
1.5	Typical training and validation loss development during supervised training.	22
A.1	Combination of sagittal and coronal projections to produce a 3D volume segmentation.	44
A.2	Segmentations are done in a waterfall fashion. The segmented volume is removed from the CT volume to reduce complexity of input images to the following step. Each row represents a task which is handled by its own CNN. The CNNs operate on 2D projections of the CT volume (left column), and the output from the CNN (middle column) is combined to produce a segmentation in the 3D volume (right column). Green and yellow frames indicate final and intermediate results, respectively.	46
A.3	Visualization of class separation using t-distributed stochastic neighbor embedding. The classes shown are for the limb-segmentation task. Left and right limbs are distinguishable in the coronal view but indistinguishable in the sagittal view.	49
A.4	CNN used for segmentation tasks. The network is based on the U-net [17] with added dropout. In addition, an optional regression head is added for segmentation tasks involving counting of class labels.	50

A.5	Training curves for classification and regression for the spine classification task. Whole lines show the training loss, and dashed lines show the validation loss. An epoch is a measure of the number of times all the training examples have been used to update the weights of the network.	54
B.1	Distribution of available data. The left panel shows the total distribution with origin of specimens. The right panel shows a typical training and validation set.	65
B.2	Architecture of the CNN. The network is in total 16 layers deep with two inception blocks.	66
B.3	Generalization of deep CNN on independent test set. The left panel shows the performance over the entire test set ($n = 30$ animals). The right panel shows the performance on the animals with a CIMF $< 6\%$, ($n = 23$ animals).	67
C.1	Variation of parameters β_n , κ_s , β_p and B relative to their respective minima. Sorted by increasing β_n	76
C.2	(A) Relative errors of models for the coefficient of nonlinearity, β_n , given in eqs. (C.13a) and (C.13b). (B) Relative errors of models of the measurable nonlinear response of the medium, β_p , given in eqs. (C.14a) and (C.14b).	83
C.3	Relationships based on thermodynamic model. The solution to optimization problem in eq. (C.37), yielding the isobaric thermal expansion coefficient in (A) and the short-range force exponent n in (B). Based on the linear relationship between n and κ_s , and α and κ_s the modeled regression in eq. (C.39) is shown in (C) and the corresponding model for the measurable nonlinear response of the medium, β_p , in (D).	84
D.1	SURF pulse complex with a 18:1 HF-to-LF frequency ratio. The two HF pulses show the positioning of the HF pulses at the center (0.0 mm) and edge (5.5 mm) of the transmit aperture. The F-number is 2, and the focus is at 22 mm.	94
D.2	The accumulated PFD (top) and NPD (bottom) for LF element heights $L_{y,LF} = \{5.1 \text{ mm}, 10.7 \text{ mm}, 13.77 \text{ mm}\}$ corresponding to $\{100\%, 210\%, 270\%\}$ of the HF element height respectively. . . .	98
D.3	Front view of the probe radiation apertures. The LF aperture is larger than the HF aperture, and located behind it. In practice, the area in front of the part of the LF aperture which is outside the HF aperture is covered with inactive HF elements.	99

D.4	Cross-sectional overview of the stack structure, showing the different sections. Each section may contain multiple layers of material. The dashed lines show the active part of the HF section, which coincides with the elevation width of the lens.	99
D.5	Normalized HF backing impedance (top) and normalized LF loading impedance (bottom) when using a single isolation layer (thin solid), when using three layers with alternating low and high impedance (dashed), and when using three layers with the thickness of the third layer tuned to achieve an impedance resonance at $\omega_T = 1/18$ (thick solid).	101
D.6	Simulated (dashed) and measured (solid) transfer functions of the LF (left) and HF (right) arrays of the Vora-T probe.	102
E.1	SURF pulse complex with 10MHz HF and 1MHz LF.	107
E.2	Simulated phantom. In (A) B-mode image of phantom beam-formed from 128 plane waves with $D_{HF} = 9\text{mm}$ transmit aperture. In (B) the measured experienced LF pressure calculated through (E.13) on the forward propagating wave. Dashed red circle indicate the location of the inclusion.	115
E.3	Resulting β_p map for simulated phantom, calculated from (E.10), for the simulations of 128 scanned plane waves.	116
E.4	Simulation results. In (A) the measured delay, τ_y , (thin) and estimated first order delay, τ_x , (thick), along the center and edge of the circular inclusion. The dashed red line show the expected delay along the edge of the inclusion. In (B) the estimated β_p along the center channel.	117
E.5	The measured experienced LF pressure, calculated from (E.13). The figure shows the deviation in decibels from the average experienced manipulation pressure over the entire region, $\bar{p}_{LF} = 280\text{kPa}$	118
E.6	<i>In-vitro</i> results with scanned plane waves. In (A), B-mode image with $D_{HF} = 9\text{mm}$ transmit aperture with corresponding β_p map in (B). Smoothed with a $3\text{mm} \times 3\text{mm}$ window.	118
E.7	<i>In vitro</i> results. In (A) the measured delay, τ_y , (thin) and estimated first order delay, τ_x , (thick), along the center and edge of the circular inclusion. In (B) the estimated β_p in a $3\text{mm} \times 3\text{mm}$ smoothed area around the center channel.	119

