Intelligent speaker

Master thesis

B&W Group Ltd
Research & Development
Elm Grove Lane
Steyning, West Sussex
BN44 3SA
United Kingdom
http://www.bowers-wilkins.co.uk/

Company mentor: Stuart Nevill
Head of Advanced Engineering
stuart.nevill@bowerswilkins.com

Delft University of Technology
Faculty Industrial Design Engineering
Landbergstraat 15
2628 CE Delft
The Netherlands
http://www.io.tudelft.nl/en/

Project chair: Prof. ir. J. van Erp
Department: Industrial Design
Section: Design Conceptualization
and Communication
J.vanErp@tudelft.nl

Project mentor: Ir. E.W. Thomassen
Department: Design Engineering
Section: Product Architecture Design
E.W.Thomassen@tudelft.nl

Roel Stein
4015657
December 14, 2017
Delft University of Technology
Faculty of Industrial Design Engineering
Integrated Product Design (MSc)
+31 624100747
roelstein@gmail.com

Special thanks to
Jon Moore, Jasper Brekelmans.

“We’re creating a world in which devices bend toward the user,

instead of users bending toward devices.”

Inon Beracha, PrimeSense chief executive

figure 1 B&W Zeppelin Wireless in context

Executive summary

British premium loudspeaker manufacturer Bowers & Wilkins hit the spot in a new category with the Zeppelin. Today, severe competition is driving the brand to ‘up-the-game’ for connectivity and user experience. B&W is a dominant player in the conservative Hi-Fi market. With a rich heritage and a strong brand image it has successfully expanded to adjacent markets. Typically, B&W carefully analyses new market explorers and then challenges the state-of-art through extensive research and development. In-store observations reveal that buyers of lifestyle products want 'instant everything'. B&W's most important discriminator 'sound quality' will only affect the consumers' experience on the long term. For in-store persuasion B&W's lifestyle products should be equipped with an iconic next generation user-product interaction.

The intelligent speaker is a 'concept speaker' that will serve as a conversation starter. It should explore possibilities and inspire. The concept speaker will embody the intended B&W interaction of the future as well as the acoustic aspiration of giving a consistent experience in any place in any room. Equipped with a circular array of drivers it will be able to change its acoustic directivity and beam width.

The objective of this project was to: "Design a unique B&W user experience for the intelligent speaker". In order to materialise the crafted vision an agile approach has been adopted. Repetitive user tests quickly gave insights into the needs and values of the intended consumer. Gesture control is the way forward for simple navigation through a music library or playlist. Voice control is the essential addition to assist the music lover with her music preferences.

The foremost aesthetic attribute is honesty. The form should 'tell the story' and put the focus on the right areas of the product. The final form communicates the principle of a directive array of drivers that is assisted by a powerful bass unit. The shape induces the consumer to position the speaker as a centrepiece to benefit from the acoustic novelties to its full extent.

The thesis concludes with a model that can be used for further testing of the interaction. The model has the intended form and finish of the intelligent speaker with a facetted grille covered with cloth. Voice- and gesture control can be mimicked as well as the intended visual feedback. The shape is designed to accommodate a core in a later stage for acoustic measurements.

The concept should inspire to develop a lifestyle speaker featuring the B&W interaction of the future. The interaction could well be adapted cross-platform. If there is a strong market pull for directivity- and beam width control this can be introduced in a separate product.

Contents

Glossary

Amp	Amplifier
BOM	Bill of materials
B&W	Bowers & Wilkins
Concept speaker	A concept that is used to explore or communicate a company's vision on future developments
Design	Widely used terminology to describe the process of deciding how something will be made or look
Directivity	The ability to control the direction of the emitted sound
Driver	The part of a loudspeaker that converts electric current into physical movement, and thus movement of air
DSP	Digital Signal Processing
FoV	Field of View
HF	High frequency sound
Hi-Fi	High fidelity, the pursuit of realistic reproduction of sound
Industrial Design	Study to develop durable products and services for people, on the basis of the balanced interests of users, industry and society.
IoT	Internet of Things, the network of physical devices
IR	Infrared
LF	Low frequency sound
MF	Mid frequency sound
Product concept	Developments that have the aim to result in new product introductions
SDK	Software Development Kit
SPL	Sound Pressure Level
SRE	B&W's Steyning Research Establishment
Sweet spot	Terminology used in Hi-Fi to describe the ideal listening position
Wireless speaker	A loudspeaker that is wirelessly connected to the music source whilst still using a power cord

Introduction

figure 2 B&W 802, one of B&W's most prestigious speakers

Introduction

British premium loudspeaker manufacturer Bowers & Wilkins hit the spot in a new category with the Zeppelin. Today, severe competition is driving the brand to 'up-the-game' for connectivity and user experience.

The conventional market of high-end audio entertainment has been disrupted by the advent of smartphones and tablets. The way of listening to music has undoubtedly changed to a less involved manner. The functionality of speakers has evolved consequently. British speaker manufacturer Bowers & Wilkins (B&W) was the first company to launch a high-end docking station in 2007. Ten years later the company is in an extremely competitive market of wirelessly controllable speakers and intends to develop a new statement piece.

Since 1966 British speaker manufacturer Bowers & Wilkins (B&W) has been known for creating high-end Hi-Fi stereo speakers. Their top of the range products are being adored in the recording industry for their excellent performance. It is with a reason

figure 3 Left: Zeppelin Docking, right: Zeppelin Wireless

that they are currently controlling over 50% of the worldwide Hi-Fi market. The mid-size company aims at up-market users worldwide.

Design challenge
With the trend of digital entertainment, a completely new product category emerged to serve these unconventional use cases. Wireless connectivity became essential within this category and the battle between manufacturers for supreme user-experience was on. B&W's product portfolio nowadays includes 'wireless' speakers and headphones, internally addressed to as the New Media category. The New Media customer is different from the previously addressed customer and desires 'instant everything'. To some extent, the ease of listening and the user experience take priority over the actual sound quality for this type of products.

Competitiveness in this lucrative market has increased over the years. In order to create awareness amongst the New Media customer B&W developed an 'Intelligent speaker' that makes a statement in every possible way. This speaker features, next to astonishing sound quality, the capability to actively sense changes in its environment. The big question was: how can B&W create leverage from the brand equity they have to differentiate in this highly competitive market?

Assignment

Whereas the acoustic architecture is being developed by B&W's acoustic engineers, the market positioning, user-product interaction and form were to be explored in this dissertation. The objective of this project was to:

"Design a unique B&W user experience for the intelligent speaker
- for the New Media target group and domestic use."

Approach

The effort resulted in a concept speaker that showcases the possibilities with regards to the current market developments. This concept serves as the basis for the intelligent speaker and as inspiration for range-wide developments.

The starting point of this thesis was a fundamental analysis on the brand's equity. The brand's strengths were then translated into a bespoke vision - not just copying the competitors' activities. The vision was justified by repetitive tests and resulted in a future user-product interaction for range-wide application of all of B&W's lifestyle speakers. To demonstrate the potential of this future user-product interaction it was then demonstrated using an application with a steerable array of drivers; *the concept intelligent speaker.*

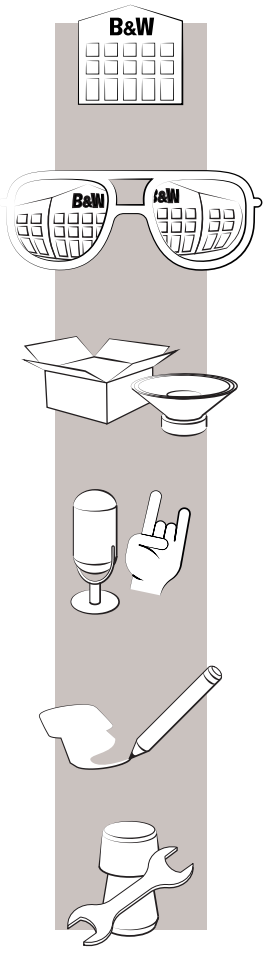


figure 4 Flowchart of report structure

Understanding brand equity

figure 5 Abbey Road Studios adopts the B&W Matrix 801 in its studios in 1988.

Understanding brand equity

B&W is a dominant player in the conservative Hi-Fi market. With a rich heritage and a strong brand image it has successfully expanded to adjacent markets. Typically, B&W carefully analyses new market explorers and then challenges the state-of-art through extensive research and development.

Content

Bowers & Wilkins' profile

Bowers & Wilkins' headquarter is based in Worthing, West Sussex on the south coast of England. The company is funded in 1966 by John Bowers whose dream was to manufacture the perfect loudspeaker. Ever since B&W's speakers help people to experience listening to music better.

Bowers & Wilkins is known for its award-winning Hi-Fi stereo speakers and used its reputation to enter adjacent markets like custom installation speakers and automotive audio (see p. 12). The brand image is imbued with 'studio quality' audio, although they do not manufacture speakers for recording studios.

In 1981 the research and development department was deliberately separated from the manufacturing plant to be able to think more freely. Industrial design has been carried out in collaboration with UK's leading designers since its early years which resulted in iconic designs like the Emphasis (1989), Nautilus (1993) and Zeppelin (2007). Nowadays the company has a global presence with strong British roots.

figure 6 From left to right: Emphasis (1989), Nautilus (1993) and Zeppelin (2007)

Mission statement

B&W's last revision of their mission statement gives an accurate description of their core values:

"To deliver an unsurpassed "true sound" experience to discerning listeners wherever they may be: In the studio, at home, or on the move."

A "true sound" experience is the key differentiator B&W leverages from. Although the experience is ultimately a matter of taste, the tech-driven company assures the most accurate representation of sound through extensive research and development.

B&W's products are positioned up-market and appeal to those who share the passion, i.e. discerning listeners. It is a niche market that is constant and consists of customers that are loyal to their most favourite brand.

The third element is a bit of a concern as it does not restrict the activities to certain markets. B&W has a broad product- and market domain and is still expanding their product offer. This lack of restriction, as well as the tech-push, result in a somewhat scattered portfolio (see p. 12).

Competitive strategy

Method
Miles and Snow business strategies

According to the Miles and Snow business strategies theory, B&W's competitive strategy would best match the description of 'analyser' which in practice means it will carefully analyse new market explorers and developments before stepping into these new markets. Once nested in a market the company challenges the state-of-art through research and development wherever they can. Results of this are the iconic diamond tweeter and Kevlar midrange, the 'Nautilus' tapering tubes and the de-coupling of drivers.

Furthermore, Miles and Snow say an 'analyser' typically focuses on product performance, not on process optimisation. B&W relaunches their traditional Hi-Fi products every 5 years and New Media products every 4 years. When competing in the New Media market the product development time could be a bottleneck. A change in organisation and approach should be considered to anticipate on that.

Focus on a certain product- or market domain could allocate resources more efficiently. B&W will try to remain in certain markets by market penetration with a broad product-market domain and as a result manufactures over 80 different products as of this writing (B&W, 2016a).

figure 7 Top: diamond tweeters and Kevlar midrange driver, bottom: the 'Nautilus' tapering tubes and the de-coupling of drivers.

Product portfolio

Overview

Bowers & Wilkins is renowned for its studio quality Hi-Fi that nowadays makes up for roughly one-third of total revenues. An almost equally important market is custom installation speakers. These cannot be found in stores but are brought to the end-user by third party installers and architects.

All other products together count up for roughly one-third of total revenues. B&W is also supplying the automotive industry with premium audio for various models of McLaren, Maserati, BMW and Volvo; an excellent opportunity for exposure to what might be a potential user of the core products.

The B&W product portfolio is plotted for the typical involvement of each product (see fig. 8). The lower on the y-axis the more accessible products are to the user. The products are explained from left to right.

Passive Hi-Fi speakers have a domestic use-case and have an involved manner of interaction where one has to sit down and listen carefully. Headphones have a very intimate, involved and individual interaction but are not bound to a static position. Custom Theatre speakers provide a private cinema experience. The Theatre solutions are the more accessible products for a home cinema experience consisting of a soundbar and satellite speakers assisted by a subwoofer. Bowers & Wilkins speakers can be chosen as top tier option in various premium cars. The MM-1 computer speakers are hardwired stereo speakers. In-ceiling and in-wall speakers are available in all sizes. Marine speakers are the water-resistant installation speakers. Outdoor speakers are installed in public spaces both in- and outdoors and provide ambient music. Last but not least, integrated 'wireless' music systems have a versatile use-case and serve a less involved and more social purpose.

The Intelligent Speaker project focusses on this last category of 'wireless' speakers. With the discontinuation of three different 'wireless' speakers, i.e. the Z2, A5 and A7, the product offer for 'wireless' speakers drastically decreased down to two products: the 10-year-old but still very relevant Zeppelin and the portable T7 Bluetooth speaker. The planned developments for this category are plotted over time on the next page.

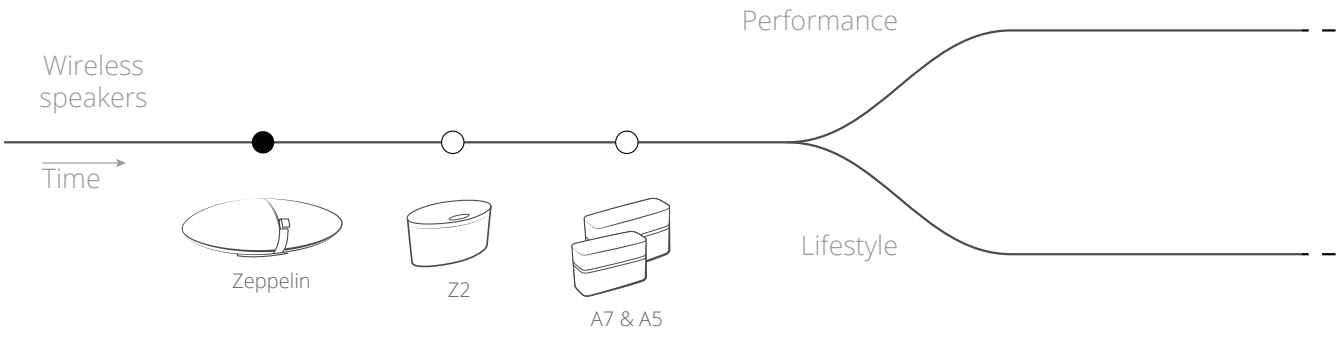
figure 8 B&W's current portfolio. X-axis the user's involvement. Y-axis accessibility.

Portfolio roadmapping for New Media Speakers

Before moving forward, the taken steps should be considered first. A decade ago, in 2007, B&W introduced the first Zeppelin; a docking station for the iPod. The Zeppelin was not the first docking station for iPods, but it outperformed all existing docking stations by a long way. It was a great success selling well over 300,000 units. More discrete models followed that were less of a centre piece, i.e. the Z2 docking station, A5 and A7. As mentioned before the latest Zeppelin is the only model standing along with the T7 Bluetooth speaker.

Soon this category will be expanded. B&W's R&D department is responding to the market demand for 'wireless' audio in a brand-fashioned way by developing a range of active products to present a new level for high-end 'wireless' audio. The product range will consist of both performance stereo solutions as well as more lifestyle oriented single box products.

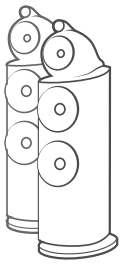
The product strategy for 'wireless' speakers is plotted over time (see fig. 9) with the near-future on the left side and time increasing further down the line. Black dots are for essential products with a strong message, white dots are for necessary additions to create a complete product offer.



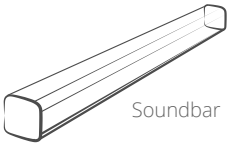
The lifestyle speakers are all single box products that are unique for their modularity. Forced by market pull, B&W want to offer all common products, i.e. a Soundbar, Subwoofer, a mini speaker and an updated version of the iconic Zeppelin.

The new intelligent speaker will make a statement in technology within the lifestyle products. It must be a one-box solution and must sound at least equally as good as the latest Zeppelin. Because of its added intelligence features it will sell at a higher price point than the $699 Zeppelin. B&W identified a gap in the market for the Intelligent Speaker at a price point between $1500 and $2000.

Potentially more high-end flagship products will be added to both the performance- and the lifestyle range. In any case this is long term thinking.



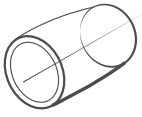
Active 802



Soundbar

Intelligent Speaker

Flagship?



NSW

N1

Zeppelin

figure 9 Product strategy for 'wireless' speakers

Market segment

Customer segment

Because of positioning and price point B&W is serving a niche-market of people that share the passion for 'true sound'. This segment is very constant and has not significantly changed over the last decade.

Customer research of independent research agencies in 2006 and 2015 reveals that B&W has consistently been addressing the same target group over the last ten years. The brand did not 'age' with its audience.

B&W's conventional Hi-Fi products generally appeal to a wealthy male buyer with a mean age of 42 years old. Surprisingly, B&W's 'wireless' speakers seem to appeal to a similar buyer but show a little more accessibility.

GENDER

It is no secret that Hi-Fi is mostly a men's interest. Visit an audio show and the number of female visitors can be counted on a single hand. However, market research agency 2CV (2015b) revealed that the gender split in the 'wireless' audio category is 50/50 which implies there is a rich territory to win.

Amongst Hi-Fi enthusiast "WAF" is an acronym for "Wife Acceptance Factor". The lady of the house has to approve the looks and expenses of her spouse's hobby.

So, what is the reason that 95% of B&W's 'wireless' audio users are men? B&W's headphones have shiny metal details and the 'wireless' speakers come in sombre colours. The masculine appearance appeals more to male buyers. A universal appeal will be vital for future growth in the New Media segment.

AGE

Consumers buying into headphones and integrated wireless music system products in general tend to be a younger buyer, but not for Bowers & Wilkins. B&W's Product Strategy Manager Andy Kerr explained: "Much as we would like to be selling our headphones to 23-year-olds, the truth is our headphones probably sell to 43-year olds. Which as an average is slightly below, but not far below the age group that our hi-fi sells into." (Personal conversation Andy Kerr, product strategy manager B&W, February 8, 2016). "Bentley don't waste their time making a car that is designed to appeal to 21-year-olds. Cause you can't make a Bentley that a 21-year-old can afford, unless he plays for Chelsea."

Andy's vision for where Bowers & Wilkins can actually drive and survive would be: "*when you're ready for us, we are ready for you*."

GEOGRAPHIC DATA

In 50 years B&W managed to establish a dealer network with global presence. An estimation of the global reach of the intelligent speaker can best be made by studying the Zeppelin's sales figures (see fig. 10). Sales of the successful Zeppelin shows a lesser global reach compared to the overall figures; 64% is harvested in Europe (B&W, 2014). There is a big opportunity to expand in the USA and China, respectively equivalent to just 14% and 5% of sales of all Zeppelin models, whilst the countries account for respectively 22% and 10% of the world's GDP (Davies et al., 2009).

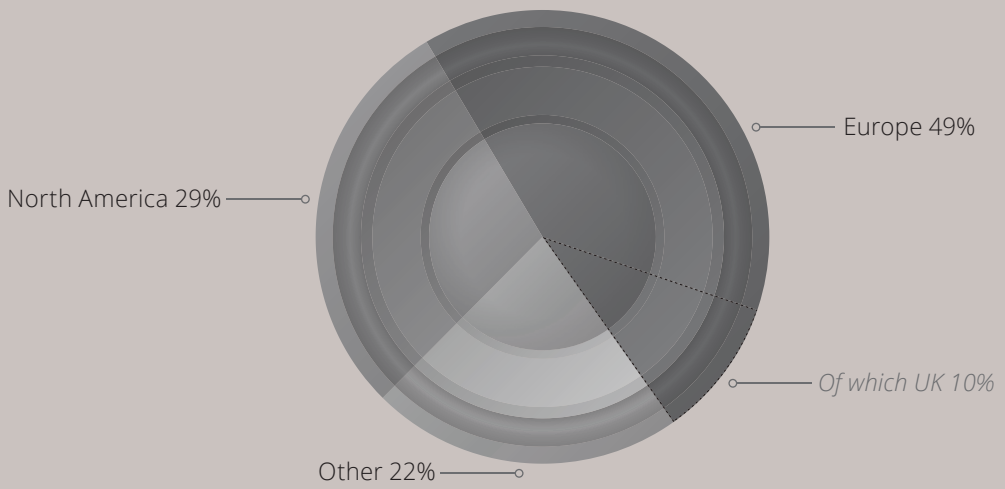
CUSTOMER SATISFACTION

B&W's focus on performance and its attention for build quality result in a high brand appreciation. Their customers are extremely loyal. 90% Of current owners reconsider the brand for their next purchase (2CV, 2015b). This is remarkably high compared to other brands.

- Bowers & Wilkins 90%
- Bose 70%
- Beats 70%
- B&O 63%

Overall 2014

(B&W, 2014)



Zeppelin (2009-2016)

(B&W, 2016b)

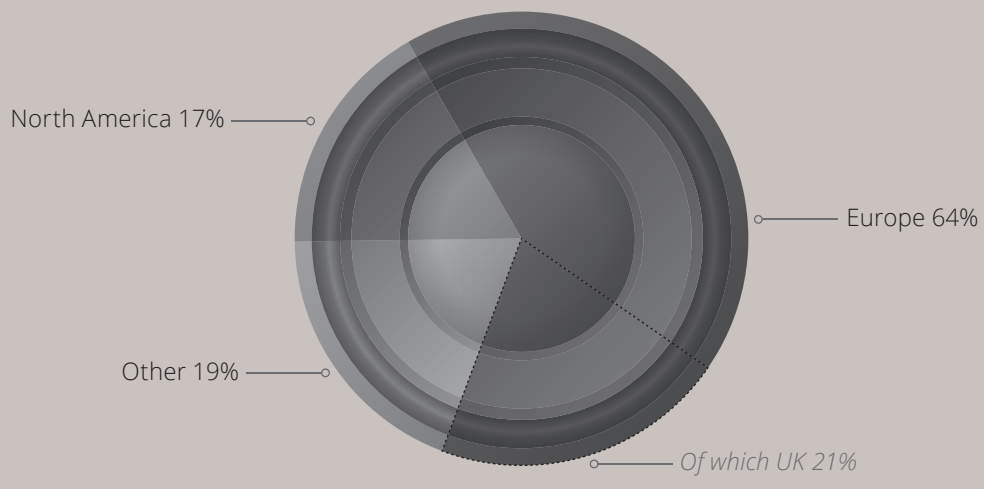


figure 10 B&W's regional sales numbers

Channels and marketing

For the premium speaker industry in the UK and USA 71% of the people ultimately buy the product in store. Of the 29% of the people that buy online 66% visited a store earlier for research purposes. So only 10% does not visit a store in their buying process (2CV, 2015a).

Purchase premium market

(2CV, 2015a)

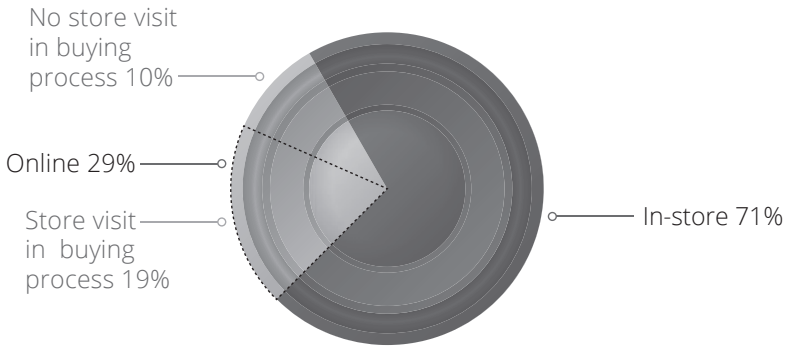


figure 11 Bricks versus clicks

Therefore, physical stores will still be very relevant for the sales of the intelligent speaker. Currently, B&W's physical presence in department stores is modest. B&W's product presentation in London's Selfridges store was a bit cluttered at the time of visit (see fig. 13). Other brands like SONOS and Bose showcase every product separately and thereby exhale a more accessible feeling (for more about its competitors see p. 20). It is a good move to discontinue the A5, A7 and Z2 and replace it with products that create more coherency.

Once again, a good case study is the Zeppelin of which 91% of all models have been sold through physical stores (B&W, 2014). Department stores like John Lewis (UK) and Walmart (USA) account for 23% of total, core channels account for 49% and Apple stores sold 19% of total.

Zeppelin (2009-2016)

(B&W, 2016b)

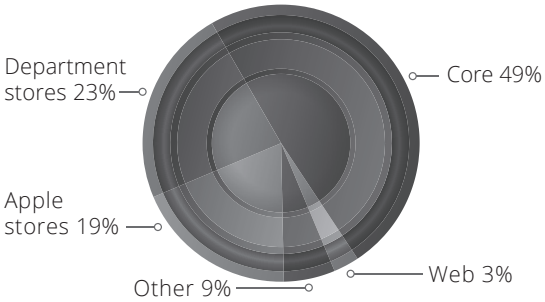


figure 12 Channels for the Zeppelin

The intelligent speaker is intended to sell through premium department stores worldwide. But there is a ceiling to the amount of money that people bring into a department store. With a price point between $1500 and $2000 USD this might be tricky.

BRAND AWARENESS

B&W has a small presence outside the Hi-Fi industry. From the estimated 12.1 million people in USA and UK that are considered part of the premium audio market, only 22 per cent knows B&W (2CV, 2015a).

In Apple stores B&W was privileged to demonstrate its products to millions of people that are looking for supreme build quality. Multiple B&W speakers were on sale in Apple stores until 2015. With the discontinuation B&W lost an important channel with a high volume.

On the other hand, the increasingly large presence in the automotive industry creates exposure with an accurate focus on potential users of one of the core products. If people buy into a premium audio system for their McLaren, Maserati, Volvo or BMW they might as well be interested in buying a premium audio system for domestic use.

Method

London shop visit February 9, 2016

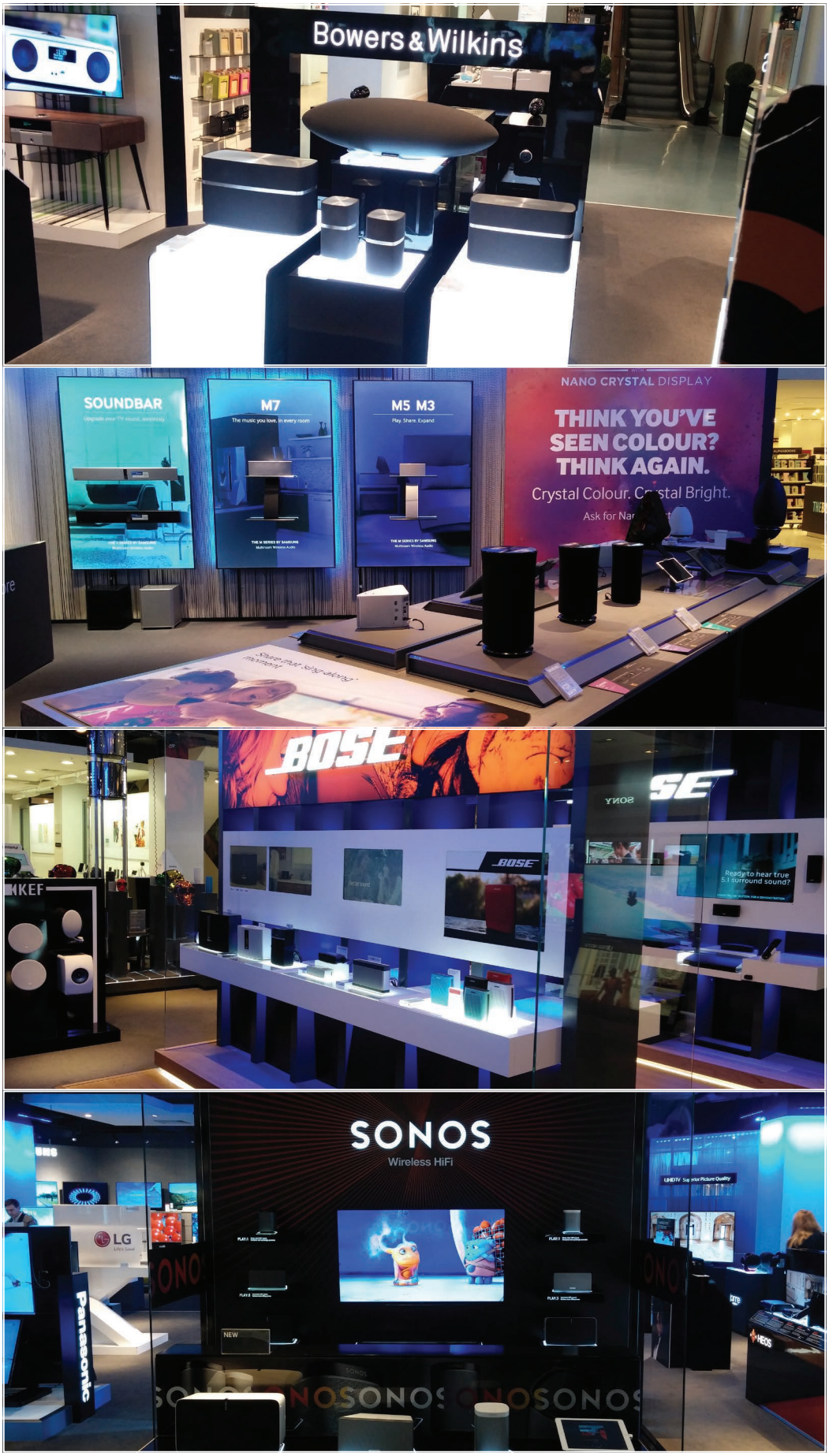


figure 13 Point of sale in Selfridges, London. From top to bottom: Bowers & Wilkins, Samsung, Bose, Sonos

Competitors

The 'wireless' speaker market is dominated by the Santa Barbara based Sonos that beats the drums for user experience. Discrete speakers with an inclusive design is what the company is shipping in high volume. With a small comprehensive portfolio Sonos achieved a yearly turnover of 926 million euros (Hijink, 2017) - roughly five times the turnover of B&W group.

On the other side of the spectrum is the Danish Bang&Olufsen that is renowned for high-end audio and video solutions with a minimalistic design. B&O speakers are a piece of art and the iconic designs makes the brand well-known.

Other names in the industry are Heos, Dynaudio, Bluesound and Naim. Big multinationals like Sony, LG and Samsung noticed the success of Sonos and are now taking on the lower segment.

B&W is compared with two iconic players from this market; Sonos and Bang&Olufsen, using a competence wheel (see fig. 14). This competence wheel remains rather subjective but does give an insight on which competences to emphasise and which to improve on. The three potential competences to improve on are: user experiences, scalability of the brand's system and brand awareness. Acoustic performance and build quality are two of the stronger company competences and these are identified as key elements of B&W's brand equity.

Method
Competence wheel

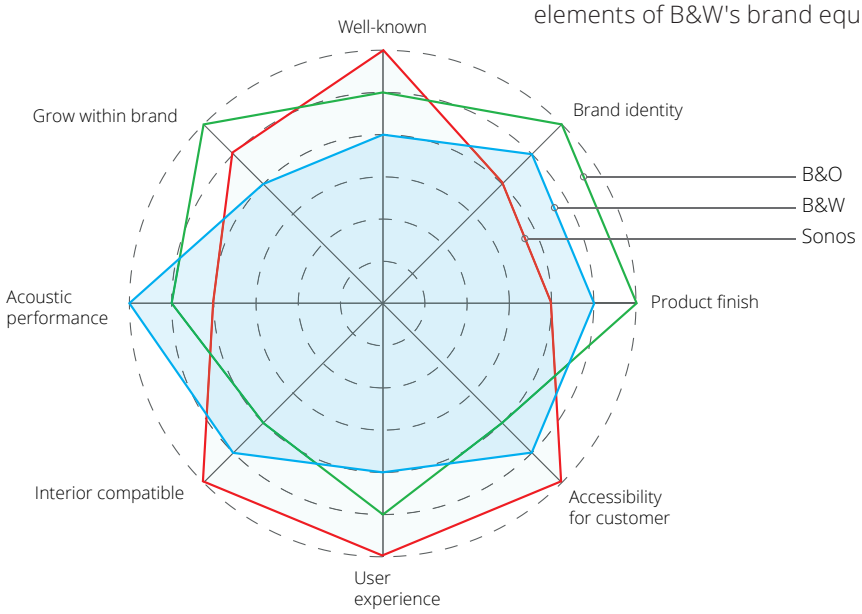


figure 14 Competence wheel to compare B&W to its direct competitors

figure 15 'Wireless' speakers from (top, left to right) Sonos, Heos, (bottom, left to right) Bang&Olufsen, Dynaudio and Naim

Bowers & Wilkins' brand equity

"In moving forward, we have to look backwards and recognise the heritage, the values and what we have done before." - (Matt Hill, 2017, 9:01. Interior Design Director for Aston Martin DB11)

Compared to its competitors Bowers & Wilkins has a robust set of assets that should be articulated more thoroughly throughout all communication. This is B&W and this is why B&W is different from its competitors.

Help every user to experience the music better

When listening to B&W speakers there is a strong sense of emotions involved. The high-quality representation of sound is for many people an excitement that makes them feel as if they experience the artist performing live.

Research – incremental steps

Every time a new product is launched, a new standard is set. B&W defines the state-of-the art by introducing new technologies like Nautilus tube, decoupled drivers, diamond tweeters and the dimpled port tube. The Steyning Research Establishment is often regarded to as the 'University of Sound'.

True sound – studio quality

B&W's speakers are rewarded as the top performing speakers with several collaborations with recording studios as a result. When buying into a B&W speaker the buyer knows he or she has got the best performance in that category.

Heritage – credibility

The company has proven to be able to provide a constant high level of quality for the last 50 years. The new flagship of the 800 series is being ordered months before its launch, showing the trust people have in the company.

Craftsmanship

The British company still makes their high-end speakers by hand in their factory in Worthing, United Kingdom. Five days a week 300 people work all around the clock to craft a speaker with supreme build quality. All made from carefully selected wood, veneers and premium materials like the diamond tweeters.

Refined, elegant and classy

More subjective attributes on a product level are that B&W's products exhale a certain character.

figure 16 The intelligent speaker will be the successor of the Zeppelin Wireless

Vision

In-store observations reveal that buyers of lifestyle products want 'instant everything'. B&W's most important discriminator 'sound quality' will only affect the consumers' experience on the long term. For in-store persuasion B&W's lifestyle products should be equipped with an iconic next generation user-product interaction.

Content

Market developments

The industry is seeing a rapid shift from conventional wired loudspeakers and docking stations to wirelessly controllable speakers. Big multinationals create a higher accessibility to speakers by delivering extremely user-friendly plug and play solutions.

With less resources compared to the multinationals mentioned in the section *Competitors* (page 20), B&W will have difficulties to keep up with the latest connectivity and user experience. Therefore, it is advisable for B&W to not just copy the state of art, but offer a brand fashioned implementation of the ongoing developments.

Whereas Sonos has a vision to make every audio product wirelessly controllable, B&W comes from the point of view that they can make any existing audio solution sound better. With the rise of the Internet of Things (IoT) wireless technology is becoming less of a discriminating factor and more of a hygiene factor (something that users expect). It is now time for B&W to make a statement for high-end sound in this market.

Customer needs

The best places to visit to assess the current situation are dealers and stores. The most trivial feedback from conversations with B&W's dealers and in-store observations are listed.

Method
Hi-Fi dealer and store visits in London.

According to premium Hi-Fi dealer Saj Afzal “The first Zeppelin hit the spot with customers, was a new price in a new category, had an iconic design to supplement and had connectivity relevant at the time” (Personal conversation Saj Afzal Audio Venue London, February 9, 2016). Now this category is much more saturated. In his opinion the Zeppelin Wireless is too conservative with less connectivity and relevance today compared to items like SONOS and the Naim Mu-So (see fig. 15).

Premium custom installer Paul Bennet working for Robert Tausig in London experienced that a customer with a high budget invests less time in the purchase. Brand credibility and trust play an important role in his or her decision taking. Consequently, it is hard to build a relationship with this kind of customer in the store.

Method
Observation premium department store John Lewis London, February 9, 2016.

The intelligent speaker will be sold through premium department stores. Customers of these stores want everything instantly. They are looking for a turn-the-key solution; any hassle for product set-up and eventual use will obstruct them from their purchase. Also in a noisy department store the first thing people do is crank up the volume. If it plays loud it will impress. It is not before it is in their living rooms that the product has to tick the other boxes (Personal observation, John Lewis London, February 9, 2016). Therefore, sound quality will make you appreciate the product and helps reconsideration of the brand for next purchase, but it will not sell in the first place.

The user's experience will always be a mix of expectations and the actual performance. The expectations are amongst others related to size. A big sound from a small box is more impressive than a big sound from a big box. Whereas the company focuses on the physics of creating the best acoustic performance it should also strive to keep the irrelevant physical volume ‘zero’.

Next to size, the shape of the product also plays a trivial role in the product’s use case. Paul Bennet argued the Zeppelin, due to its shape, is a centrepiece that requires some breathing space around it (Personal conversation Paul Bennet, Robert Taussig London, February 9, 2016). Not every household can accommodate a product with a physical footprint like that.

Relevant market trends

The presentation of the concept speaker is intended for Q1 2019. In order to develop a product with a ten-year lifetime, one should ask: what will the world look like a decade from now?

> "No one can predict the future exactly, but we know two things: it's going to be different, and it must be rooted in today's world."
> (Thiel, 2014, p.6).

A lot will change in ten years. In perspective, a decade ago smart phones did not exist. Apple's iPods were the next big thing, B&W did not have its Zeppelin docking station. The first iPhone was released in 2007.

That said, at this point in the design process the intelligent speaker's specs should be a little out of B&W's comfort-zone to make it a statement in technology. Developments must not be restricted to incremental changes.

In any case, the future "must be rooted in today's world" (Thiel, 2014, p.6). Today's trends in the field of society, markets and technology are mapped on a midi level (market level with a time horizon of one to five years). The full market trend analysis can be found in appendix A. A selection of trends is highlighted below.

Method
Trend analysis

Intuitive interaction
Over the next years, user-product interaction will gain importance in New Product Development. Smartphone apps are counting up quickly and are therefore becoming a hassle. People get used to a new standard easily. The few seconds it takes to unlock your phone and fire up an app are becoming a hassle. In the far future (beyond the scope of this project) products will get a greater intelligence and will serve your needs without taking the step to tell them.

High-res audio
With the market of streaming services becoming saturated, offering high-res streaming is now a distinguishing factor but will become standard in the future.

Machine learning
Products will be sold with a basic know-how and will learn from your input. Thereby better understand your habits and preferences and create a more intimate experience since the product is tailored to the user.

IoT platforms
There will be a greater cooperative interaction between the deluge of IoT devices. Platforms will unite IoT devices with cross-platform software. Making it equal to control your music system from phone, tablet, laptop or fridge.

Circular Economy
Companies in all sorts of markets will move from selling products to offering services. Hereby creating a more competitive environment with a more eco-efficient product use as result. In practice this could mean not selling a speaker but offering the service of high-end sound.

Virtual Reality
Today multiple major players are investing in VR. The next years the first applications of VR will be implemented to for instance assist doctors in patients' therapies.

The dream

Method
Brainstorm

Allow us to think limitless for a moment; what would be the dream? In a short brainstorm session B&W employees shared their take on what the future will bring. The results are summarised in table 2.

Whereas 'wireless' speakers nowadays are stand-alone units or part of a brand's scalable system in the near future they will start to communicate with other IoT devices in the house. The music system will inform the user on the status of for instance the ready-meal in the microwave and you can talk to it. Ultimately the music system will become the main interface of your IoT world wherever you go.

Furthermore, where services are now building databases of user profiles to understand the user's preferences, in the future music systems will be able to understand its audience on the spot. The loudspeaker or the mesh of IoT products in your house together will adjust to whoever is in the house at that moment, opposed to anticipating on a single user via an app. In practice that means the system will be capable of sensing your mood and activity and adjusting the content accordingly.

As mentioned in the section *Opportunities* (page 24), the user-product interaction will change as well. Apps will be replaced by natural gesture- and voice commands. Over time these sensors will become more accurate and will for instance know what you are looking at by analysing your iris (IB Times, 2016). In the ultimate dream the system will be able to read our minds.

	Now	Up next	The dream
Function	Is a stand-alone unit or part of a brand's scalable system.	Communicates with other IoT devices in the house.	Is the main interface of your IoT world wherever you go.
Source	Can play audio if you connect to it.	Is agnostic to source and will always be connected without pairing.	Will continue stream cross-platform for e.g. television, headphone and car.
Content	Building a single user profile with an online music collection.	Combines content from multiple user profiles according to the people in the room.	Understands preferences of every person in the room by sensing it on the spot and does not require user profiles.
Sound	Can be calibrated at initial setup to compensate for the room characteristics.	Is aware of its environment and compensates actively for changes in the room.	Gives a nearly equal experience in any place in any room by steering audio to the right places.
Control	Needs to be instructed through an app.	Is controllable through natural voice- and gesture commands.	Reads your mind.

table 2 The results of a brainstorm on future developments

Method
Golden circle Simon Sinek. Reason from the inside-out: why - how - what. (Sinek, 2009)

"The goal is to do business with people who believe what you believe." (Sinek, 2009, 5:43).

Vision

Vision (Why?)

From the establishment of Bowers & Wilkins in 1966 John Bowers' vision was to make existing solutions sound better. Up to this day the company is still carrying out its business according to this vision. In every market the company stepped in, it challenges the status quo. For this project the interpretation of that vision would be:

*"B&W will enable **more people** to enjoy a **studio quality audio experience**."*

Mission (How?)

The vision consists of two major elements: 'more people' and 'studio quality audio experience'. The latter is attained through extensive research and development and as a result B&W delivers an unsurpassed true sound experience.

The somewhat fuzzy element 'more people' is more of a concern since B&W is currently addressing a niche market of discerning listeners. However, a large share of B&W's brand equity (see p. 21) consists of assets that do not solely appeal to discerning listeners, but more generally to anybody that appreciates quality.

B&W should lower the barrier for the target group to identify with a strong B&W image imbued with the emotion of music and pure, uncompromised sound. What B&W is lacking most at the moment is a state-of-the-art user-product interaction and thus the mission statement reads:

*"B&W wants **everybody who likes quality** to feel part of the brand by offering a **next generation product interaction**."*

Strategy (What?)

The intelligent speaker is the foremost opportunity to demonstrate a next generation interaction that introduces a new buyer to the new range of lifestyle speakers. The speaker will be controlled so intuitively it will appeal to both the early adopter as well as to the less tech-savvy, because it makes it so easy to use.

Consumer electronics shifted from buttons on the device – to remotes – to apps.
With our current demand even apps are becoming a hassle. Eventually products will get intelligent and will think for you. They will learn from your input and preferences and perform our every-day tasks without the need to tell them. The remaining communication will happen without the need of physical contact. The product should feature an interaction that is as close as possible to human interaction and therefore the strategy is:

"To develop a statement concept product with no buttons, no remote and no app."

More specifically B&W will develop a concept product that:

- Is a statement in technology.
- Impresses within the minute. By having an innovative interaction next to great sound quality and a remarkable industrial design.
- Addresses a large audience. Appeal to a new user and less male oriented.
- Provides instant everything. A one box solution which requires low involvement and offers time to music = 'zero'.
- Ensures a 10-year lifetime. The hardware will support future functionalities and updates.
- Is part of a scalable system. The system can be expanded with other B&W products so one can make a custom configuration.
- Has premium build quality. Made of premium materials with a refined finish.
- Enables skill transfer. Technologies and data could also be implemented in other B&W products in the future.
- Serves a domestic use case.

Product definition

figure 17 Assembly of a continuum driver in Worthing, England

Product definition

This thesis concludes in a 'concept speaker' that will serve as a conversation starter. It should explore possibilities and inspire. The concept speaker will embody the intended B&W interaction of the future as well as the acoustic aspiration of giving a consistent experience in any place in any room. Equipped with a circular array of drivers it will be able to change its acoustic directivity and beam width.

Content

Concept speaker

The assignment of this thesis is to "Design a unique B&W user experience for the intelligent speaker". During the course of the project the term 'intelligent speaker' meandered along with the developments. There were no hard restrictions to the definition. Therefore, this chapter is dedicated to restrict the acoustics of the speaker and create a solid starting point to design an interaction (see next chapter) for.

The intelligent speaker will remain a concept speaker with the main purpose to be a conversation starter and to explore technological possibilities. The ultimate goal is to see a range-wide improvement of the user-product interaction.

The concept shows what B&W's vision is for its future developments. It demonstrates their technological capabilities. And last but not least, it brings coherency to its ongoing operations.

Acoustics

Raison d'être

When listening to a live performance the sound of the band is directed straight towards you. Perhaps some band members are positioned a little more off-axis than others. However, the acoustic colour of the venue is determined by reverberations and distortions that bounce of the wall.

The intelligent speaker will unlike any other speaker not have its drivers in a single plane or direction. It will feature a set of relatively smaller drivers all-around. A circular array of drivers can jointly project the sound on the right places in the room, to create a more realistic image of the recording. It can beam the sound of the band towards the listener and project the reverberations of the venue across the room.

Acoustic principles explained

Drivers are fundamentally not directive. In practice, this means that an ordinary speaker with drivers on a single side will emit sound in all directions. The sound that bounces of the wall creates a time-delayed and distorted version of the original sound.

By adding drivers on the rear of the speaker in a 180-degrees phase shift the sound on the rear can be cancelled out. The overall sound pressure level (SPL) decreases, but the sound does become slightly more directive. This is called a cardioid response. Sound pressure levels can be measured for every angle (see fig. 18b). Note that high frequencies (HF) are more directive than low frequencies (LF). This is because low frequencies have longer wavelengths.

To get a more directive response it is possible to let a 360-degree array of drivers collaboratively emit sound. Here every driver radiates sound as a function of the position relative to the desired direction of sound. The beam of sound gets narrower (see fig. 18c, cyan). Again, the mid frequencies are effected the most by this principle. In earlier revious research Moller et al. (2010) achieved desired directivity patterns from 500 – 4000 Hz.

A circular array allows to steer this narrow beam in any direction by changing each driver's input signal. Furthermore, it can simultaneously emit other polar responses, like for instance a di-pole (see fig. 18c, green). From any stereo recording the spatial information can be extracted. Using a di-pole, this spatial information can then be emitted to the sides. When desired, an omnidirectional pattern is also possible.

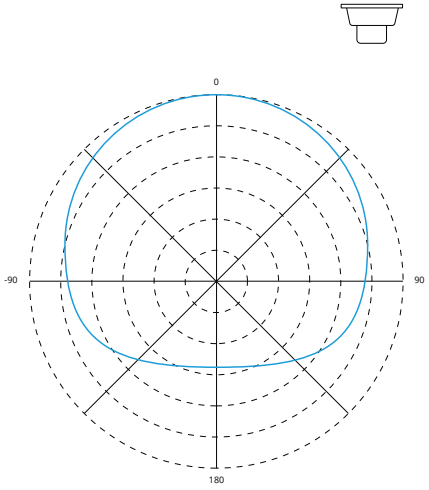


figure 18a Simplified Sound Pressure Level vs Angle of normal speaker.

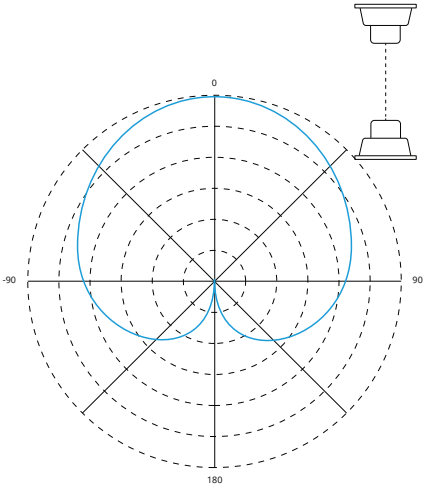


figure 18b Simplified Sound Pressure Level vs Angle of cardioid pair.

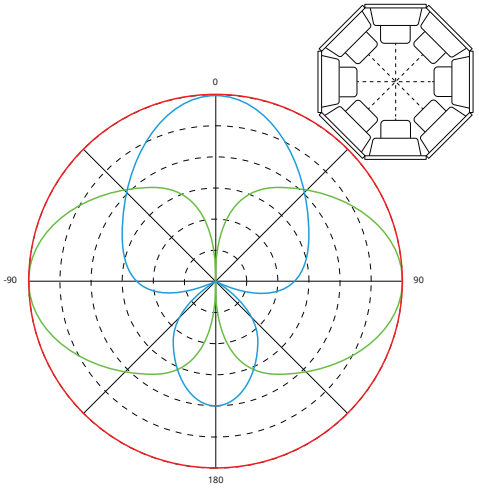


figure 18c Simplified Sound Pressure Level vs Angle of 360-degree array. Cyan: Hypercardioid, green: di-pole, red: omnidirectional.

Intelligence

Now, every speaker system has an optimal listening position. The user of this lifestyle speaker does most likely not stay in the ideal listening position (or sweet spot) for most of the time. The sound experience decreases when listening outside the sweet spot and so the speaker has to adjust.

The speaker must know the position of the listener(s) in the room. This can be achieved using different types of sensors and will be discussed in the next chapters. B&W's concept engineer Jon Moore has developed a principle that would be able to independently steer each driver relative to the desired direction of sound. The speaker can change its directivity so it puts the listener always in the optimal position.

Applications
All together the speaker will assist the user to get the most consistent experience he or she can get. These are some of the benefits:

Ideal listening position
It will track the location of the user(s) and adjust the directivity accordingly. Think of stereo camera tracking to distinguish a cat or plant from the actual user. This function also enables proximity sensing, meaning that the volume could drop a fraction when the user is close.

Adjust to number of listeners
By changing the Digital Signal Processing (DSP) to each driver the speaker can control the envelopment. If the users are far apart the envelopment will become wider whereas a single user would result in intimate sound. Potentially the intimacy or involvement can also be changed on demand to create a background sound.

Room proofing
It will dynamically compensate for changes in room character such as open patio doors, closed curtains or having friends over to create a consistent experience regardless of the environment. The way of measuring this is not defined in the project description.

Positioning proofing
Speakers usually require some breathing space. If a normal speaker is placed too close to the wall low frequencies create a boom from bouncing of the wall. With a 360-degree array of drivers the radiation to the wall can be cancelled out.

Configuration

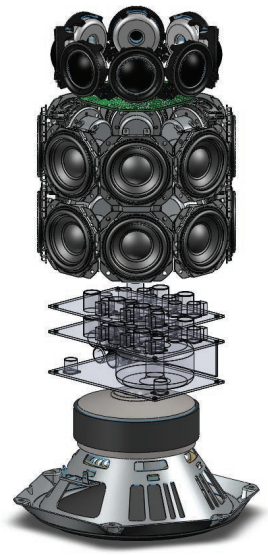
There are some rules of thumb that apply to a circular array like this:

- Large drivers cannot move fast enough to produce HF sound due to inertia.
- A single small driver can produce MF and HF but does not have the required surface area to move enough air to create LF. However, by letting multiple of these smaller drivers emit the same signal the surface area adds up and can thus create LF.
- Drivers need to be as close as possible to create directivity. So, with the physical dimensions of the drivers, the outline of the product is also configured.
- The drivers need to be perfectly cylindrical to keep the sound constant when you walk around the product.
- A higher number of drivers creates a more fluent transition when walking around.

That said, from an acoustic point of view the product would be a cylinder that is covered in tiny drivers (see fig. 19, top left). However, the costs of a configuration will quickly ramp up so the chosen configuration will always be a compromise. Also, directional bass is extremely difficult and is therefore out of the scope for this project.

figure 19 Different configurations with an omnidirectional coverage

project B&W's concept engineer Jon Moore proposed a preliminary configuration that would be the best basis for further testing (see fig. 20).



8x HF (M1: 27mm)

16x MF (T7: 50mm)

18 channel amp
450W power supply

1x LF (Zeppelin Wireless: 150mm)

figure 20 Preliminary configuration for the intelligent concept speaker

Existing drivers are used from the T7, M1 and Zeppelin Wireless. The chosen configuration is two times eight 50mm full range units (T7), eight 27mm tweeters (M1) and a 150mm omnidirectional bass unit (Zeppelin Wireless). The tweeters are powered by eight 25W amplifier channels, another eight for the mids and a powerful 50W channel for the bass.

The Zeppelin Wireless 150 bass unit would ideally have an enclosure of 5 to 8 litres. In the Zeppelin it is even squeezed in a 4.5 litres box. The 4.5 litres was taken as the minimal required volume. The other drivers require such a small volume that it should not be limiting at this stage.

For a full list of requirements see Appendix B.

User-product interaction

figure 21 The physical user-product interaction on the existing B&W Zeppelin Wireless

User-product interaction

In order to materialise the crafted vision an agile approach has been adopted. Repetitive user tests quickly gave insights into the needs and values of the intended consumer. Gesture control is the way forward for simple navigation through a music library or playlist. Voice control is the essential addition to assist the music lover with her music preferences.

Content

The speaker is defined as a speaker with a directive array of drivers that senses the position of the listener(s) and adapts accordingly. Moreover, it will be a "statement concept product with no buttons, no remote and no app". The user-product interaction will be unlike any existing solution. What interaction suits this use case without alienating its audience?

Voice enabled speakers are gaining the consumers' interest. Amazon Echo is forecasted to dominate over 71% of the market in 2017 followed by Google Home with 24% ("Alexa Say What", 2017). Both Amazon and Google focus on voice control of the smart home using their own digital assistant software, respectively Amazon's Alexa and Google Now.

Apple on the other hand, announced their own voice-enabled speaker called Apple Homepod that will be available December 2017 ("Homepod", 2017). The Apple Homepod

figure 22 Voice-enabled speakers; Google Home, Apple Homepod and Amazon Echo

is a speaker first. The Siri voice control assists listeners to choose a tune from their Apple Music collection. According to independent research, Siri is not nearly as accurate as third-party platforms like Amazon's Alexa, Google Now and Microsoft's Cortana (Enge, 2017).

Sonos speakers do not feature voice control as of this writing but the company is giving hints that they might in the near future. Their speakers are already equipped with microphones. Chief of Design Mieko Kusano has been appointed as their new Chief Voice. Now, Sonos decided to pair up with Amazon Echo (Hijink, 2017). This collaboration could potentially combine the best of two worlds.

Voice control supports a wide range of commands to navigate through music collections, playlists or to jump from tune to tune. However, when the listener is in proximity, voice control might not be the most convenient way to perform simple commands.

Gesture control could ease the simple controls and could supplement voice control in its shortcomings. Think of commands like *Next*, *Pause*, *Add to list* and possibly *"Don't you ever play this song for me again"*.

The combination of voice- and gesture control could hypothetically replace the combination of apps, remotes and physical buttons. But does it really provide a better experience in practice? The hypothesis is tested in a pilot.

Exploring ease of use

Method
Pilot

A pilot was conducted to validate the idea of voice- and gesture control replacing apps and buttons. The hypothesis for this pilot is:

Hypothesis: A combination of gesture- and voice control can provide enough freedom for a discerning listener to navigate through his or her music preferences.

In order to confirm or knock down the above hypothesis a pilot should answer the following four questions.

Q1: What level of involvement does a discerning listener typically have using a lifestyle speaker?

A lifestyle speaker fundamentally serves a less involved interaction compare to the performance range. The user is assumed to mostly listen to playlists as opposed to songs.

Q2: What is according to a discerning listener the most pleasurable way to start a music stream? (app, discovery mode, voice command, other..)

There is a difference in involvement between choosing the right starting point and navigating to the right songs within a playlist. The higher the listener's involvement, the higher is the need for freedom.

Q3: When listening to a bespoke playlist or radio can a discerning listener give enough guidance to the music choice with some simple gesture- and voice commands?

Q4: Are the proposed gestures intuitive and understandable?

Method

Four respondents participated in a 60-minute pilot in which they performed numerous tasks with a music system that features make-belief gesture- and voice control (see fig. 23). The respondents' input was interpreted by the host who would then take the associated actions manually. For more information on the pilot's method, prototype and results, please refer to Appendix C. The pilot's main purpose was to quickly assess the viability of the application of gesture- and voice control for speakers before heavily investing time and effort in this direction.

Results

The outcomes of just four tests was found to be enough incitement to invest in making a stand-alone functioning prototype.

Q1: What level of involvement does a discerning listener typically have using a lifestyle speaker?

First and foremost, this is different for every user. The typical use case of a lifestyle speaker sees a clear difference in involvement in different stages of use. A user of a lifestyle speaker listens mostly to playlists as opposed to single songs. In order to enjoy listening to these playlists the listener first has to invest time and effort to explore and/ or collect songs to build the playlist. In a scenario with high involvement the listener is carefully looking for an album, artist or playlist to meet his or her desire. In between phases of high involvement there is the actual listening phase that only involves basic controls like skipping a song, change volume and saving/ forgetting a song (see fig. 24).

Q2: What is according to a discerning listener the most pleasurable way to start a music stream? (app, discovery mode, voice command, other..)

All respondents acknowledged that they would feel comfortable if the music system starts playing their saved songs in 'shuffle' mode after a simple voice command or gesture. In this case there is no need for an app or remote. One of the respondents explained: "With the time it takes to unlock the phone, open the app, wait for the app to load content: gestures would be much quicker". For all respondents this is different from the way they would start a music stream at the moment.

figure 23 Prototype to facilitate pilot on the added value of intuitive control

Q3: When listening to a bespoke playlist or radio can a discerning listener give enough guidance to the music choice with some simple gesture- and voice commands?

The respondents felt comfortable to navigate through the music using gesture- and voice controls. However, the discerning listener that B&W addresses with its products is interested in high-content information about an artist or album. This information is hard to communicate without visual feedback.

Q4: Are the proposed gestures intuitive and understandable?

There is room for improvement for some of the gestures. Since all respondents were English it does not prove to the world-wide acceptance and understanding of the gestures.

Conclusions

Does all of the above provide enough reason to believe the hypothesis is true? Once again, the hypothesis was: *A combination of gesture- and voice control can provide enough freedom for a discerning listener to navigate through his or her music preferences.*

Long story short; a combination of voice- and gesture control would not suffice for the discerning listener when executing complex tasks that require high involvement.

A combination of gesture- and voice control can provide a breakthrough for the ease of use by shortening the time for basic navigation (e.g. next track, adjust volume, pause/play or save a track to your favourites). Gesture control is the quickest way to perform simple commands. Voice control is the additive when it comes to more radical changes in the listener's preferences. However, in situations with extremely high involvement like for instance set-up, the combination of gesture- and voice control does not suffice.

However, the system always has to be connected to the listener's database in order to get content. The database, like for instance a Spotify account, can perfectly work in parallel to the system. With a new or empty Spotify account the system will basically recommend any track, making it hard to find the right starting point. From the section *Customer needs* (page 23) we know users desire instant everything. With this learning curve users need to commit an undesirable level of investment. It does not sell in a minute. Managing the first tunes of your music collection might be more convenient to do with use of a screen, e.g. an app. When the database takes shape, the proposed content gets more accurate. The eventual use of an app will be incidental.

A speaker system cannot completely eliminate the use of an app in the beginning. However, when the listener's database is rich enough the speaker can play a bespoke playlist, making the use of an app (almost) redundant over time.

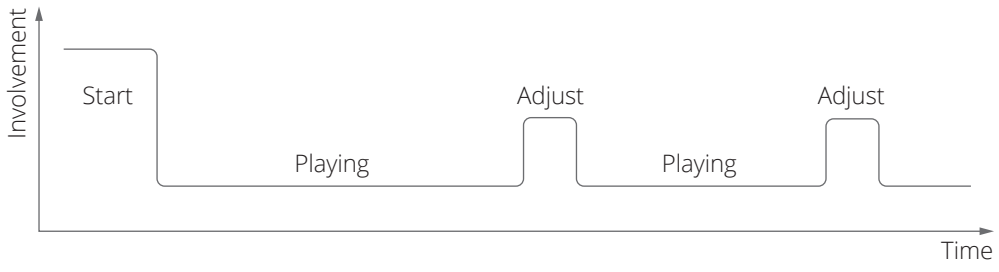


figure 24 Involvement of a lifestyle speaker

Feasibility study gesture control

Method
Pilot

Gesture control can be a success or failure depending on the accuracy of the detection. But how does it really feel when the volume increases along with the movement of your hand? And how does it feel when it does not? This pilot served to explore what it would take to implement gesture control and to take the idea one step closer to a tangible proposal. This pilot should solve the following questions:

Q5: How accurate can gesture control potentially be?
How accident-prone can it potentially be? How big or small can the gesture be? Is it accurate enough to detect engagement (looking at it)?

Q6: What are the limitations of the Field of View (FoV)?
What are the limitations for distance and envelopment? How does that affect the use case?

Q7: How would people use the product?
Where would people position the product in their homes? Do people like the feeling of immediate feedback? What gestures would not interfere with daily used gestures?

Method

For the second pilot a prototype was developed that features gesture recognition. In as little as two weeks software was written that connected an infrared depth sensor with a speaker. The music content was streamed from Spotify. Because of time restriction the prototype did not use an API, but just the Spotify desktop app instead. The software simulated keyboard strokes and was thus limited to the few keyboard shortcuts Spotify knows. Four respondents were asked to play around with the prototype in their very own living rooms. All details about the pilot's method, prototype, supported gestures and results can be found in Appendix D.

Results

Q5: How accurate can gesture control potentially be?
Of course, this prototype (of which the database has been made in as little as two weeks) is not a proof of the final achievable accuracy. However, it was remarkable how much could be achieved in this time.

To say something about how accident-prone it can it potentially be, the two types of errors should be explained first, i.e. a false positive or a false negative. A false positive is when the wrong gesture is detected and a false negative is when the gesture is not detected at all. According to Microsoft's documentation the error rate for false positives could be reduced to approximately 5% and false negatives even to 2% (Microsoft, 2014) meaning an accuracy of over 95% would be achievable. Furthermore, the Software Development Kit (SDK) comes with a tool for facial recognition which is remarkably accurate. When in a fixed position in front of the sensor it can even understand facial expressions and measure heart rate by looking at your skin.

At a first attempt pointing somewhere resulted in a false negative error. Moreover, the prototype only listened to the one person that was featured in the clips. After updating the database the gestures were based on a double movement; first reaching out, pausing for a second and then make the movement. This made it easier for the software to distinguish a gesture form any other movement. Also, the database was made using clips from two different people, opposed to a single person in the first try. The outcomes were reasonably accurate considering how little time it took to build the databases.

Currently, the database is based on clips from just two different persons in five different positions. For greater accuracy the database

should be constructed with more people in different environments. Microsoft (2014) explains "It is important to record people in different environments, and with different heights and tilt angles of the Kinect Sensor. [...] Wearing different types of clothing can produce different skeletal data, even for the same person." With a little refinement the gestures can be as small as a hand's length.

The accuracy can however be affected negatively in a few cases: reflective surfaces such as photo frames, mirrors or windows, exposure to direct sunlight and incandescent lights shining into the camera. The Kinect 2.0 is a time-of-flight camera which means it emits an IR laser pattern throughout the room which is reflected on an object's or person's surface and then sensed by the IR camera. Any of the above will interfere with the emitted IR pattern in that area making it unable to read the gestures.

Q6: What are the limitations of the Field of View (FoV)?
The near-range recognition was low. Because of the narrow field of view, you find yourself relatively quickly outside of the sight of the camera. The gestures worked op to a distance of an impressive five meters. In one case the living room was much bigger than 5 meters making it hard to understand when you are in range of the camera.

The horizontal envelopment was very narrow: 70.6 degrees. If you are standing on the edge of the camera's view the system has difficulties to recognise your skeleton and can therefore not cope with the gestures. The vertical envelopment is no problem whatsoever.

It does affect the use case because you cannot sit close or aside of the speaker. For further development it is required to look for a 180-degrees FoV.

Q7: How would people use the product?
There is a bit of a learning curve to the gestures since the gestures are so specific. It is hard to take natural gestures as a base since they are mostly used for indicating other things in daily life.

All respondents would place the product 'on the side' or even behind the sitting area. In all cases not in front of the user. Furthermore, the respondents want the gesture to be consistent no matter what position. That means they do not want to bend towards the product in order to make the gesture but they want to be able to do it at all times. This was something that was not supported for this prototype.

figure 25 Prototype for the second pilot on the feasibility of gesture control

Conclusions

The gesture control only works when the user is in sight of the camera. When the user is standing around the corner he or she needs to rely on voice commands.

This prototype was limiting the use because it could not notify whether or not the user was in range. The user would expect the product to work, but it did not respond. This happens when the range of the product does not cover the entire room (see fig. 26). The range of the Kinect 2.0 sensor is very limited. The Kinect 2.0 was chosen for this prototype because the well-developed SDK resulted in speedy developments. The sensor was introduced on November 22, 2013. With the ongoing developments of autonomous driving cars and artificial intelligence it is expected that this kind of sensor will see rapid improvements in the next years. The longitudinal range (depth) will not obstruct the use of gesture control for domestic use in the future.

Furthermore, the field-of-view of a Kinect 2.0 is 70.6 degrees. In order to recognise gestures from the user in any place in the room the field-of-view should be at least 180 degrees. A single camera in the speaker could rotate to follow the user's movement. A fisheye lens could increase the field of view. There are ways to increase the FoV of a single camera by using a 360-degree mirror (Bourke, 2009). Previous work of Maimone and Fuchs (2012) demonstrated ways to increase the FoV by using five Kinect sensors from the same location, looking at different angles. Multiple cameras can be integrated in the speaker. Schönauer and Kaufmann (2013) successfully managed to link up multiple Kinect sensors that are positioned in different rooms throughout the house. The latter is not an option as part of the system because it would jeopardise the 'turn-the-key' solution. Potentially there could be a close integration with existing IoT solutions to increase range to different rooms. There even are experimental ways to determine gestures from fluctuations in the in-house Wi-Fi signal if the user is standing between the router and the speaker. There is no reason to assume why there would be no solution for the lateral field-of-view.

The consumer hardware that is currently available does not provide the required range. However, anticipating on future developments it is assumed there will be a bespoke solution in the future that meets the requirements. For further developments the depth camera is assumed to have a 360-degree vision and is visualised as a ring of high gloss plastic that holds the cameras.

Method

Consulting expert Julian Kooij, MSc. Artificial Intelligence, Assistant Professor at the Intelligent Vehicles and Cognitive Robotics section of the TU Delft.

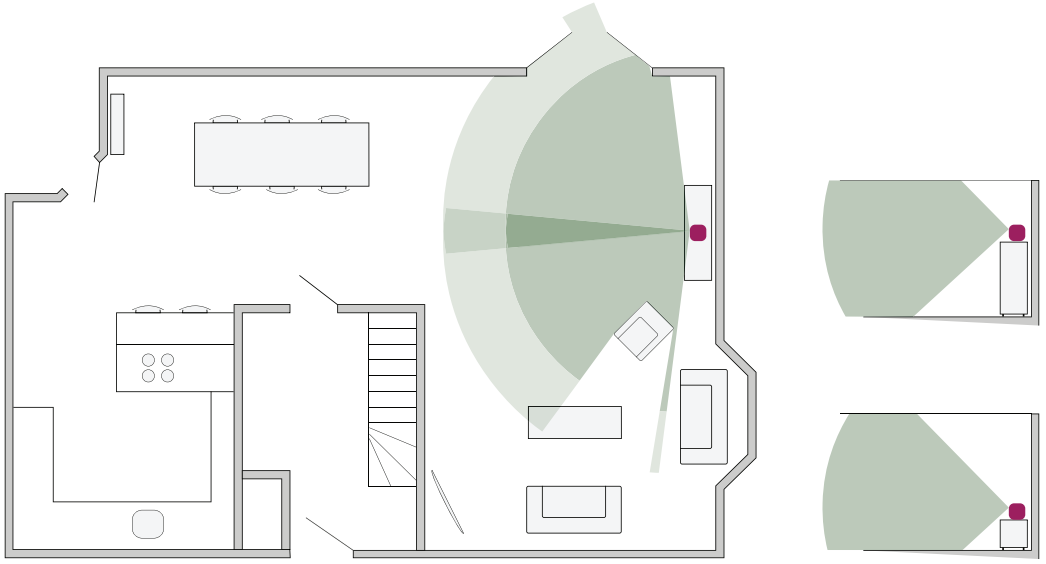


figure 26 Positioning in the room

User applications for 2020

The intelligent speaker is a statement in technology and user experience. The circular array of drivers creates a new level of freedom. Where traditionally the manufacturer's influence on sound experience reached as far as the front door, the intelligent speaker automatically compensates for imperfections in the room acoustics and moreover the position of the listeners. The user experience is much closer to the aspired performance.

The possibilities seem to be endless when designing a speaker with these specs. What user applications are to be introduced at the concept speaker's presentation in 2020? A stepwise approach to the intended user applications is illustrated in figure 27. The intelligent concept speaker will feature the essentials of voice- and gesture control. The gesture commands only work in case of eye contact with the device to prevent false positives. After a few years' time a software update will enable biometric authentication and recognition of moods and habits. The hardware installed in the product facilitates these functionalities.

The speaker's advantages over any other speaker are split in two; those due to the array of drivers/ those due to the future UI of lifestyle speakers. The user-product interaction as tailored to the intelligent speaker might also be applied to any other lifestyle speaker B&W will make in the future. The interaction does not fit the performance range products as they require a more involved interaction.

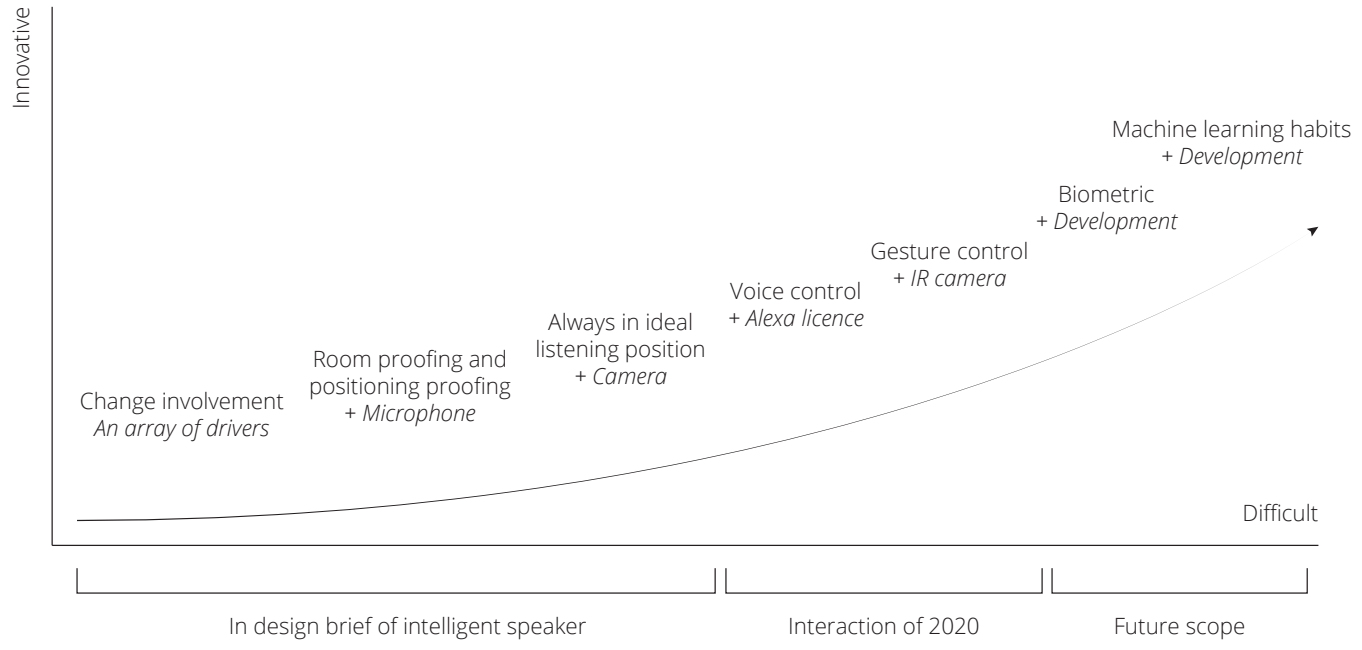


figure 27 Minimal Viable Product applications and future scope

Interaction of 2020

Method
Interaction statement; a step in the Vision in Product Design method from Lloyd, Hekkert & Van Dijk.

Interaction statement

What emotion should the user-product interaction express? To make this vision tangible it is formulated in a mission statement. This vision is usually a metaphor that has the envisioned interaction qualities.

Morgan Freeman is one of today's most iconic actors and narrators. His distinctive voice brings a rich emotion. Talking to Morgan Freeman reflects the intended user-product interaction in many ways.

QUALITIES

The interaction should express a warm, familiar feeling that makes the user feel at ease. When you are ready, the device is ready. However, the device is not too involved and does not require high-maintenance. Instead it is unforced and laid-back. The device is not a softy, it is not humble. It expresses confidence and shows the user it is operating with the greatest ease. It is imposing while still remaining modest. The device's feedback is thoughtful and positive. Transitions are fluent and smooth. After a hard day of work the use of the speaker must make you relax.

Using the product should feel like "talking to Morgan Freeman".

figure 28 Using the product should feel like "talking to Morgan Freeman"

Managing music content

A new interaction requires a bespoke way to manage music content. With a single voice- or gesture command the device plays your favourite tunes. This implies the use of screens as a mean of primary visual communication is not available. Therefore, the discovery element plays a more important role. New content is pushed to the user who can then decide to save the songs for later.

For every new song the user can decide to:

- ++ Queue up related songs: “track station”
- \+ Save it for later: “save”
- +/- Listen to it once, no command
- \- I don’t want to listen to this (now): “next”

For every saved song the user can decide to:

- ++ Queue up related songs: “track station”
- +/- Listen to it once, no command
- \- I don’t want to listen to this (now): “next”
- -- "Forget" song

Content from any list, radio, podcast or mix can be saved using algorithms similar to Shazam and Soundhound.

The user is able to change the ratio between saved songs and discovery songs by a simple gesture.

The proposed content is accurate with an extensive list of favourites or saved songs. At the initial set-up the user has to go to a webpage to connect the speaker to the Wi-Fi network. Next, the speaker needs to be linked to a database like Spotify or Tidal. If the user profile is not extensive enough, the user is asked to pick his or her preferences. The rest of the library will grow as the user "saves" and "forgets" songs.

The speaker tracks the listener(s) position(s) and directs the visual feedback accordingly. This transforms a speaker from a rather passive product into a more intelligent, aware and personal product. Imagine a speaker that will light up when you look at it and is always ready to go.

Gestures

All gestures are unimanual, meaning the user can operate the device with one hand. With the lack of screens, commonly used gestures do not apply. Turning a virtual switch mid-air does not make sense. Piumsomboon et al. (2013) researched gestures for Augmented Reality using user-centred design. For their research 40 respondents performed 20 different tasks whilst picking their own gestures. The 11 gestures that were used the most are listed in figure 29 and form an inspiration to those of the intelligent speaker.

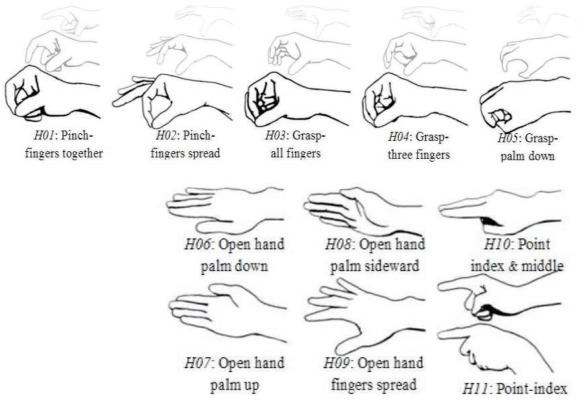


figure 29 Eleven most understandable gestures (Piumsomboon et al., 2013)

The interaction can be explained best as an imaginary head-up display (see fig. 30). The music feed is coming from the right-hand side. The album covers are moving to the left as the music plays. The cursor hovers above the song that is being played with the exact location being how far it is into the song.

The cursor is the part that can be controlled by gestures. To start or pause the feed, the cursor simply needs to be tapped upon.

Swipe sideways with open hand palm to push cursor to the start of the desired track. A swipe to the right will make the feed go to the next track. Note that in an app one would have to swipe the album cover to the left or press the button on the right-hand side.

A pinch of index finger and thump with the other fingers together (fig. 29, H01) would grasp the cursor for more subtle movements. Hold and slide sideways to skip forwards or backwards in steps of 10 seconds. Let go when the cursor is in the desired position.

The next most important thing is the ability to save or forget any played song with a single gesture. For saving a song the semantic meaning of “embracing” something is used. The colour changes from a cold to a warm colour tone. To forget a song the album cover has to be moved away, using the semantic meaning for "letting go". The colour changes from a warm to a cold colour tone.

All states are explained in more detail in table 3 and all gestures are explained in table 4.

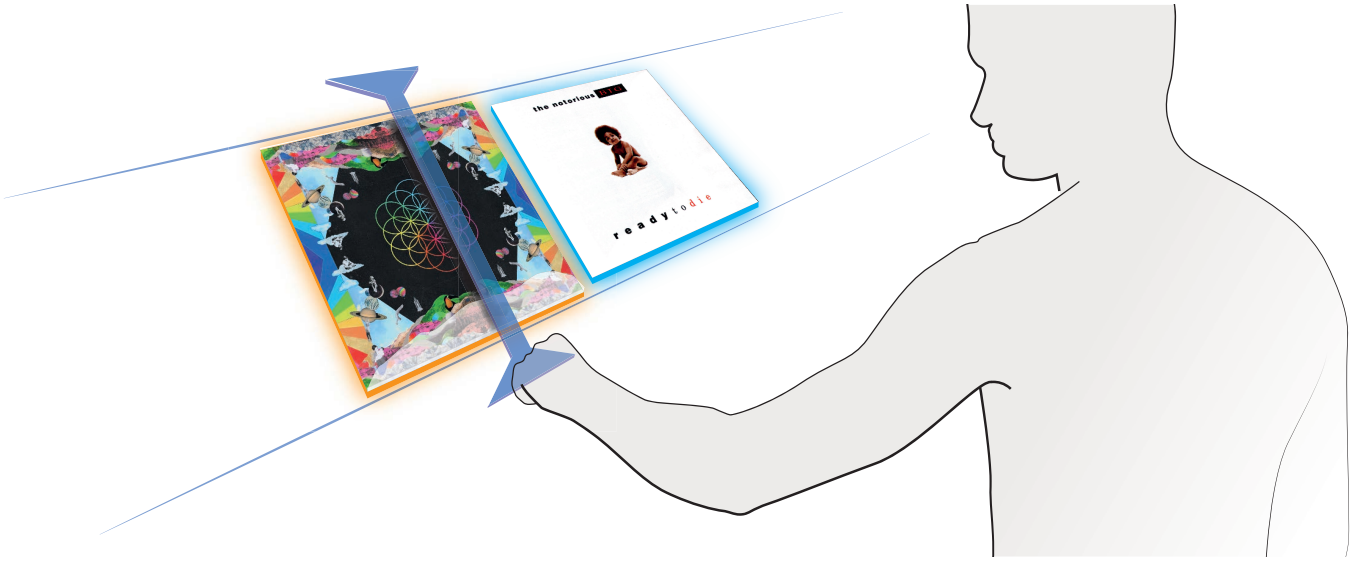


figure 30 Visual representation of the mid-air interaction using gesture commands

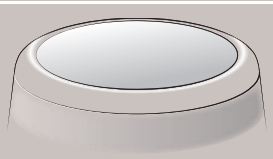
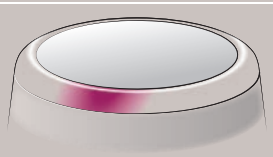
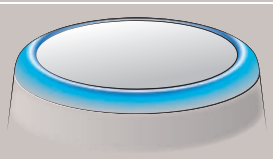
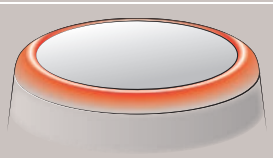
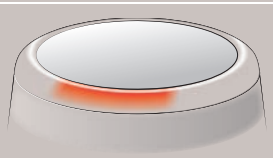
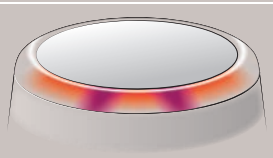
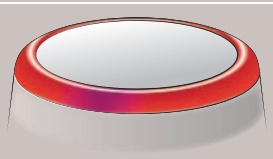
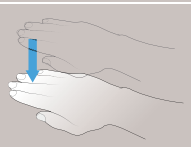
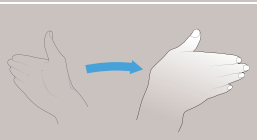
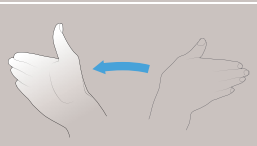
State	Function	Input	Audio	Visual feedback
Quiet/ sleep	Enters mode after being in Pause mode and not having detected a user for 1 minute.	-	-	 Light fades out, then off
Attention / Ready	Detected eye-contact. Means user is in-range. State required for all gestures.	Look at the speaker.	-	 Bright magenta spot pointing at each listener.
Playing discovery song	Playing proposed song that can be saved.	-	-	 Colour tone is cold turquoise. Brightness represents directivity.
Playing song from list	Indicates the song playing is part of inventory.	-	-	 Colour tone is warm orange. Brightness represents directivity.
Playing listed song for single listener, not engaged	Very directive. Sound and visual feedback follow listener.	-	-	 Narrow warm orange beam.
Playing listed song for two listeners, both engaged	Compromises for two users using two separate beams. Sound and visual feedback follow listeners.	Look at the speaker.	-	 Wide warm orange beam with magenta spots for each engaged user.
Error	Indicates there is an uneasy error	- Various inputs possible -	-	 Colour tone changes to red.

table 3 States and visual feedback as designed for the intelligent speaker

Command	Function	Input	Audio	Visual feedback
Resume last/ Pause	Resumes from last stop. Next day discovery level is reset to 0%, i.e. plays from favourites.	 Tap on an imaginary button mid-air in lign of sight with device.	Resumes feed	Colour fades in to either warm orange or cold turquoise.
	Pause		Music pauses	Colour fades out.
Volume up	Increases volume in 8 steps.	 Reach out to centre and move hand up fluently with palm facing up	Volume goes up + 'apple sound'	Circle fills clockwise in eight steps with white colour tone. Then returns to previous state.
Volume down	Decreases volume in 8 steps.	 Reach out to centre and move hand down fluently with palm facing down	Volume goes down + 'apple sound'	One-eighth of the circle fades out counter clockwise with white colour tone. Then returns to previous state.
Next track	Skips to next in queue. Listed songs are suggested a less often. Suggested songs won't be suggested in a while.	 Reach out to centre and move hand fluently to the right	Music skips to next track	A single swivel to the right.
Previous track	Skips back to beginning of song. Do it twice to go to previous song and so forth.	 Reach out to centre and move hand fluently to the left	Music skips to previous track	A single swivel to the left.

Save song	In discovery mode: save song. Stick song to list.	 Grasp at centre and move towards chest	-	Colour tone changes from cold turqoise to warm orange.
Forget song	While playing from list: remove song from list. Let go of it.	 Grasp at chest and move towards centre	Music skips to next track	Colour tone changes from warm orange to cold turqoise.

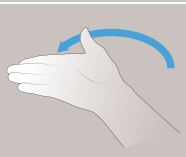
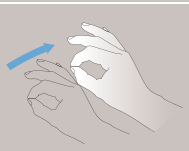
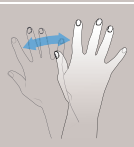
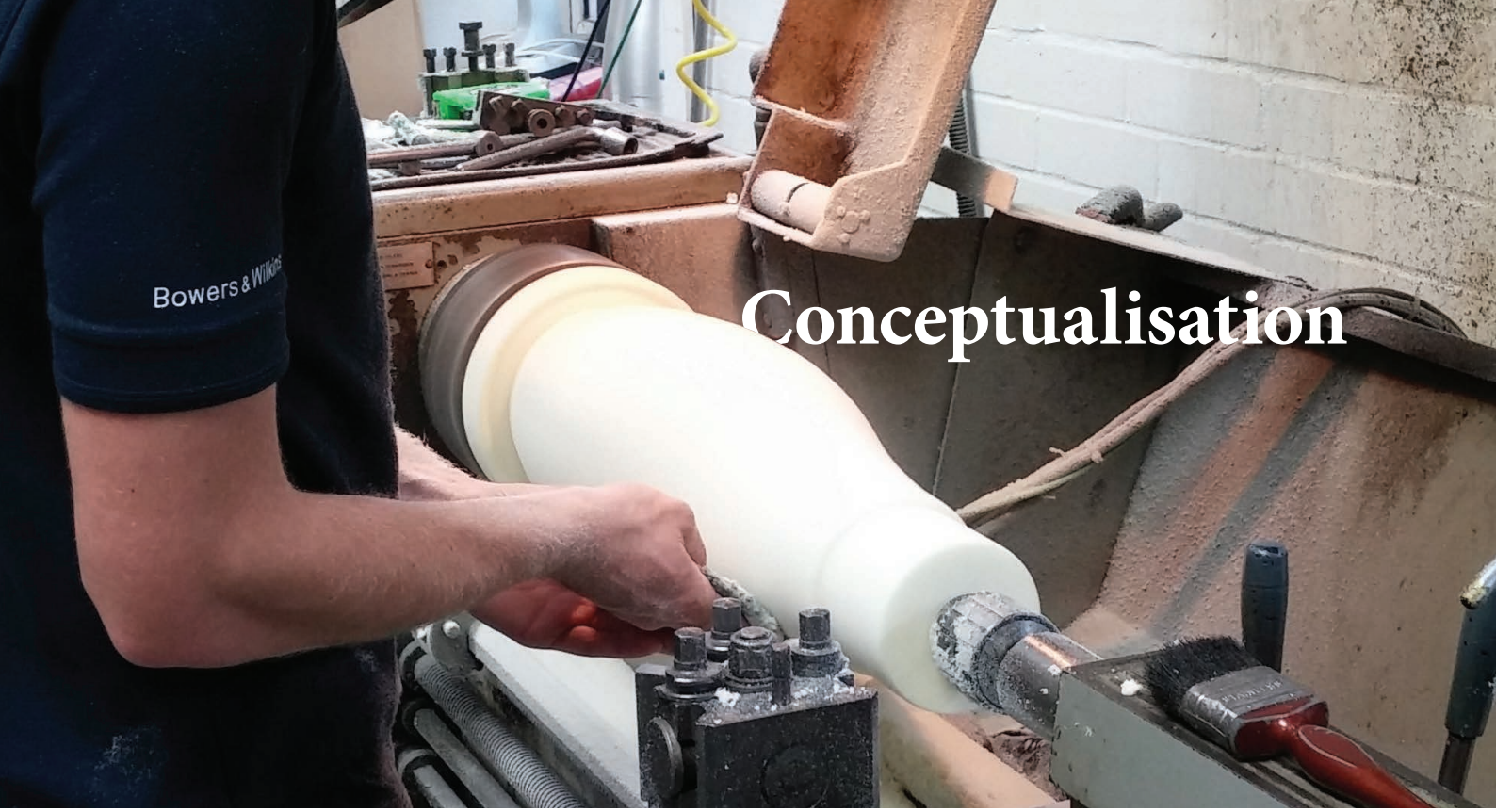
Command	Function	Input	Audio	Visual feedback
Play related next	Start track station. Queues up related songs. Discovery level changes to 100%.	 Grasp at eye height and hold for the ring to fill (2,5 seconds)	-	Ring fills with colourfull tones (purple-orange-yellow), then the original tone swivels continuously
Increase feed's discovery level	Sets discovery level to 0% -50% - 100%	 Wrap arm around 'music feed' and pull 'feed' towards you	"Playing from favorites", "Playing mix", "Playing full discovery mode"	-
Decrease feed's discovery level	Sets discovery level to 0% -50% - 100%	 Reach out to centre and push 'feed' away from you.	"Playing from favorites", "Playing mix", "Playing full discovery mode"	-
Fast forward	Fasts forward with movement to right.	 Reach out, then slowly move hand left or right with thumb and index finger together.	Music skips, volume fades low	A continuous pulsing swivel to the right.
Rewind	Rewinds with movement to left		Music skips, volume fades low	A continuous pulsing swivel to the left.
Undo input	Undo latest command. Continue at previous state. Works for both gesture and voice.	 Wave quickly at eye height with fingers apart	Returns to previous state	Quickly pulses for three times
Voice	Wake up voice control.	 Say "Alexa, ... " or raise hand (eye height, outside shoulder width) and hold for 2 seconds with falm facing the device	Volume fades low + Siri-like tone	Colour tone jumps to navy blue

table 4 Gesture commands and feedback as designed for the intelligent speaker



Conceptualisation

figure 31 Form study by sculpting foam

Conceptualisation

The foremost aesthetic attribute is honesty. The form should 'tell the story' and put the focus on the right areas of the product. The final form communicates the principle of a directive array of drivers that is assisted by a powerful bass unit. The shape induces the consumer to position the speaker as a centrepiece to benefit from the acoustic novelties to its full extent.

Story

The form should communicate this is a speaker with an array of drivers that not only emits sound in all directions but is also able to direct the sound towards the listener(s). A powerful bass unit then supports this directive array for the lower frequencies with an omnidirectional dispersion. The materials and finish should communicate that it is a concept speaker and not an actual product.

How

This speaker will be complex from the inside featuring state-of-the art electronics to sense the environment and as much as 25 drivers. But it must be simple on the outside and simple to use. Every aspect of the form should add to the story in a way. Every aspect that is irrelevant to the story should be eliminated to keep the product simple and understandable.

The awareness and intelligence should be communicated in a comfortable way expressing the same qualities as the UI (see page 40). Applied to aesthetics that means the visual feedback should be fluent and not overly present.

The form should invite the user to place the speaker anywhere in the room. The beauty of the speaker is it can adjust to any situation, so the use case should not be limited to specific placements in the room.

Requirements

The proportions of the concept are closely related to the hardware that is specified in section Product definition on page 31. 'Form follows function' does very much apply in this case. The overall shape will be a chunky mass with a 20cm diameter and 45cm height. The shape must draw the user's focus to the top part that produces the directive sound.

Additional requirement regarding to form can be found in the List of Requirements (see Appendix B).

Process and ideation

Method
Visit London perfume stores

The ideation phase was deconstructed in a few steps in order to diverge and converge. A visit to London perfume and cosmetic stores was an effective way of mapping existing shapes of similar proportions. The findings were then supplemented with search results from the internet. The most interesting shapes are aligned in Appendix F.

Method
Form study

At the start of this phase, while still designing for a configuration that was later considered sub-ultimate, the London store visit provoked a stimulus to make a full-size model. By shaping foam on a lathe a silhouette was created that gave insight in proportions, scale and form language. Later this shape became obsolete when the configuration changed to the one that is described on page 31. Nonetheless it should be mentioned as it contributed to the sketches.

Similarly, the sketches that were made early in the process became obsolete when the proportions changed due to changes in hardware. The initial work can be found in appendix G. With the latest configuration heavily restricting the variety of possible shapes the ideation could be channelled very effectively. It is for that reason that all sketches show great similarity (see fig. 33).

To be able to compare the sketches in a fair way all sketches were coloured with the same tones. The black rings represent the way the camera could potentially look after development.

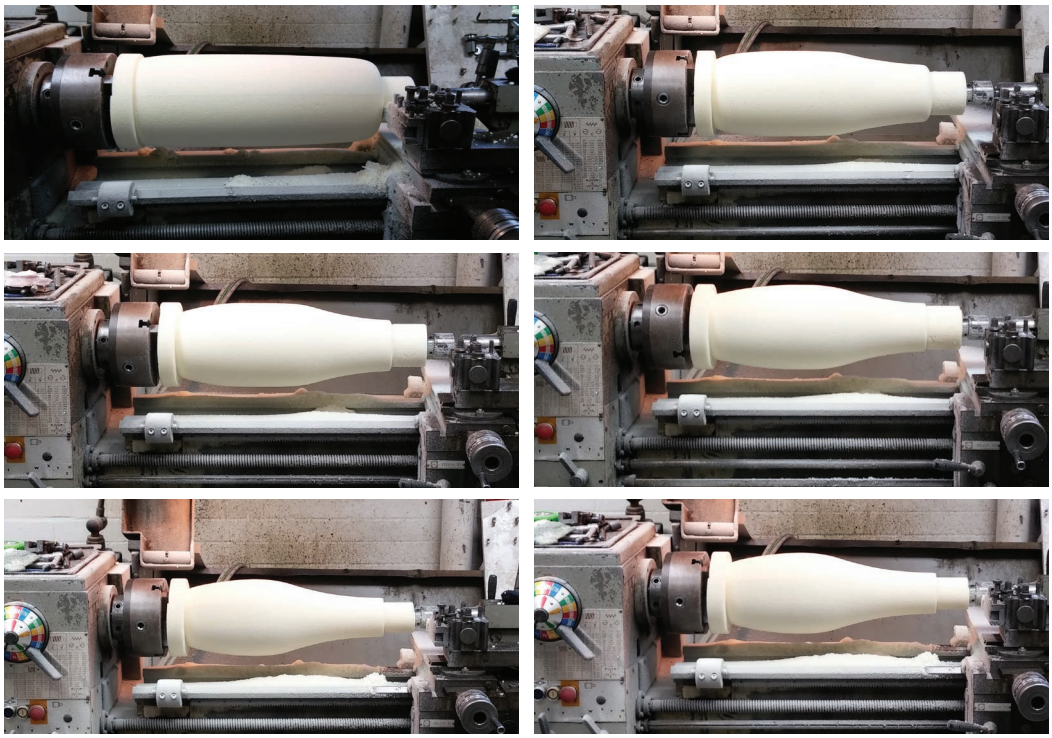


figure 32 The process of sculpting a shape out of foam

figure 33 Ideation sketches

Trade-off

The shapes were grouped in similarly looking shapes and the five most promising ones are compared for their overall value. The shapes' overall value is assessed on three main objectives, i.e. "aesthetic", "understandable" and "simple". The objective "Aesthetic" is evident. "Understandable is another way of saying the product should tell the story. "Simple" is the essential part to not alienate B&W's customers with such a complex concept. Scores are assigned from 1 to 5 with 5 being the highest. Ultimately, the weighted totals are compared.

Sketch 5 is very understandable but everything but simple. Sketch 4 explains the sound is radiated in all directions but does not allude to that is being done. Instead it puts a lot of focus on the lower part of the shape. Shape 3 is easy to manufacture but the faceted faces have no relation to the story that needs to be conveyed. Sketch 1 and 2 actually have a great similarity. Sketch 2 manages to draw the attention on the right part quite well and that only adds to the understandability. However, at the same time there is a lot going on what makes sketch 1 excel in simplicity. Therefore, shape 1 is chosen for further embodiment.

Method
Weighted objectives evaluation method

1 2 3 4 5

	Weight	Score	Total	Score	Total	Score	Total	Score	Total	Score	Total
Aesthetic	30	4	**120**	3	90	4	120	3	90	1	30
Understandable	50	3	**150**	4	200	3	150	3	150	5	250
Simple	20	5	**100**	3	60	3	60	2	40	1	20
	100		**370**		350		330		280		300

table 5 Trade-off of different form ideas

Concept detailing

The shape is very calm and simple but does not make it obvious what is going on the inside. By creating a split-line underneath the directive array, halfway down the product, the choice of materials could place more emphasis on the top part. However, manufacturing the concave shape of the mid-section would be complex.

The solution was to separate the core from the grille. The core functions as the cabinet for all 25 drivers and holds all the electronics. The grille consists of two halves that jointly create the silhouette of the product. Once assembled, a seamless knitted textile sock slides over the grille halves. Ultimately, both ends are sealed with an end cap, i.e. a bottom plate and a top plate (see fig. 36).

Light feedback should be fluent (see p. 40). Philips Luminous Textile creates this experience (see fig. 35). However, in order to achieve a similar consistency, the LEDs require a large spacing to the grille or the number of LEDs need to be increased drastically. Other ways have been investigated to fully illuminate the cabinet or the grille. Optic fibres or light guides were also considered. However, under the pressure of time the light feedback is communicated using a simple but effective ring of LEDs that is visible through a frosted acrylic diffuser.

In order to create coherency between ongoing developments the grille features a pattern that is used for other products as well. It makes the concept slightly less calm and there is no acoustic reason to choose such pattern. However, the concept has to share B&W's iconic form language.

figure 34 The chosen shape

figure 35 Philips luminous textiles (Philips, n.d.)

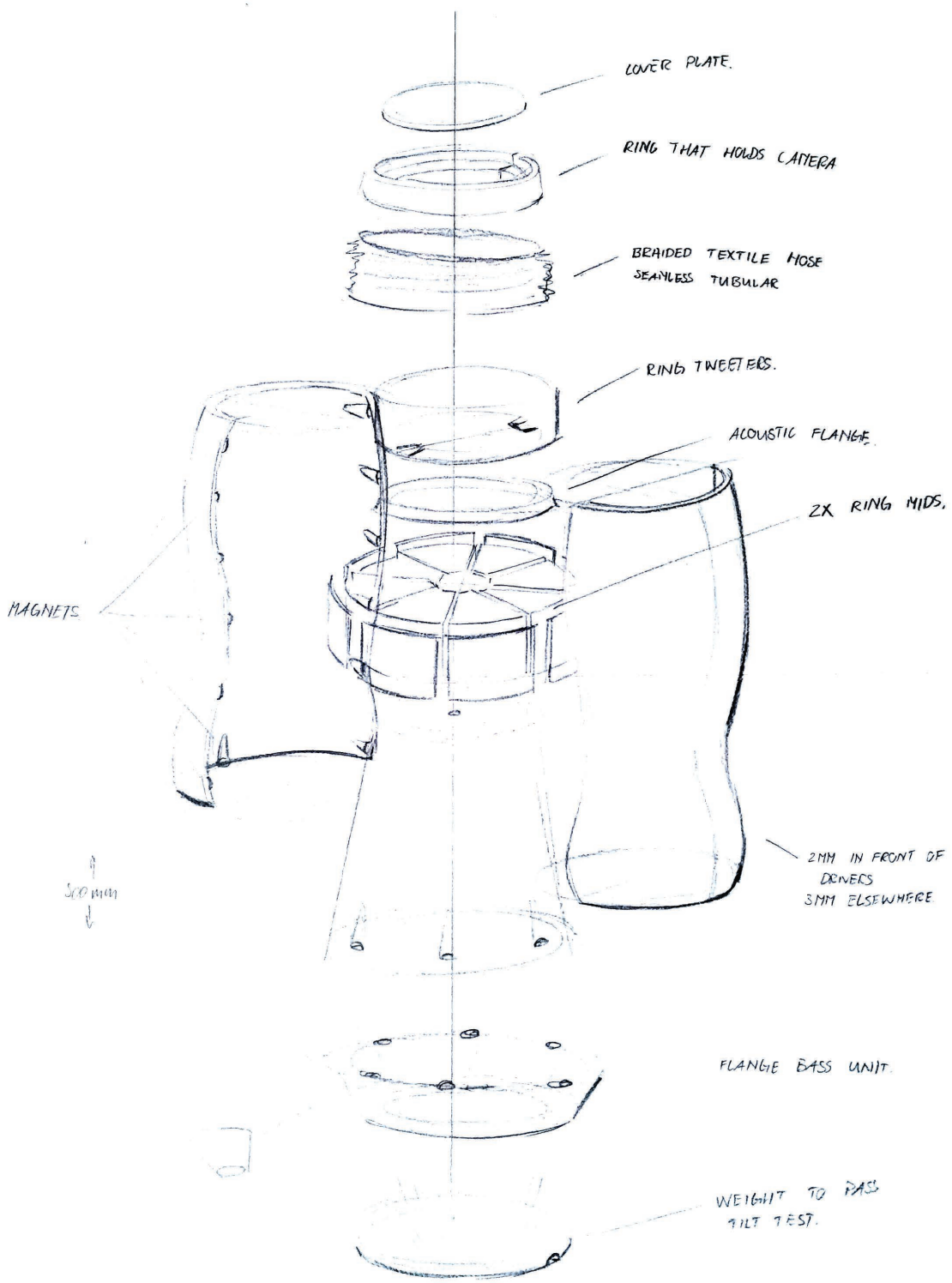
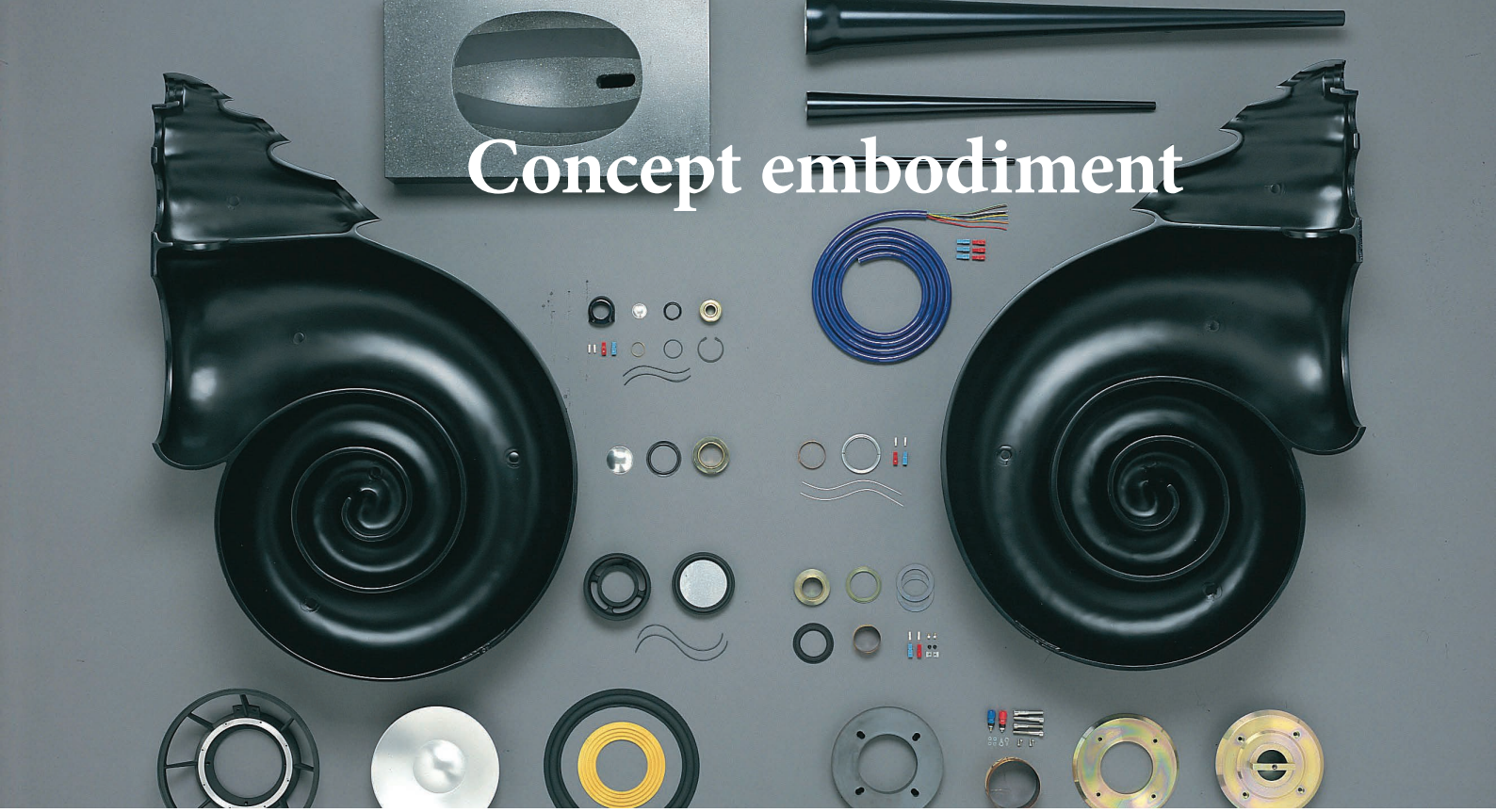


figure 36 First sketch of product assembly



Concept embodiment

figure 37 Exploded view of the B&W Nautilus

Concept embodiment

The final design has a facetted grille and will come in three different finishes. For further testing a prototype is developed in the Cotton White finish. The exterior is used to proof acceptance and can later be equipped with a core to do acoustic measurements with.

Content

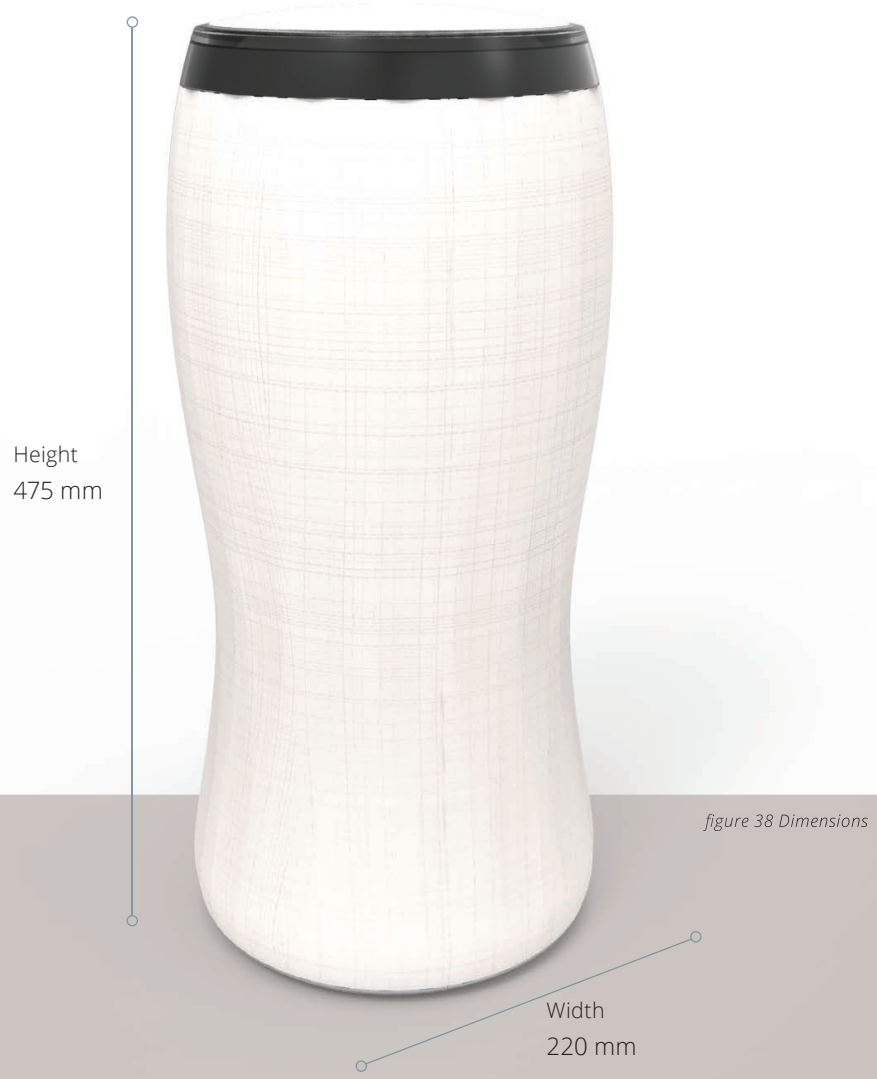


figure 38 Dimensions

Model
Intelligent speaker

Description
Directive speaker with voice- and gesture control

Drive units
8 x 25mm (1 in) tweeter
16 x 50mm (2 in) midrange
1 x 150mm (6 in) woofer

Amplifier output
8 x 25W tweeter
8 x 25W midrange
1 x 50W woofer

Microphones
Eight-microphone array for voice control and room sensing.

Optical sensors
4 x 90-degree infrared sensor
4 x 90-degree infrared emitter

Input voltage
100V – 240V – 50/60Hz

Height
475mm (18.7 in)

Width
220mm (8.7 in)

Weight
9.6kg (21.2 pounds)

Product finish
Midnight Blue or Cotton White

Weight

To estimate wieght drive units are measured, The weight of the plastic parts are extracted from the 3D solidworks model. All other parts are compared to data from the Zeppelin. The product's weight is estimated on a total of 9,6kg.

Part	Zeppelin	Intelligent speaker
Drivers:		
1x Zep LF	1540	1540
Zeppelin HF and MF	1140	n/a
16x T7	n/a	1344
8x M1 HF	n/a	344
Electronic Module:		
Amp board	100	300
Signal Processing Board	150	150
Optical Environmental Sensor	n/a	± 500
Microphone and Display board	n/a	± 200
Other	230	230
Power Supply	250	750
Plastics/ Metal	2500	2766 (CAD)
Grille	390	704 (CAD)
Packaging	n/a	n/a
Labels/ Gaskets/ Glue	-	-
Cables/ AC cord	200	800
Total weight	6,500 g	9,628 g

table 6 Product weight

Tilt test

Test 1: Position on a surface on a 10 degree slope while being rotated 360 degrees.

Assumed concept is axisymmetric with C_G = (0, 0, 235) mm.
Tipping point is when C_G and A are aligned vertically.
d_{Gy} = 235 mm
d_{Gx} = 93.5 mm
$\vartheta_{TIP} = TAN^{-1}$ (93.5 / 235) = 23.1°

PASS: Tips over at 21.7° slope

Test 2: Subject to a sideways force in the least stable position on the highest point of the product. The applied force equals 13% of the product weight.

Starts to tilt when F_C = 0 N.
m = 9628 g
$F_G = m \cdot g$
d_B = 465 mm
d_G = 93.5 mm
ΣM = 0
$F_B \cdot d_B = F_G \cdot d_G$
$F_B = 9.628 \cdot 9.81 \cdot 93.5 / 465 = 19.0$ N
$F_{B\text{-}MAX}$ = 19.0 N what equals 1.94 kg
1.94 kg is 20.1% of the concept's mass

PASS: Tips over when 20% of mass is applied sideways

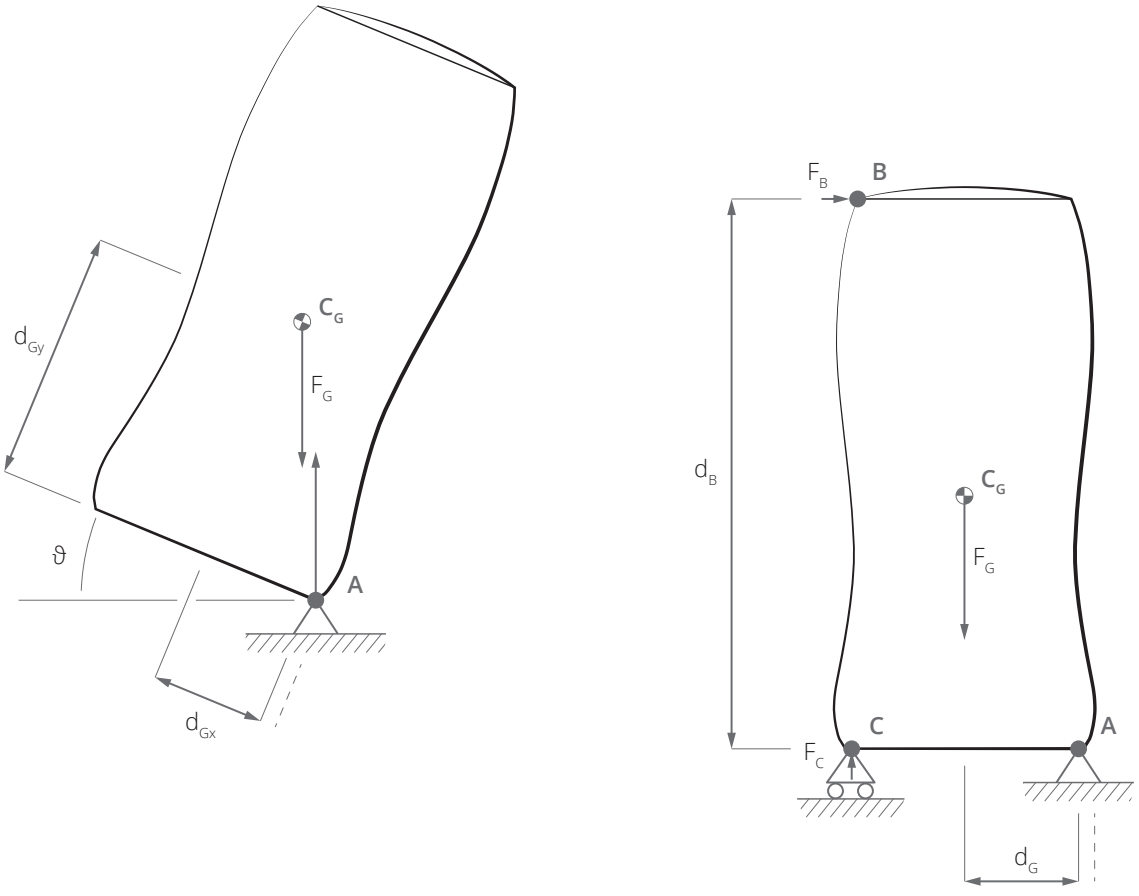


figure 39 From left to right: tilt test 1 and 2

Aesthetics

Aesthetical values are already defined on page 46. The chosen form is elegant and simple. The choice of material still follows the interaction statement.

For this product there is no black or space grey finish available. The shape is not masculine and the finish is not either. As stated on page 27 the product should be less male oriented than previous activities.

The intelligent speaker comes in three colours; cotton white, silver grey and midnight blue. Every variant is mainly cloth because of its ideal acoustic transparency. The cotton white and silver-grey version come with a soft top plate. The midnight blue version comes with an etched aluminium top. All aluminium parts are etched to add a nice tactile experience and a premium finish.

A prototype will be developed for a proof of acceptance of amongst others the shape and finish. The prototype that is being developed will be Cotton White as it makes the speaker look progressive and state-of-the-art.

figure 40 Product finishes

Manufacturability

Production costs

The BOM costs are compared to that of the existing Zeppelin. With the targeted retail price a mark-up ratio of over 3 could be possible.

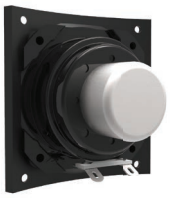
BOM cost

Part	Intelligent speaker
Drivers:	
1x Zep LF	6.88
Zeppelin HF and MF	n/a
16x T7	69.44
8x M1 HF	26.88
Electronic Module:	
Amp board	55.44 (18ch)
Signal Processing Board	30.87
Silicon Module	± 30.00
Optical Environmental Sensor	± 100.00
Microphone and Display board	± 20.00
Other	16.39
Power Supply	39.30 (450W)
Plastics/ Metal	118.71 (300%)
Packaging	15.30 (150%)
Labels/ Gaskets/ Glue	3.42
Cables/ AC cord	9.72
Licensing	3.60
OEM/ Manufacturer cost	40.00
BOM cost	$585.95
Retail price	$1,799.00
Ratio	3.07

Table 7 BOM costs

Assembly

Sub-assembly 1: Install electronics to LF cabinet section and wire all components. Qty: 1



Sub-assembly 2: T7 Flange. Screw T7 unit to flange. Wire driver and dcrew flange and driver onto cabinet. Qty: 16

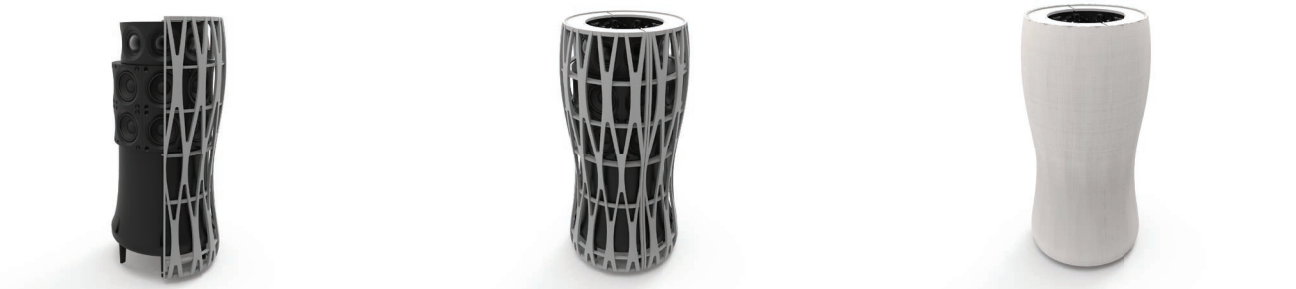
Sub-assembly 3: M-1 tweeter flange. Screw M-1 tweeter to flange. Qty: 8

Sub-assembly 4: Wrap Adafruit Neopixel string around LED flange using dubble sided tape and glue 6 Neodymium magnets in place. Glue LED diffuser to LED flange. Glue camera shades to LED diffuser. Qty: 1

Sub-assembly 5: Glue 6 Neodymium magnets to top plate. Wrap cloth around top plate and weld it around the corners, then cut off excess. Qty: 1

Position LF flange on a jig that makes the pins float. Screw Zeppelin bass unit into place. Screw sub-assembly 1 into place.

Add 8 times sub-assembly 2 and wire drivers. Add another 8 on top. Add 8 times sub-assembly 3 and wire all drivers.



Slide one grille-half on side-ways. Slide the second half on (both halves have a form fit but are still floating at this point. Slide over the seamless grille cloth as a sock and weld edges to grille.

Screw bottom plate to LF flange and screw both grilles to bottom plate. Cover up screw holes with adhesive rubber. Turn product around.

Add sub-assembly 4 and screw to top of grille parts. Add Arduino and place top plate. The concept is now ready for testing.

figure 41 Product assembly

Market introduction

It's an audio system, that sounds great anywhere in the room. B&W's product management set the price point of the intelligent speaker project between $1500 and $2000.

Compared to its direct competitors (see fig. 43) it is by far the most advanced and automated speaker.

The customer pays for a statement in technology and user experience. Would audio enthusiasts appreciate voice- and gesture control? Would someone that like quality appreciate sound beaming?

Considering the price point, the intelligent speaker will mainly sell through the core channels. Only a fraction will sell through premium department stores.

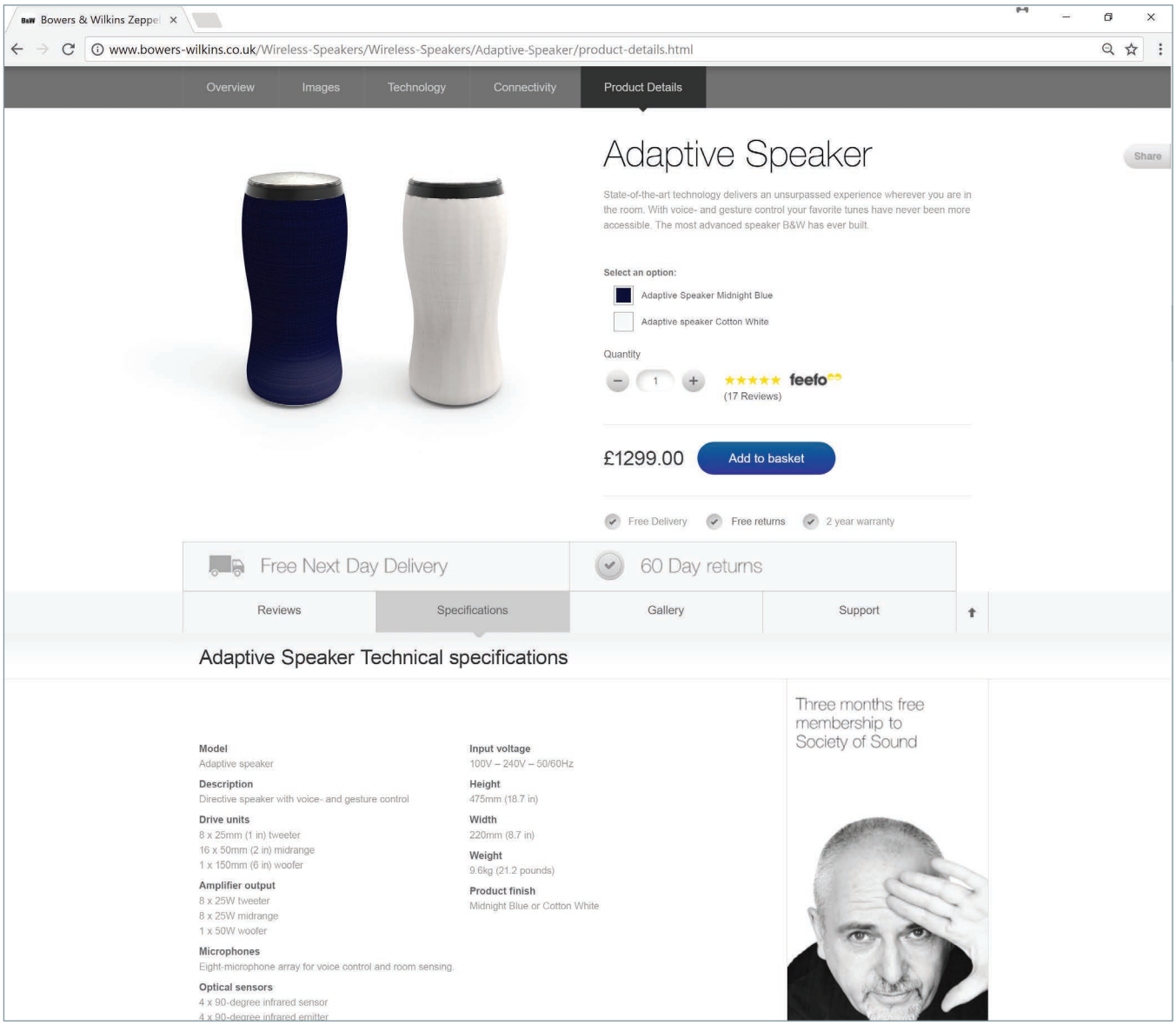


figure 42 What the website could look like after product launch

Name	B&O Beolab 90
Price	£58.820
Steerable array	Yes, manual
Interaction	App

Name	Apple HomePod
Price	$349
Steerable array	Yes
Interaction	Siri

Name	Dynaudio Music 7
Price	£875
Steerable array	No
Interaction	Buttons for favourites

Name	Lexicon SL-1
Price	--
Steerable array	Yes
Interaction	Unknown

Name	Sonos One
Price	£199
Steerable array	No
Interaction	Amazon Alexa

Name	B&W Intelligent Speaker
Price	£1299
Steerable array	Yes
Interaction	Voice- and gesture control

figure 43 Product landscape

Cross-platform

The beauty of the interaction is that it is not limited to the intelligent speaker. The interaction, or parts of it, can be applied to any of B&W's lifestyle speakers.

Gesture control and voice control even more so can be integrated in any product by integrating microphones, cameras and some LEDs for visual feedback. The vast majority of users would put the speaker back to- or close to the wall. So a 180-degree camera coverage would fulfill.

In a more simplified way the speaker can sense the user's presence. Imagine you walk in the lounge after a long day at work. At the moment you enter the room the speaker recognises you and increases the brightness of the ring of light and is ready to go.

Some elements of the customer applications could even be applied to B&W's wireless performance speakers.

A microphone array would be a first step towards room proofing.

With a stereo pair of speakers, a simple RGB-camera with no IR filter and some clever DSP a primitive version of beamfollowing can be achieved. In practive this means sound from one channel can be delayed in order to put the listener in the ideal listening position.

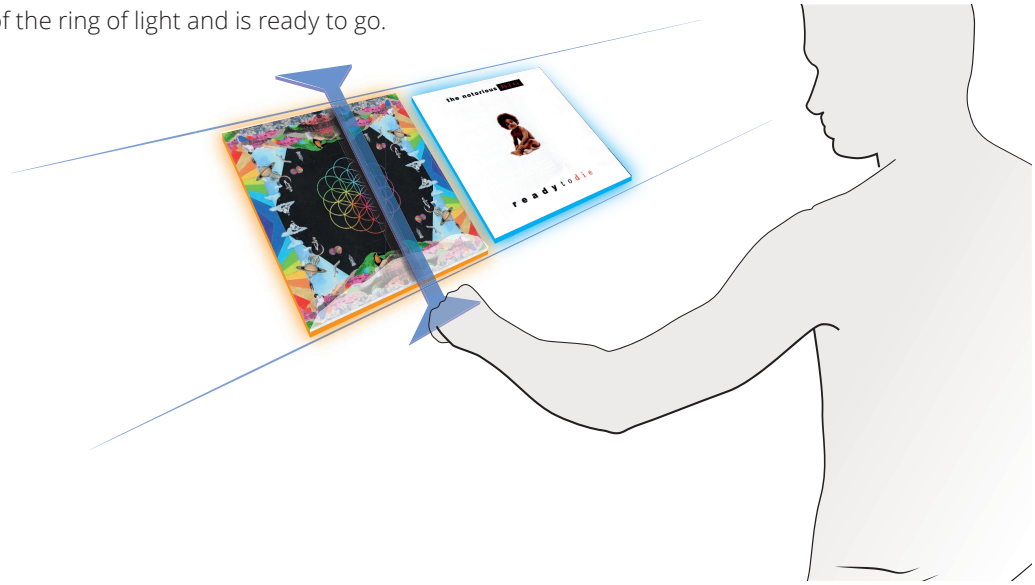


figure 44 A soundbar from B&W's wireless lifestyle range outfitted with the Intelligent Speaker interaction

Prototype

A last prototype is under development to proof acceptance of the exterior and interaction. It is basically a shell that can be outfitted with a core to do acoustic measurements with. These measurements should proof the concept of the acoustic principles and associated benefits before further developments.

The prototype is a subset from the complete model. It basically consists of all visible parts. The prototype uses a Microsoft Kinect 2.0 for gesture control, an Amazon Alexa for voice commands and an existing speaker for the acoustics.

figure 46 Aesthetic shell and acoustic core

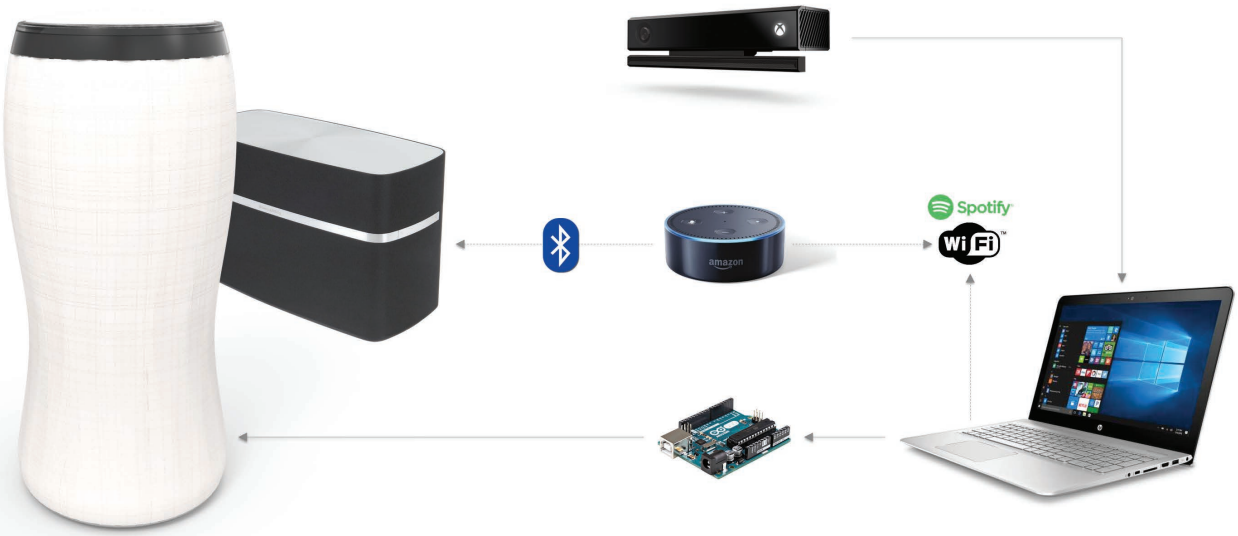


figure 47 Diagram of working prototype

Conclusion

figure 48 Suggested product range

Conclusion

The concept should inspire to develop a lifestyle speaker featuring the B&W interaction of the future. The interaction could well be adapted cross-platform. If there is a strong market pull for directivity- and beam width control this can be introduced in a separate product.

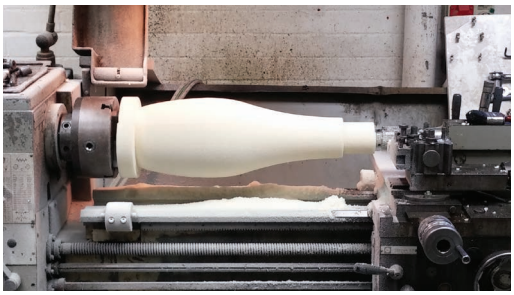
The objective for this project was to: "Design a unique B&W user experience for the intelligent speaker". The truth of the matter is that the intelligent speaker is a few steps too far ahead for B&W as of this writing. The project's activities were carefully re-framed as concept speaker and form an inspiration of a possible direction to take. The intelligent speaker really is an all-in-one solution that is not quite an incremental step forward from B&W's current activities. In reality there are most probably another two product cycles required to confidently present a solution with the intelligent speaker's complexity.

During the course of the project it has been confirmed in many ways that the project is going in the right direction. In the meantime, Sonos has launched a voice-controlled speaker; the Sonos One. In January 2018 Apple will launch a directive speaker with voice control; the Homepod.

Recommendations

Due to the complexity it is recommended to split developments in three categories; user-product interaction, steerable array and an adaptive platform.

The adaptive platform is a sensible first step forward from current developments and can become a range-wide implementation of statement technology. A second step would be to invest in user-product interaction. A steerable array can follow last.



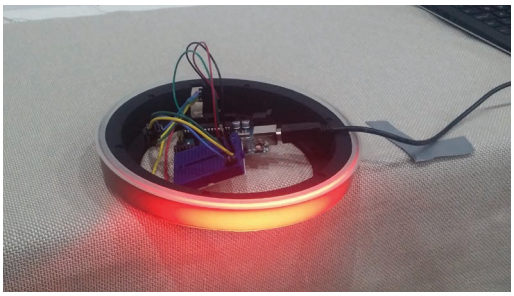


figure 49 Different prototypes used for the thesis

References

2CV (2015a, November). “Bowers & Wilkins Opportunity Exploration – Integrated Results”. Presentation for B&W on market research by a 5-minute survey among n=2000 respondents in both the USA and UK.

2CV (2015b, November). “Bowers & Wilkins Opportunity Exploration – Integrated Results”. Presentation for B&W on market research by a 30-minute survey among n=750 premium headphone owners / intenders and n=750 premium speaker owners / intenders in both the USA & UK + ~700 Bowers customers in each market.

Alexa, Say What?! Voice-Enabled Speaker Usage to Grow Nearly 130% This Year (2017). Retrieved from: https://www.emarketer.com/Article/Alexa-Say-What-Voice-Enabled-Speaker-Usage-Grow-Nearly-130-This-Year/1015812

B&W (2014, September). Internal presentations and documents

B&W (2016a, January). Internal suggested retail price list January 2016

B&W (2016b). Internal Zeppelin sales figures.

Bourke, P. (April, 2009). Processing images from the Lucy 360-degree video camera. Retrieved Sep-tember 21, 2017, from http://paulbourke.net/dome/LucyCamera/

Bryant, M. (2016). “Samsung’s new fridge has a built-in grocery store”. Retrieved March 4, 2016, from http://thenextweb.com/gadgets/2016/01/05/samsungs-new-fridge-has-a-built-in-grocery-store/

C&R (2006, July). “Bowers & Wilkins Customer Segmentation Study”. Presentation for B&W on market research by conducting nearly 1,200 online surveys among B&W customers.

Charlton, A (2016). “Volvo to replace car keys with smartphone app in 2017”. Retrieved March 4, 2016, from http://www.ibtimes.co.uk/volvo-replace-car-keys-smartphone-app-2017-1544644

Cox, J. (2015). “Your encryption will be useless against hackers with quantum computers”. Retrieved March 4, 2016, from http://motherboard.vice.com/read/your-encryption-will-be-useless-against-hackers-with-quantum-computers

Cox, J. (2016). “Prizm plays music based on tastes of people in the room”. Retrieved March 4, 2016, from http://www.whathifi.com/news/prizm-plays-music-based-tastes-people-in-room

Davies, J.B., Sandström, S., Shorrocks, A.B., Wolff, E.N. (2009). "The Level and Distribution of Global Household Wealth". Retrieved July 26, 2017, from http://www.nber.org/papers/w15508

Enge, E. (2017). " Rating the Smarts of the Digital Personal Assistants". Retrieved August 15, 2017, from https://www.stonetemple.com/digital-personal-assistants-test

Frog Design (2016). “Tech Trends 2016”. Retrieved March 4, 2016, from http://www.frogdesign.com/techtrends2016/

Funnekotter, W (2016). “Avegant Glyph Preview - Mobiele bioscoop verpakt in een koptelefoon”. Retrieved March 4, 2016, from http://tweakers.net/reviews/4311/avegant-glyph-mobiele-bioscoop-verpakt-in-een-koptelefoon.html

Gartner (2015). “Gartner Identifies the Top 10 Strategic Technology Trends for 2016”. Retrieved March 4, 2016, from http://www.gartner.com/newsroom/id/3143521

Gartner (2016). “Gartner Identifies the Top 10 Internet of Things Technologies for 2017 and 2018”. Retrieved March 4, 2016, from http://www.gartner.com/newsroom/id/3221818

Garun, N. (2015). “Amazon Echo review: Alexa’s a great listener, but is awful at search”
Retrieved March 4, 2016, from http://thenextweb.com/gadgets/2015/07/08/alexa-y-u-no-answer/

Ghoshal, A. (2016). “These $299 wireless earphones track your steps and heart rate”. Retrieved March 4, 2016, from http://thenextweb.com/insider/2016/01/06/these-299-wireless-earphones-track-your-steps-and-heart-rate/

Gonser, T. (2016). “5 Things That Will Disappear In 5 Years”. Retrieved March 4, 2016, from http://techcrunch.com/2016/01/03/5-things-that-will-disappear-in-5-years/#.xwdlfw9:9cSa

Hijink, M (2017, March 22). De Apple van de audio. NRC, pp E5.

Hill, M (2017). Meet the Designers: Aston Martin DB11 [Video File]. Retrieved from https://www.formtrends.com/in-depth-look-aston-martin-db11-design/

Homepod (2017). Retrieved from https://www.apple.com/homepod/

IB Times (2016). “Apple TVs Could Be Controlled Via Eye And Hand Gestures, Apple Inc. Patent Reviels”, retrieved March 17, 2016, from http://www.ibtimes.com/apple-tvs-could-be-controlled-eye-hand-gestures-apple-inc-patent-reveals-2336781

InfoComm International (2016). “Market research makes sense of big data”, retrieved March 4, 2016, from http://www.infocomm.org/pub/img/BigDataInfographic.pdf

Ipsos (2013, July). “Brand Growth Study UK & US Headphones and Wireless Speakers”. Presentation for B&W on market research by a 20 minute online self-completion survey in the UK and US among n=800 premium headphone owners or intenders and n=400 wireless speaker owners or intenders.

Lopez, N. (2016). “Razer’s $999 Blade Stealth is basically a Macbook Air made for gamers”. Retrieved March 4, 2016, from http://thenextweb.com/gadgets/2016/01/08/razers-999-blade-stealth-is-basically-a-macbook-air-made-for-gamers/

Maimone, A. and Fuchs, H. (2012). Real-Time Volumetric 3D Capture Of Room-Sized scenes For Telepresence.

Microsoft (2014). “Visual Gesture Builder: A Data-Driven Solution to Gesture Detection”, retrieved April 13, 2016, from http://aka.ms/k4wv2vgb

Møller, M., Olsen, M., Agerkvist, F., Dyreby, J. and Munch, G. (2010). Circular Loudspeaker Array with Controllable Directivity (Audio Engineering Society convention paper 8012). Retrieved March 4, 2016, from http://www.aes.org/e-lib/browse.cfm?elib=15309

Ossic X (2016). “OSSIC X: The first 3D audio headphones calibrated to you”. Retrieved March 4, 2016, from https://www.kickstarter.com/projects/248983394/ossic-x-the-first-3d-audio-headphones-calibrated-t

Philips (n.d.). Play with light, texture and dynamic content. Retrieved October 30,2017, from http://www.lighting.philips.ca/products/luminous-textile

Piumsomboon, T., Clark, A., Billinghurst, M. and Cockburn, A. (2013). User-Defined Gestures for Augmented Reality. Retrieved October 20, 2017, from https://hal.inria.fr/hal-01501749/document

Rau, T. (Guest speaker). (2015, November 8). VPRO Tegenlicht [Television series episode]. Einde van bezit. The Netherlands: VPRO. Retrieved March 4, 2016, from http://tegenlicht.vpro.nl/afleveringen/2015-2016/einde-van-bezit.html

Schönauer, C. and Kaufmann, H. (2013). Wide Area Motion Tracking Using Consumer Hardware. Pro-ceedings of the The International Journal Of Virtual Reality, Vol. 12, No. 1.

Sinek, S. (2009, September). How great leaders inspire action [Video file]. Retrieved from https://www.ted.com/talks/simon_sinek_how_great_leaders_inspire_action

Thiel, P. (2014). Zero to one: notes on startups, or how to build the future. New York City, NY: Crown Business.

WhatHiFi (2016). “MQA hands on”. Retrieved March 4, 2016, from http://www.whathifi.com/mqa/review

Wokke, A (2016). “Google voegt optie voor 'ruimtelijke audio' toe aan sdk vr-bril Cardboard”. Retrieved March 4, 2016, from http://tweakers.net/nieuws/107324/google-voegt-optie-voor-ruimtelijke-audio-toe-aan-sdk-vr-bril-cardboard.html

Woods, B. (2016a). “The next version of Wi-Fi doubles range, but it’ll be years before you can use it”. Re-trieved March 4, 2016, from http://thenextweb.com/insider/2016/01/04/the-next-version-of-wi-fi-doubles-range-but-itll-be-years-before-you-can-use-it/

Woods, B. (2016b). “BMW’s heads-up display for helmets could make riding a motorbike safer”. Retrieved March 4, 2016, from http://thenextweb.com/gadgets/2016/01/06/bmws-heads-up-display-for-helmets-could-make-riding-a-motorbike-safer/

Williams, O. (2016). “This radical rethink of how computer interfaces work by a 21-year-old designer is amazing”. Retrieved March 4, 2016, from http://thenextweb.com/dd/2016/01/21/this-radical-rethink-of-how-computer-interfaces-work-by-a-21-year-old-designer-is-amazing/#gref

Appendices

Appendix A - Market trends

In order to get a feel for the design context of the intelligent speaker relevant market trends are identified. The following list is a bundle of personal observations of market developments, latest tech-products and predictions of professional trend studies.

Home entertainment in 2020 will see...

... user interfaces becoming more intuitively.
The way we currently interact with screens will change, potentially making dialog windows superfluous (Williams, 2016).

... high-res streaming audio as standard.
High resolution audio gets available to the mass. Creating more value for the speakers that can orchestrate this.
The Master Quality Authenticated (MQA) codec already makes 24-bit/192 kHz audio easier to stream. Services like Tidal are winning ground (WhatHiFi, 2016).

... learning products.
Devices for Smart Home will learn from your input. Audio devices that learn from the users' input (like Prizm) will gain interest (Cox, 2016).

... an ambient user experience.
Software will be cross-device and cross-platform. Making use equal for phone, tablet, laptop or fridge. Enabling us to access our data from an expanding number of devices (Gartner, 2015).

... a shift from product to service.
Our identity is not linked to our devices anymore but to our data, making the device just an enabler. Similarly, we do not buy light bulbs and pay the energy bill. We just buy light (Rau, 2015).

... first applications of virtual reality.
"In the near future, VR will become a common treatment option for PTSD, exposure therapy, chronic pain, and other conditions" (Frog Design, 2016).
"Online webcams are eclipsed by connected VR cams, which allow viewers to virtually transport themselves to points in space around the world and interact with the people there" (Frog Design, 2016).

... an increased value of big data.
Information of everything: information analysis techniques will bring meaning to the often-chaotic deluge of information. There will be a greater cooperative interaction between devices (Gartner, 2015).

... China as biggest cinema market of the world.
The global pro-AV market currently is twice as big in Asia-Pacific and North-America as it is in Europe (InfoComm International, 2016).

... multi-room audio as a part of IoT.
More and more suppliers of smart home devices will also offer multi-room audio.

... autonomously personalised music streams.
Listening habits have changed over the last years from listening to your old record to assembling a playlist. Streaming services will add intelligence to their products that tailors the content to the user's mood. More people

will swap their music collection for streaming services.

... cloud computing for the mass.
All computing power will be utilised somewhere in the cloud. The only thing we need is a device that serves as an interface to our data and a plan that decides how many cores you can use (Lopez, 2016a).

... less wires.
Cables are from the past. Headphones and now even earplugs can be wireless. More headphone jacks will become old-fashioned and home devices will become wireless with the possibility to charge wirelessly (Ghoshal, 2016).

... fridges with a built-in grocery store.
Samsung already took the first steps (Bryant, 2016).

... a run for the most desired IoT platform app.
Smart Home devices are counting up quickly in numbers, now requiring managing apps (personal observation ISE fair Amsterdam, February 11, 2016) (Gartner, 2015).

... closer collaboration and integration of products.
Conventional speakers are made 'wireless'. Amplifiers, sources and connectivity devices will be integrated (Gartner, 2015).

... an autonomous agent as main user interface.
"An autonomous agent becomes the main user interface for ambient user experience" (Garun, 2015). Voice controlled personal assistants (like Amazon Echo) will see more applications.

... some heavily loaded Wi-Fi networks.
Wi-Fi networks will be heavily loaded, and evolve to keep up (Woods, 2016a).

... no more remote controls and keys.
Remote controls will disappear. Volvo's next model will replace car keys with smartphone app (Charlton, 2016) (Gonser, 2016).

... immersive audio for VR.
First 3D headphones will be there in 2017. Main application for VR goggles (Ossic X, 2016) (Wokke, 2016) (Funnekotter, 2016).

... bow-kite shaped TV screens.
Screens are already curved but may morph to even more realistic shapes. Screens will to go from flat, to curved, to panoramic, to an approximation of a sphere.

... haptic feedback.
"This year, haptic feedback changed the way we interact with smartphone screens, watches, and laptop mouse pads – helping bring the sensation of texture, vibration, and motion to those interactions" (Frog Design, 2016).

... moral debates about safety concern with IoT.
Security risks on IoT will remain an ongoing topic in 2017 and 2018 (Gartner, 2016).

... no more passwords.
No password will be safe with quantum computing. Password will make room for a more secure way of access (Cox, 2015).

... heads-up displays in motorcycle helmets.
And bicycle glasses (Woods, 2016b).

... ear plugs with live EQ tuner.
Doppler Labs' Here is a live EQ tuner that you wear as ear plugs.

... gadgets that monitor your health.
With enhanced accessibility of information, data on your own body gets more meaning (Lopez, 2016a).

Appendix B - List of requirements

List legend
Regular font: requirements
Italic font: wishes

General requirements lifestyle products

1. The product must fit the B&W portfolio.
1.1. It must comply with form language of current developments.
2. The product must sell through B&W's existing channels.
2.1. The product must sell through premium department stores
2.2. The product must sell through B&W Hi-Fi dealers.
3. It should address a wider audience than the conventional Hi-Fi.
3.1. It should appeal to both the male and female buyer.
4. The customer must be able to scale the system.
4.1. The product must be able to communicate with other B&W 'wireless' products.
5. It must allow for use from multiple users.
6. It must pass two tilt tests.
6.1. It may not fall over when placed on a surface with a 10 degree slope while being rotated 360 degrees.
6.2. It may not fall over when subjected to a sideways force in the least stable position on the highest point of the product. The applied force equals 13% of the product weight.

Directive array - acoustics

7. It must have drivers all-round.
8. It must be able to steer each driver independently.
9. It must know the position of the listener(s) in the room.
10. The product must be considerate of its environment.
11. It must dynamically compensate for changes in room character.
12. The sound quality must at least equal that of the existing Zeppelin Wireless.
13. It must have 8 M1 tweeters, 16 T7 midrange units and 1 Zeppelin base unit.
14. The tweeters must be on top, the bass unit at the bottom.
15. The bass unit must be firing downwards.
16. The tweeters, midrange units and the bass unit must have at least three different enclosures.
17. It must be able to dissipate the heat of a 450W power supply.
18. The HF drivers should ideally be positioned at ear height, i.e. between 105cm and 120cm from the floor.

Directive array – marketing

19. It must be a single box solution.
20. It must serve a domestic use case.
21. The price point must be within the $1,500 to $2,000 range.
21.1. It must retail at a mark-up ratio of at least 3.0.
21.2. The BOM cost may not exceed $500 USD.
21.3. Would ideally retail at a mark-up ratio of 4.0.
22. It must be a statement in technology.

22.1. It must feature technology that is beyond what is available for today's speakers.
23. It should have a lifetime of over 10 years.
23.1. The hardware should be architected to support future updates.
24. It should feature the latest connectivity.
24.1. It should integrate with any IoT platform.
25. It should not directly take on the mid-segment of 'wireless' audio.
25.1. It should be positioned in the premium spectrum.
25.2. It should address a discerning listener.
25.3. It should address a music lover.

Interaction specific

26. It may not have buttons, or use remotes and apps.
27. It should feature an intuitive interaction.
28. It should be a turn-the-key solution.
28.1. Time to music should approach 'zero' seconds.
28.2. Operating the system should require low-involvement.
29. It should make a statement in every possible way.
29.1. It should impress within the minute.
29.2. It should have the X-factor to sell it.
30. It should give the customer a unique audio experience.
31. It should allow for skill transfer.
31.1. Other B&W 'wireless' products should be able to adopt the intelligence platform of the speaker (not necessarily the directivity).

Form

32. The form should be immediately communicative about the intelligence and directivity.
32.1. The exterior should support the user applications.
33. The exterior should be refined, elegant and classy.
34. The choice of materials should reflect the craftsmanship that goes into the product.
35. It should have a small physical footprint.
36. The physical volume that is not used for acoustic purposes or to accommodate the hardware should be as close to 'zero' as possible (in order to manage customer expectations).

Appendix C - Pilot 1

Content

Method

Four respondents were invited to one of the listening rooms of the Research Establishment Steyning for a 60-minute timeslot. The respondents were asked to get acquainted with a music system that features make-belief gesture control and were asked for their opinion afterwards.

Location:	B&W AV Room Steyning
Apparatus:	Laptop, beamer, iPad, prototype, Zeppelin Wireless speaker.
Duration:	50 minutes
Sample size:	n=4

INTRODUCTION 10 MIN

All four respondents came in completely blank so the entire test was supported with a Power Point presentation. After an introduction on the subject and the limitations of the set-up, the respondents were asked about their listening habits to get to know him or her as a listener. Questions included:

- How important is listening to music for you?
- How important is sound quality for you?
- How do you typically browse for music?
- What is your typical involvement in different scenarios?

HAND GESTURES 10 MIN

A set of different hand gestures is explained by showing illustrations (see table 8) on the projector screen. Afterwards the respondents are asked to perform some commands in order to check full understanding.

SCENARIOS 15 MIN

Once the respondent got acquainted with the hand gestures two scenarios were mimicked. First, the respondent is asked to describe the typical situation in which he or she appreciates listening to a music system the most. The setting is very chilled and the respondent is assisted where required.

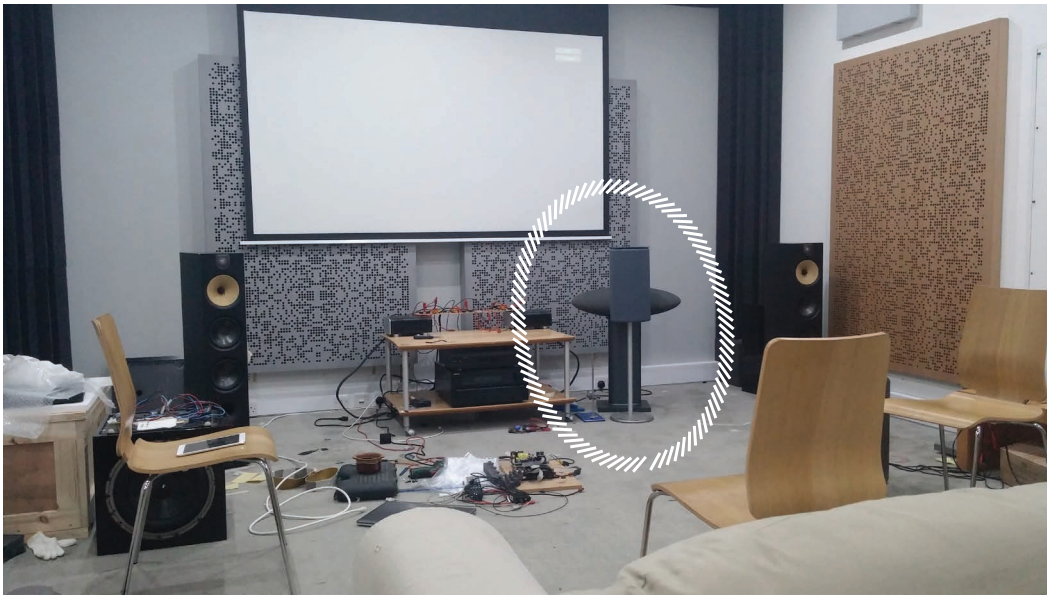


figure 50 Setup for pilot 1 with the mock-up slightly on the right of the centre

Secondly, the respondent will perform a prescribed scenario: “You’ve been using the product for quite a while now. It starts predicting your music choice and habits better and better. Today some friends are coming over and you want to share your favourite songs. Start the system without using the app. The system will play you a bespoke playlist. Try to navigate to your favourite song.”

Note that the set-up also included voice commands.

EVALUATION 15 MIN

Afterwards the respondent is asked the following questions:

- What is your first reaction?
- Did you feel comfortable to navigate through the music?
- What is your typical way to listen to music? How is that different for high- and low- involvement?
- If the system would start by playing your saved tracks randomly, would you feel comfortable to turn it on without an app or remote?
- Did the gestures feel natural?
- What feature would be a standout for you?

figure 51 The mock-up in normal state (white) and 'this song is saved' mode (warm orange colour)

Prototype

The system should in all cases not distract from the intended purpose. This prototype was intended to research user habits and should not arouse discussions about form or sound quality.

In essence, the system consisted of three parts, i.e. a mock-up to represent what it could look like, a Zeppelin Wireless speaker to provide a representative level of audio quality and an iPad to stream the content wirelessly to the loudspeaker (see fig. 50). The mock-up could provide visual feedback through a strip of individually controllable LEDs (see fig. 51).

The input from every respondent was interpreted by the host who would then take the associated actions manually. With a gesture ‘next track’ the host would quickly press the ‘next track’ button on the iPad. If the gesture did not require auditory feedback with e.g. the ‘save song’ gesture the host would press the associated button on the keypad in this case changing the colour of the light to a warm orange. The potentiometer was used to manually aim the beam of light at the user (see fig. 52).

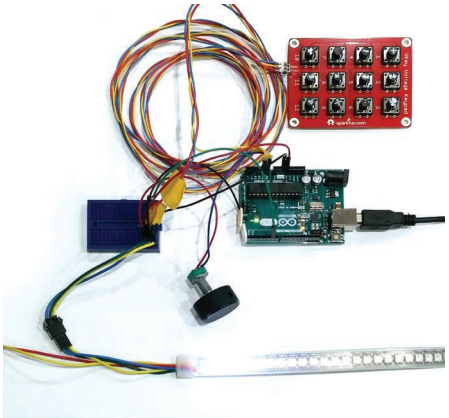
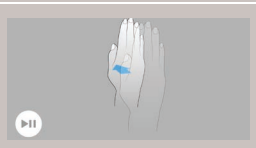
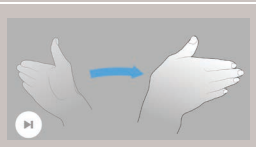
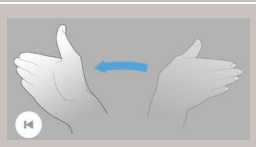
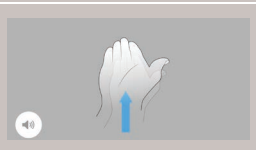
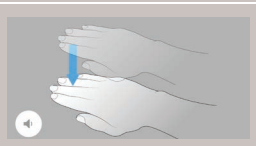


figure 52 The keypad to provide visual feedback through the means of LEDs. An Arduino controlled a circular LED strip.

Interactions

	Command	Input	Auditory feedback	LED feedback
	Attention	Look at the speaker	--	Increase brightness
Primary	Resume last	 Tap hand palm towards speaker	Music starts playing	Circle fills
	Pause	 Tap hand palm towards speaker	Music pauses	Brightness drops low
	Next track	 Move hand fluently to your right	Music skips to next track	A single blink
	Previous track	 Move hand fluently to your left	Music skips to previous track	A single blink
	Fast forward	 Move hand repetitively to your right	Music skips, volume fades low	Light dims during action
	Rewind	 Move hand repetitively to your left	Music skips, volume fades low	Light dims during action
	Volume up	 Move hand up with palm facing up	Volume goes up + 'apple sound'	A single blink
	Volume down	 Move hand down with palm down	Volume goes down + 'apple sound'	A single blink

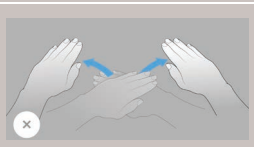
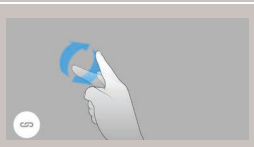
	Command	Input	Auditory feedback	LED feedback
Secondary	Save song	 Grasp and hold for 2 seconds	--	Colour changes to warm orange (state) (see fig. 51)
	Forget song	 Grasp and hold for 2 seconds	--	Colour changes back to white (state)
	Dislike song	 Two hands planar movement	Music skips to next track	Single pulse to cold cyan
	Start track station next	 Circular motions of the index finger	--	Colour swivels
	Play related now	'Start track station next' + 'Next track'	--	--
Tertiary	Voice control	 Raise hand and hold for 2 seconds	Volume fades low + Siri-like tone	Blue light 'following' user
	Undo input	 Wave quickly	Returns to previous state	Quickly pulses for three times
	Play suggested now	n/a	n/a	n/a
	Discovery level [%]	n/a	n/a	n/a

table 8 The different interactions for the first pilot

Results

Did you feel comfortable to navigate through the music?

Yes
Yes
Yes
Yes. Loved the voice control

What is your typical way to listen to music? How is that different for high- and low- involvement?

Start by actively exploring music by looking for artist or label. Then save the tracks I like to my playlist. When I am doing things I put on my playlist on shuffle.
Start off with putting on a specific song from 'recently added' and playing the album. Skip tracks on that album that I do not like. Then look for related albums and so on.
Putting on music CD collection. Never change volume or skip track. Start hoovering instead.
Look for an artist...

If the system would start by playing your saved tracks randomly, would you feel comfortable to turn it on without an app or remote?

Yes
Yes
Yes
Yes. With the time it takes to unlock the phone, open the app, wait for the app to load content: gestures would be much quicker.

Did the gestures feel natural?

Pause Play is a bit dull. Volume height not intuitive. Also prefers identical gestures for up and down, regardless of palm direction. Dislike song and Undo are confusing.
Fast forward / Rewind could be a circular motion. Look at Apple multi-touch gestures. Look into AVRCP controls.
Sound too slow. Liked save and forget
Really liked the voice control

What feature would be a standout for you?

Resume last	2/4	Save/ forget song	2/4
Pause	3/4	Dislike song	0/4
Previous/ next track	4/4	Start track station	0/4
Skip back/ forward	0/4	Voice control	2/4
Volume up/ down	3/4	Undo	0/4

Other feedback

How does the gesture control work with 20 people in the room? Is there an option to stop the track station and switch to random mode again?
It would be cool to have a gesture for telling you the artist and song title. "This sounds good, tell me what it is". Suggest to specify 'song radio' for the song, artist, album, label, genre. All with different gestures. Maybe in 2019 a large share of the people would use smart watches. This would be a great opportunity to extract difficult processes out of the speaker.
--
Invite respondents for a second time to check if the novelty of the gestures is still appreciated.

Appendix D - Pilot 2

Content

Method

The prototype was first tested with five different colleagues, but soon was agreed that the situation was not realistic. Therefore, the test was continued with respondents in their home situation. The respondents were asked to put the prototype in their rooms, get acquainted with the gestures and just play around with it in an informal setting.

Location:	The respondents' living rooms
Apparatus:	Laptop, prototype.
Duration:	30 minutes
Sample size:	n=4

Overall there was a less prescribed procedure compared to pilot 1. Every test consisted of the following steps:

INTRODUCTION

Firstly, the respondents were explained the purpose of the pilot. The respondents were then asked where they would put a device like this in their house. A typical situation in that room was then discussed and used at a later stage in the test.

The set of gestures were explained. The six gestures that were used for this pilot are illustrated in table 9.

SCENARIO

The respondent was now asked to mimic a realistic situation; just like he/she would normally do it. All activities were observed from a distance.

EVALUATION

Afterwards the respondents were asked for their opinion on a number of questions:

- What is your first reaction?
- Do you consider the gesture control to be an added value?
- Are you willing to pay extra for this feature?

figure 53 Prototype with active gesture recognition used for pilot 2

Prototype

The system consisted of a speaker and a depth sensor that are connected via a laptop. The depth sensor constantly measures a 3D infrared point cloud. The laptop processes the data and sends the music stream to the speaker. If the laptop detects a gesture from the sensor data, it acts accordingly. Note: this prototype did not feature visual feedback.

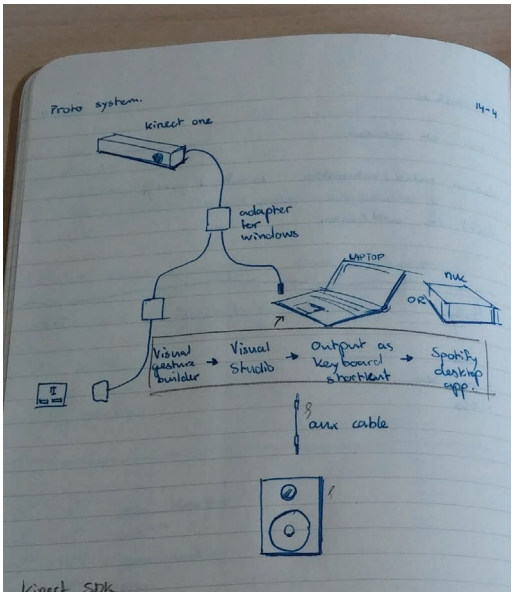


figure 54 System for second pilot

For the depth sensor a Microsoft Kinect One is used because it has a fairly long range on paper and it comes with a well-developed Software Development Kit (SDK) (see Appendix E for the trade-off). The sensor uses three IR emitters to project a point cloud of to the human eye invisible light in the room (see fig. 55). An IR sensor then detects the relative location of every dot. Because every IR emitter projects from a slightly different origin the distance of between the dots is different for every depth.

figure 55 IR point cloud of Kinect v2

The laptop will then process this data. In parallel it will run the Spotify desktop app. If it recognises a gesture it will put the Spotify desktop app on the foreground and simulate a keyboard stroke accordingly. The laptop is wirelessly connected to a speaker through Bluetooth. Due to the provisional solution the gestures were limited by the available Spotify keyboard hotkeys which only included: pause, play, next, previous, volume up and down.

So how can the system recognise a gesture? The Kinect One's SDK comes with a tool called Visual Gesture Builder. Visual Gesture Builder can import recorded clips and asks for each and every frame whether the value is false or true. From this data it can then build a bespoke database through Machine Learning. This database contains data of the movement of every body part for a particular gesture.

The Kinect One's SDK comes with an example code called DiscreteGestureBasics that is capable of comparing the data from the depth sensor with the database and output if a gesture is detected and with what level of confidence. By adding a few lines of code this information could then be translated into the right keyboard stroke to control the Spotify desktop app. The main element of the added code looks as illustrated in fig. 56. The confidence threshold was set to 35%.

The Kinect V2 sensor has a range from 0.8 to 4.0 meter with a 57° horizontal field of view. The gesture database was built from footage from two different people. standing in different positions. For the best performance the database should be constructed from footage of different people wearing different clothes. The prototype was developed in as little as two weeks.

```
if (this.Detected && detectionConfidence > 0.35f)
{
switch (currentGesture.Name)
{
case "Next": SpotifyHelper.controlSpotify(SpotifyHelper.SpotifyAction.Next);
case "Play": SpotifyHelper.controlSpotify(SpotifyHelper.SpotifyAction.PlayPause);
case "Previous": SpotifyHelper.controlSpotify(SpotifyHelper.SpotifyAction.Previous);
case "Voldown": SpotifyHelper.controlSpotify(SpotifyHelper.SpotifyAction.VolumeDown);
case "Volup": SpotifyHelper.controlSpotify(SpotifyHelper.SpotifyAction.VolumeUp);
}
}
```

figure 56 Simplified version of the working element of code to send keyboard strokes to the Spotify app

Interactions

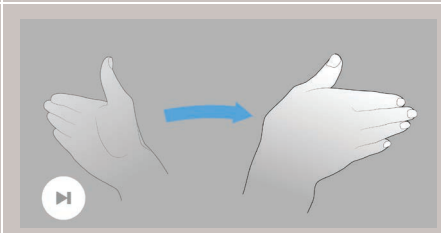
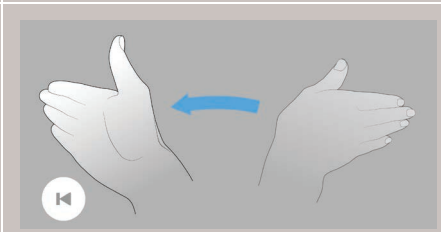
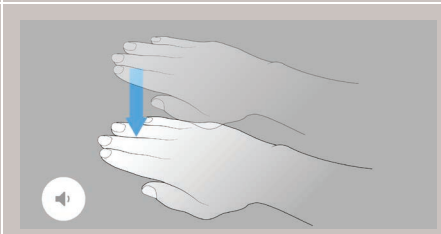
	Command	Input	Auditory feedback
Primary	Resume last	Clap hands twice at shoulder height	Music starts playing
	Pause	Clap hands twice at shoulder height	Music pauses
	Next track	 Move hand fluently to your right	Music skips to next track
	Previous track	 Move hand fluently to your left	Music skips to previous track
	Volume up	 Move hand up with palm facing up	Volume goes up + 'apple sound'
	Volume down	 Move hand down with palm down	Volume goes down + 'apple sound'

table 9 The supported gestures for the second pilot

Results

Placement in room
Miles could think of three positions to place this speaker in. On the bookshelf behind the sofa, on the table on the side of the sofa or in front of his television. For this experiment the speaker was placed on the table on the side of the sofa.
On the shelf near the window. On the left side of the listener.
Daphne would normally place a speaker near her bed because the first thing she likes to do in the morning is switch on the radio. However, when she experienced how well the gestures worked she could as well have placed it above the fireplace.
Anna would put the speaker either in the kitchen or living room. For this pilot the speaker was placed on the window sill near the sitting area. This is not the area where the TV is.

Do you consider the gesture control to be an added value?
Only when the gestures work really accurate. Miles bought a cheap remote control at some point that features gesture recognition. He stopped using those features because it did not always work. Most annoyingly were the false positives where he would for instance put the remote control on the ground and it would mute audio.
Yes, if it works.
As long as the gestures do not make any sound. The clapping gesture would make too much noise and could wake up her housemates in the morning.
As long as the dogs do not interfere with the gestures, yes. It would be an added value to be able to switch the speaker on, change volume etc; similar to the buttons on a music system.

Are you willing to pay extra for this feature?
If there would be two almost identical products of which one features gesture control he would be willing to pay just a little more because of his experiences earlier. The idea could either be a success or failure depending on how accurately it understands your input. Considering B&W's prestigious image it could add a high-tech experience when it works really accurately. When it doesn't, it will risk making the product childish and playful.
Yes
No. Daphne currently listens a lot of radio through her Sonos systems and quite likes to press a button that is physically there. She admits to be impressed by how this technology is possible. She would like to show it off to her friends, but does not get the benefits for herself. Hence, she is not willing to pay extra for it.
Anna admits to not be very skilled with a cell phone. She likes the idea of a stand-alone system that does not require the use of cell phones. She does not carry a phone on her in the house. And what if the cell phone breaks down?

figure 57 Prototype in respondents' living rooms

Practical notes

- The pause/ play clap did not work accurate.
- The system got confused when Miles reached out but did not move his hand.
- When Miles was in a position on the edge of the field of view the system would identify a lot of false positives.
- The gestures worked best if Miles was positioning his upper body perpendicular to the camera.
- The gestures only worked while doing it in the direction of the camera, as opposed to a position that would be most comfortable for you. In the final product it should be clear where the camera is.

- The shiny armchair caused a confusing image.
- The gestures did not work for Adeline in the first place simply because she would move her arm differently and the system would not recognise it. Understandably, this really frustrated her. And it was not possible to execute the experiment properly for this reason.
- In close proximity the height of the gesture became an issue.

- It did not indicate when the system was off. Therefore, kept waving while the system was not 'paying attention'. It is crucial to have a visual feedback that gives this kind of information.
- Tried to make gestures bigger and faster when it did not work.

- The living room was much longer than the sensor's range allowed for. Making it hard to understand when you are in range.
- The view of the sensor was blocked by armchairs and plants.

Recommendations for development

- The product should get 'pre-school' to understand the basics of gestures and learn from and adjust to the user.
- Miles was very much interested in putting on a song through voice commands.

- The gestures should be recognised from the side.

- When you throw a party there should be a way to prevent people from messing around with it.
- Also anticipate on older people because it can make things easier for them. It could appeal to a lot of target groups.

- It would be nice to hang the speaker from the ceiling. It would be the best place to keep track of the users and the least change-prone in terms of acoustics.
- The gestures are stiff, but clear.
- She is worried that the speaker will respond to her movements if she starts dancing.

figure 58 Prototype in respondents' living rooms

Appendix E - Depth sensors

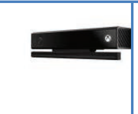
												
	Microsoft Kinect One	**SoftKinetic DS525**	**Google ATAP Soli**	**Leap Motion controller**	**Intel Realsense R200**	**Lytro camera**	**Asus Xtion Pro**	**Creative Senz3D (intel)**	**Creative BlasterX Senz3D**	**Stereolabs Zed**	**Orbbec Astra Pro**	**Structure**
Components	3x IR blasters 1x IR camera 1x RGB camera 2x microphone	3x IR laser diffuser 1x IR camera 1x RGB camera 2x microphone	1x radar	2x cameras 3x IR LEDs	1x IR laser projector 2x IR camera 1x RGB camera	1x light field sensor	1x IR blaster 1x IR camera 1x RGB camera 2x microphone	1x IR laser projector 1x IR camera 1x RGB camera 2x microphone	1x IR laser projector 1x IR camera 1x RGB camera 2x microphone	2x RGB camera	3x IR projector 1x IR camera 1x RGB camera 2x microphone	1x IR projector 2x IR LEDs 1x IR camera
Resolution			N/A									640×480
Range	0.8 – 4.0 m *(Or 0.4 – 3.0 m in near range mode)*	0.15 – 1.0 m	0.0 – 0.2 m	0.025 - 0.6 m	0.5-3.5 m	Unknown	0.8 - 3.5m	0.2 – 1.0 m	Unknown	1.0 – 15.0 m	0.4 – 8.0 m	0.4 – 3.5
Field of view	57° horizontal, 43° vertical	63° horizontal, 49° vertical, 75° diagonal	360°	150°	59° horizontal, 46° vertical, 70° diagonal	Unknown	58° horizontal, 45° vertical, 70° diagonal	74°	Unknown	110° diagonal	60° horizontal, 49.5° vertical, 73° diagonal	58° horizontal, 45° vertical
Size	24.9 x 6.6 x 6.7 cm	10.5 x 3.0 x 2.3 cm	As small as a watch	7.6 x 3.0 x 1.3 cm	102x9.5x3.8mm (integrated version)	Unknown	18 x 3.5 x 5 cm	10.8 x 5.2 x 5.4 cm	Unknown	175 x 30 x 33 mm	160 x 30 x 40 mm	119 x 28 x 29 mm
Release date	November 22, 2013	+/- June 3, 2013	Not yet released	July 22, 2013	January 12, 2016	February 29, 2012	+/- July 18, 2011	+/- June 4, 2013	Q2 2016	19 May 2015	Unknown	Nov 2013
Retail price	$99.99	$164.00	N/A	$ 69.99	$99.00	$399	$169.99	Discontinued?	$199.99	$449	$149.99	$379
Pro's	Large open source community		Recognises the tiniest gestures	High detail, desktop applications	Very compact, also comes as OEM	Adjust focus after capture				Suitable for outdoor use. 4M pixels. Long range.	Persee version has integrated computer.	Onboard Li-Po battery. LEDs enable night vision.
Con's	Bulky		Short range			Enormous data stream				No visibility in the dark		
Link	https://msdn.microsoft.com/en-us/library/hh973074.aspx	http://www.softkinetic.com/Portals/0/Documents/PDF/WEB_20131230_SK_DS525_Datasheet_V2.1.pdf	https://www.google.com/atap/project-soli/	https://www.leapmotion.com/product/desktop	https://software.intel.com/en-us/articles/realsense-r200-camera	https://www.lytro.com/press/releases/lytro-inc-unveils-the-worlds-first-consumer-light-field-camera	https://www.asus.com/3D-Sensor/Xtion_PRO/specifications/	http://uk.creative.com/p/web-cameras/creative-senz3d	http://www.creative.com/corporate/pressroom/?id=13492	https://www.stereolabs.com/	https://orbbec3d.com/product-astra-pro/	http://structure.io/developers

table 10 Available sensors

Appendix F - Inspiration cosmetics

Method

Visit to London cosmetics shops.


OR ROUGE
SERUM


ARMANI
code


CHAMPAGNE
HENRIOT
BLANC DE BLANCS


CHAMPAGNE
HENRIOT
BRUT SOUVERAIN


LANCÔME
ÉNERGIE DE VIE
THE SMOOTHING & PLUMPING
PEARLY LOTION


LANCÔME
DREAMTONE
ULTIMATE DARK SPOT
CORRECTOR
BEAUTIFUL SKIN TONE
CREATOR
3
PIGMENTATION PROFILE
DARK SKIN TONE


SHISEIDO
BENEFIANCE
NutriPerfect
Pro-Fortifying
Softener
Lotion Adoucissante
Pro-Reconstituante

Appendix G - First sketches