

Tiina Männistö-Funk  
De som lyssnar till maskiner.  
Talsyntes och hörande som vetenskaplig praktik

kapitel 9 ur boken  
Ingemar Pettersson & Daniel Normark (red.)  
*Sinnen i arbete*

Arkiv förlag 2025  
Pandoraserien XXXIII

FÖRSLAG PÅ KÄLLANGIVELSE:

Männistö-Funk, Tiina (2025) ”De som lyssnar till maskiner. Talsyntes och hörande som vetenskaplig praktik”, i Ingemar Pettersson & Daniel Normark (red.), *Sinnen i arbete*, s. 225–250, Lund: Arkiv förlag, <https://doi.org/10.13068/9789179243968>.

Det här kapitlet ur en e-bok från Arkiv förlag distribueras fritt över internet genom *open access*. Titeln finns också tillgänglig i tryckt utgåva med ISBN: 978 91 7924 395 1.

Verket är upphovsskyddat enligt en upphovsrättslicens från Creative Commons: Erkännande-Ickekommersiell-IngaBearbetningar, som medger ickekommersiell användning och spridning i oförändrat skick så länge källan anges.

Arkiv förlag · Box 1559 · 221 01 Lund · BESÖK Stora Gråbrödersgatan 17 a  
046-13 39 20 · [arkiv@arkiv.nu](mailto:arkiv@arkiv.nu) · [www.arkiv.nu](http://www.arkiv.nu)

© Författarna/Arkiv förlag 2025  
E-boksutgåva (PDF) 2025

Beständig länk till hela boken: <https://doi.org/10.13068/9789179243968>

ISBN: 978 91 7924 396 8

ISSN: 1404-000X



## 9. De som lyssnar till maskiner. Talsyntes och hörande som vetenskaplig praktik

TIINA MÄNNISTÖ-FUNK

OVE: How are you?

PAT: What did you say before that?

OVE: I love you.

PAT: What did you say before that?

OVE: How are you?

PAT (*viskar*): What did you say before that?

OVE: I love you.

PAT (*sjunger*): What did you say before that?

Det här är en konversation mellan två maskiner, kaskadsyntetiseraren Ove I, en förkortning av "Orator Verbalis Electricus", och parallellsyntetiseraren Pat, "Parametric Artificial Talker".<sup>1</sup> Dessa talsyntesapparater samtalande under ett forskningsmöte vid Massachusetts Institute of Technology (MIT) sommaren 1956.<sup>2</sup> Ove I hade utvecklats av Gunnar Fant vid Kungliga Tekniska högskolan (KTH) i Stockholm och presenterats för första gången 1952. Pat fullbordades 1953 av Walter Lawrence, forskare vid Signals Research and Development Establishment, den brittiska regeringens forskningsenhet för militärkommunikation. Ove och Pat hade utvecklats som oberoende projekt och hade vissa tekniska skillnader, men de vilade på samma teoretiska grund. Båda var så kallade formantsyntetiserare, som tekniskt gick ut på att generera de distinkta ljud vi uppfattar som tal, de så kallade formanterna, på liknande sätt som vi gör med våra kroppar. Båda maskinerna utvecklades under en internationellt aktiv period inom forskning i och modellering av det mänskliga talets fysiska och akustiska särdrag.

---

Forskningen bakom det här kapitlet var en del av *Talking machines*-projektet (2018–2022), som finansierades av Konestiftelsen.

Det är synd att transkriptionen ovan inte låter dig som läsare höra de något uttänjda och besynnerliga, men människolika och lättbegripliga, meningarna som yttrades av maskinerna. I text är vi hänvisade till ordens betydelse. Parenteserna som anger viskningar och sång är så långt detta medium når. Frånvaron av direkt lyssnande leder oss dock till artikelns huvudtema. Hörande och lyssnande spelade en viktig roll för den forskning som skapade formantsyntetiserarna, men uppmärksammades sällan när resultaten publicerades. I fonetikartiklar figurerar lyssnande som anden av en utsagd praktik, som bara kan anas mellan de textuellt och grafiskt uttryckta resultaten. Praktiken är därmed svår att lokalisera. Genom att studera talsyntetiserarnas historia kan man emellertid konstatera att lyssnandet påverkade såväl den dagliga praktiken som det samtidiga bevittnandet av forskningsresultaten. Dialogen mellan Pat och Ove är ett bra exempel på detta: den demonstrerade den fonetiska forskningens starka utveckling under 1950-talet, men bara för dem som till följd av praktiskt vetenskapligt arbete visste hur man lyssnar på talande maskiner.

Efter den första konversationen 1956 fick Ove I och Pat ett nytt tillfälle att träffas: när Walter Lawrence var opponent på Gunnar Fants doktorsavhandling i Stockholm 1958 tog han med sig Pat, och under disputationen förde maskinerna på nytt en dialog.<sup>3</sup> Liknande uppvisningar följde. I augusti 1962 välkomnades deltagarna i ett internationellt seminarium i talkommunikation av Pat och Ove II, som båda yttrade samma mening: ”Welcome to the Stockholm Speech Communication Seminar”, nu väldigt människolikt.<sup>4</sup> Dessa festliga tillfällen understryker den roll som talsyntetiserarna spelade i fonetik och talforskning. Talsyntes kom inte i kommersiellt bruk förrän i slutet av 1970-talet.

Att lyssna till syntetiserat tal var först och främst en inomakademisk syssla. Utifrån sett kan maskindialogerna framstå som lustiga anekdoter. Men för forskare som hade tillägnat sig särskilda färdigheter i att lyssna gjorde de talande maskinerna det möjligt att direkt bevittna en nivå av vetenskaplig förståelse. Att höra några enkla fraser yttras av syntetiserarna innebar inget mindre än att lyssna till vetenskapliga fakta. Sådant bevittnande är en väsentlig del av experimentell kunskap, som Steven Shapin och Simon Schaffer har påpekat.<sup>5</sup>

Lyssnandet till apparater som Pat och Ove blev ett standardverktyg för fonetisk forskning och andra former av talforskning under senare hälften av 1900-talet. För att förstå kopplingen mellan lyssnande och talsyntes under perioden 1950–1980 fokuserar jag särskilt på Ove. Genom

att anlägga ett perspektiv som kombinerar teknikhistoria och vetenskaps-historia belyser jag inte bara talsyntesens historia, utan även lyssnande som vetenskaplig praktik. Jag frågar: Vilken roll spelade lyssnande i Gunnar Fants syntesarbete som ledde fram till Ove? Vilken roll spelade lyssnande i akademisk fonetisk forskning och undervisning där Ove användes vid sidan om Fants arbete? Vilken roll spelade lyssnande i framtagandet och mottagandet av senare kommersiella tillämpningar av talsyntes? Hur och varför fonetiker lyssnade till syntetiskt tal kan hjälpa oss att förstå både instrumentens och sinnenas roll i den vetenskapliga praktiken.

I de allmänna historieverk och handböcker i fonetik som är författade av talforskarna själva beskrivs ofta talsyntesmaskinerna som inslag i utvecklingen av fonetik.<sup>6</sup> Till och med de äldsta mekaniska formerna av talsyntes, som Wolfgang von Kempelens talmaskin från slutet av 1700-talet, behandlas som en del av fonetik- och talforskningens genealogi och självförståelse.<sup>7</sup> Historiska arbeten om ljudteknologi, som främst är inriktade på den sociotekniska historien om reproduktion och överföring av tal och musik, framställer däremot talsyntes som excentriska detaljer i uppfinnarnas strävan efter ljud som gick att spela in och transmitta.<sup>8</sup> Senare talteknologier under 1900-talet, såsom voderen och vokoderen, studeras ofta inom ramarna för sin användning i populärkultur som film och musik.<sup>9</sup>

Ove-enheterna har dock inte studerats i så stor omfattning från dessa perspektiv eller överhuvudtaget från ett historiskt perspektiv, trots att maskinerna uppmärksammas i åtskilliga historiska översikter. Inte heller har lyssnandets roll som nyckelpraktik inom talsyntes studerats. Därför är det här den första historiska granskningen av en sensorisk vetenskaplig praktik kopplad till Ove, lyssnandet. Det viktigaste källmaterialet består av forskningsartiklar och andra texter författade av forskare som arbetade med Ove-maskinerna. Jag använder mig även av fotografiska källor, artefakter och intervjuer för att rekonstruera talsyntesteknologins användning och dess relation till lyssnande.

## Jakten på det naturtrogna ljudet

Den mekaniska talsyntesens historia sträcker sig över många århundraden, men ofta anses voderen vara startpunkten för elektriska maskiner som kunde generera naturligt tal.<sup>10</sup> Den utvecklades under 1930-talet på Bell Telephone Laboratories (Bell Labs) och presenterades för allmänheten på världsutställningen i New York 1939. Det var ingen liten bedrift:

förberedelserna innefattade att 24 telefonister övade i ett år på att manövrera voderens tangenter, pedaler och kontroller.<sup>11</sup> Enligt medie- och ljudforskaren Mara Mills var voderen i själva verket en ”dekorativ avkomling” av vokodern, som Homer Dudley hade börjat utveckla på Bell Labs i slutet av 1920-talet och färdigställt 1935. Nuförtiden är vokodern främst känd för sin användning inom musik och film, där den ger mänskligt tal en robotliknande ton, men den utvecklades ursprungligen för att möjliggöra effektivare röstöverföring genom att packa, koda och avkoda ljudet. Dudleys mål var att minska antalet frekvenser som krävdes vid talöverföring via till exempel telefonlinjer.<sup>12</sup> Vokoderns kodnings- och avkodningsförmåga användes under andra världskriget för att kryptera telefonsamtal mellan Winston Churchill och Franklin D. Roosevelt samt andra statliga och militära företrädare.<sup>13</sup>

Intresset för talmodellering och syntesteknologi drevs åtminstone delvis av militära intressen.<sup>14</sup> Under första världskriget arbetade Bell Labs föregångare, AT&T:s utvecklingsavdelning, med flera ljudteknologier för militär användning, vilket stakade ut vägen för att ljud började uppfattas som signaler samt för olika möjligheter att generera och reproducera ljud i elektriska nätverk.<sup>15</sup> Walter Lawrence utvecklade sin Pat-syntetiserare medan han arbetade på en statlig militärforskningsenhet i Storbritannien. Intresset för talkommunikation var även kopplat till kalla krigets cybernetikforskning, som föddes ur utvecklingen av militärteknologi under andra världskriget och strävade efter att göra gränssnitten mellan människa och maskin sömlösa.<sup>16</sup> I detta sammanhang betraktades talkommunikation som en del av ett allmänt kommunikationssystem. I ett av de grundläggande verken inom cybernetik, *The human use of human beings*, skrev författaren Norbert Wiener 1950: ”I viss mening slutar alla kommunikationssystem i maskiner men de vanliga språkliga kommunikationssystemen slutar i en alldeles speciell maskin som kallas människa.”<sup>17</sup> Samma år deltog Wiener i en konferens om talkommunikation vid MIT anordnad av Acoustical Society of America. Evenemanget blev en milstolpe i den nya tvärvetenskapliga talforskning som förenade ingenjörskonst, akustik, matematik, fysik, fonetik, lingvistik, psykologi och neurologi.<sup>18</sup> Vid konferensen föreläste Wiener om möjliga lösningar för telefonkommunikation mellan människor och maskiner vid exempelvis kraftverk.<sup>19</sup> För att synkronisera kommunikationen mellan människor och maskiner krävdes bättre teknisk kunskap om mänskligt tal.

På en praktisk nivå hade forskningen om tal och den tekniska utvecklingen varit kopplade till telefoni, som var den huvudsakliga tekniken

för att överföra mänskligt tal. Kopplingen mellan talsyntes och elektriska kommunikationssätt hade funnits redan under senare halvan av 1800-talet, inklusive idéer om syntetiserare som skulle transformera telegrafmeddelanden till tal samt Alexander Graham Bells syntesexperiment innan han patenterade telefonen.<sup>20</sup> Som fallet med voderen exemplifierar motiverades de ansträngningar som sedan 1930-talet gjorts för att förstå talproduktion och reception av det tekniska målet att klippa av och förpacka tal i enlighet med telefonens behov.<sup>21</sup> Praktiska tekniska frågor gjorde talsyntes till ett intressant tema för dåtidens största forskningsinstitutioner, såsom Bell Labs, MIT och militära forskningsenheter. Men Dudley framhävde också syntesens viktiga roll för talforskning genom att publicera en artikel om Kempelens talmaskin och andra historiska försök att reproducera tal.<sup>22</sup>

Tre decennier senare definierade Dennis Klatt, pionjär inom digital text-till-tal-syntes, voderen som början på talsyntesens utveckling, en utveckling som ledde till datorlösningar för text-till-tal, däribland hans egen. I detta narrativ följde Lawrences Pat och Fants Ove hack i häl på voderen och Haskins Laboratories Pattern Playback-maskin (den första kända elektroniska syntesmaskinen för vetenskapliga syften). Tillsammans utgjorde de talsyntesens tidiga milstolpar.<sup>23</sup> Det stämmer på en nivå, eftersom alla dessa syntetiserare byggde på regler om tal som talforskningen hade fastslagit. Men Ove och Pat var något nytt i jämförelse med Pattern Playpack, eftersom de inte bara härmade tal utan också återskapade den fysiologiska talprocessen. Och till skillnad från Klatts system var de icke-digitala och krävde, i likhet med voderen, specialister för att producera tal.

Gunnar Fant var på rätt plats vid rätt tid för att inta en central roll i det framväxande fältet för talsyntes. Han hade tagit civilingenjörsexamen i elektroteknik vid KTH 1945, med specialisering inom telegrafi och telefoni. Hans examensarbete handlade om den mänskliga röstens akustiska egenskaper, i synnerhet i vokalljud. Efter studierna fortsatte han att forska inom ämnet vid Telefonaktiebolaget L.M. Ericssons akustiklaboratorium.<sup>24</sup> Företaget ville reservera en del av telefonlinjernas bandbredd för signaleringssyften och ville därför veta hur inskränknigen av frekvensen skulle påverka hur tal uppfattades. På L.M. arbetade Fant 1946–1949 med att identifiera frekvenser i svenska talljud. På basis av insamlade audiogramdata definierade han frekvens- och intensitetsområdet för tal, som han publicerade 1954.<sup>25</sup> Den region av hörseln som är viktig för talförståelse har i audiogram en bananliknande form, och ”tallbanan” har blivit ett grundbegrepp inom talforskningen.<sup>26</sup>

Innan Fant publicerade dessa resultat tillbringade han några viktiga år i USA. Han hade blivit inbjuden till MIT av Leo Beranek, som hade besökt L.M. Ericsson 1948, precis efter sin utnämning till professor i kommunikationsteknik vid nämnda lärosäte.<sup>27</sup> Fant vistades vid MIT-laboratoriet för akustik från 1949 till 1951, precis när den nya tvärvetenskapliga talforskningen började ta form. Därmed fick han uppleva forskningsområdets framväxt ”som en förgrening av lingvistik, elektrisk kretsteori, psykoakustik och informationsteori”, som han senare beskrev det.<sup>28</sup> Fant började utveckla syntesmodeller med Ken Stevens, som senare skulle leda Speech Communication Group vid MIT,<sup>29</sup> och lärde känna bland andra James Flanagan på Bell Labs, en framstående gestalt inom talforskning som hade patent på ett modernt artificiellt struphuvud. Fant besökte även Haskins Laboratories för att testa det nya Pattern Playback-systemet. Med hjälp av en fotocell kunde maskinen skapa en ljudvåg utifrån en spektrografisk representation av tal ritad på folie.<sup>30</sup> Fant skrev senare att hans första möte med en talsyntesmaskin var en fascinerande upplevelse: ”Med min internaliserade kunskap om spektral representation i svenskan kunde jag måla en mening och få den uppspelad för mig. Det var monotont, men begripligt.”<sup>31</sup> Det var första gången han var med om förvandlingen av teoretisk kunskap om tal till syntetiskt tal som han kunde lyssna på, något som skulle bli centralt för hans karriär.

Väl tillbaka i Stockholm grundade Fant år 1951 Taltransmissionslaboratoriet vid KTH, och 1952 slutförde han sitt arbete med en talsyntesmaskin. Apparaten kunde skapa vokalljud och vissa konsonanter som v och l. När Fant presenterade maskinen 1953 intervjuades han av Sveriges Radio och blev tillfrågad om maskinen hade ett namn. Eftersom den kunde uttala ”Ove” valde Fant det som namn, och efteråt tolkade han det som en akronym för ”Orator Verbalis Electricus”.<sup>32</sup> Nästa version, Ove II, stod klar 1962. Den kunde generera obegränsat, sammanhängande tal och programmerades med linjer som ritades i konduktivt bläck på ett ark av plast. Med Ove II kunde Fant och medarbetare i hans laboratorium generera naturtroget tal, något som imponerade på det internationella forskarsamfundet när de samlades på seminariet i tal-kommunikation 1962.<sup>33</sup>

Transmissionslaboratoriet blev en viktig samlingspunkt för nya forskare inom fältet. Johan Liljencrants anslöt sig till laboratoriet som magisterstudent precis efter slutförandet av Ove II.<sup>34</sup> Han blev en banbrytare i synteshårdvarans fortsatta utveckling och gick även i bräschen för datorisering genom att införa datorer på laboratoriet 1967.<sup>35</sup> Liljencrants



utvecklade Ove III som ett parallellsystem till Ove II. Den väsentliga skillnaden mot föregångarna var att Ove III kontrollerades med ett datorprogram. Dessutom tillverkade och sålde Liljencrants Ove III samt en uppdaterad version av Ove I kommersiellt genom sitt företag Fonema.<sup>36</sup>

Både talsyntesens och Fants egen historia belyser en viktig poäng om Ove och andra besläktade talmaskiner: syftet var inte att skapa ljud som lät så mänskligt som möjligt, utan snarare att skapa maskinellt tal på ett så människoliknande sätt som möjligt för att förstå den mänskliga talprocessens teori och praktik bättre.

Under 1960-talet krävde generering av syntetiskt tal fortfarande massor av hårt arbete och omfattande teknisk expertis. Enligt Liljencrants hågkomster senare under 1970-talet kunde Ove II ”producera syntetiskt tal av häpnadsvärd god kvalitet”, men ”man måste sätta in en månads arbete för att skapa två sekunder av tal som lät bra”.<sup>37</sup> Med andra ord var de korta välkomstfraserna och den absurda dialogen mellan Ove och Pat som en akademisk publik fick höra i mitten av 1950-talet och början av 1960-talet resultatet av hisnande teknologiska insatser. Beredvilligheten att jobba så intensivt för några sekunders tal är ett tecken på den vetenskapliga betydelsen av talsyntes som man kunde lyssna på och som lät bra. Den var ett hörbart bevis för teoretisk kunskap och utveckling. Pat och Ove demonstrerade den mest avancerade dåtida kunskapen om tal, och de gjorde denna kunskap direkt tillgänglig för alla som visste hur de skulle lyssna på talsyntes.

## Hörbart kunskapsläge

Vår vardagliga användning av språket gör att vi ofta missar de skillnader som finns mellan talat och skrivet språk. För talforskare är tal mycket mer komplicerat och nyanserat än vad skrivna bokstäver och ord kan förmedla. En stor del av deras arbete på 1900-talet bestod i att skapa system som kunde fånga talets variationer och mönster på begripliga sätt. ”De fysiologiska rörelser som är inblandade i talet kan presenteras för ögat eller örat”, skrev Homer Dudley 1950.<sup>38</sup> Det örat uppfattade var talsyntes, medan ögat kunde avkoda fonetiska alfabet. Dudley hänvisade även till boken *Visible speech* från 1947 för att visa på ljudspektrografens möjligheter. Spektrografen hade utvecklats som ett krigsprojekt på Bell Labs och blivit kommersiellt tillgänglig under varumärket Sonograph.<sup>39</sup> Med hjälp av maskinen kunde talet inte bara återges i form av alfabet, fysiologiska illustrationer eller röntgenbilder av talröret (eller ansatsröret, det vill säga

kaviteterna mellan struphuvudet och läpparna, som genom sin form och funktion skapar olika språkljud), utan även i utskrivna mönster, diagram som visade talets frekvens och intensitet.<sup>40</sup> Fant själv fick höra om den nya tekniken medan han arbetade på L.M. Ericsson, där han manuellt skapade spektrogram över talets intensitet och frekvens genom att kombinera oscillogram från en våganalysator som mätte intensitet med data från frekvensband.<sup>41</sup> I början av 1960-talet ansåg Fant att ljudspektrografien var det viktigaste verktyget inom talforskning, viktigare än syntes eller digitala datorer, och senare beskrev han den som den viktigaste utvecklingen i talforskningens historia.<sup>42</sup>

I och med att talsyntetisering började bli ett etablerat verktyg för fonetisk forskning förenades två sensoriska teknologier som skapade representationer av ljud för ögat respektive örat. När det hörbara talet förvandlades till visuella spektrogram kunde de användas för att identifiera formanternas frekvenstoppar. Utifrån dessa kunde en formantsyntetiserare generera en hörbar version som forskarna kunde lyssna på och bevittna.

Man kan säga att Ove först och främst var en modell av tal. Dessutom var den en talande modell av tal; i linje med Dudleys tanke ovan presenterade Ove talapparaten fysiologiska funktioner ”för örat”. Medan den talsyntesen som vi nuförtiden möter oftast baseras på stora mängder förinspelat mänskligt tal (vilket kallas konkatenativ syntes eller konkatenationssyntes) är formantsyntes, som i Ove-syntetiserarna, fullkomligt maskinbaserad. Dessutom utgår den från det fonetiska begreppet ”formanter” som distinkta komponenter hos talet. Formanterna låg i talforskningens kärna i mitten av 1900-talet, när forskningen försökte komma underfund med hur olika talljud genererades och uppfattades, varifrån de uppfattade skillnaderna i talljud härstammade och vilka komponenter som var nödvändiga för att generera begripligt tal. Det var här som Gunnar Fant kom med sitt viktigaste akademiska bidrag: den så kallade källa-filter-teorin. Han publicerade teorin i boken *Acoustic theory of speech production*, som utkom 1960 och blev en milstolpe inom talforskningen.<sup>43</sup>

Formanter är de distinkta komponenterna i talet, som frambringas av resonans i talapparaten. I sin bok definierar Fant formanter som ”spektrala toppar i ljudets spektrum”.<sup>44</sup> Fant hade påbörjat sin analys av formanter redan i samband med arbetet på L.M. Ericsson för att identifiera formantfrekvenserna i svenska talljud. I USA hade han gjort röntgenundersökningar av en rysk invandrare då hen uttalade olika ryska ljud,

vilket sedan blev till datamaterial för Fants bok.<sup>45</sup> Redan långt tidigare hade forskare identifierat formanter som väsentliga komponenter i talet, och deras kopplingar till fysiologiska fenomen i den mänskliga talapparaten hade studerats i årtionden, men trots det fanns ingen omfattande teori.<sup>46</sup> Fants bok var således banbrytande i det att den tillhandahöll en modell över formantfrekvensernas koppling till talapparatus parametrar.<sup>47</sup> Källa-filter-teorin förklarade fonem som en kombination av ljudkällan (stämbanden) och det akustiska filtrets form (talröret). Formantsyntesen följde denna modell genom att simulera talets källor och filter för att generera syntetiskt tal.<sup>48</sup> Genom att tekniskt återskapa stämbandets och talrörets effekter och få fram ljud som lyssnare kunde känna igen som tal bevisade den källa-filter-teorin.

I motsats till ljudspektrografen, som återskapade ljudet i form av diagram på papper, var syntesapparaterna inte enbart ämnade för att reproducera tal. Som modeller av den mänskliga talapparaten var de konkreta och hörbara forskningsresultat, på ett sätt som spektrografen inte var. Det var tydligt att Fant uppfattade den teoretiska insatsen och det konkreta syntesarbetet som två sidor av samma mynt. Det återspeglades i de tekniska lösningar som han tillämpade i sina syntesapparater. Oves konstruktion gjorde det möjligt att individuellt kontrollera frekvensen för flera elektriska resonatorer samt källans frekvens och amplitud, vilket direkt motsvarade källa-filter-teorin. Walter Lawrences Pat följde liknande principer, men medan Oves resonanskretsar var kaskadkopplade var Pats parallellkopplade. Det förra var mer likt den mänskliga talapparatus sätt att fungera och kunde skapa ljud som lät "naturligare"; det senare erbjöd bättre kontroll över resultatet.<sup>49</sup>

Man började hänvisa till båda maskinerna som "formantsyntetiserare" eftersom de byggde på idén att kontrollera skapandet av formanter. De var exempel på regelsyntes – de följde explicita skriftliga regler för talproduktion – men till skillnad från Pattern Playback-syntetiseraren, den första regelbaserade syntesapparaten som kom i bruk, fungerade de även som tekniska modeller av talprocessen.<sup>50</sup> Ove var en materiell del av en teori, vilket gjorde den särskilt användbar för experiment.

I slutet av 1970-talet formulerade Johan Liljencrants den fråga som ständigt ställts under utvecklingen av talsyntes: "Är det en bra modell av verkligheten, eller är det bara en bra modell?"<sup>51</sup> Formuleringen hänvisar till uppfattningen att Ove-syntetiserarna inte bara modellerade en teori utan var ett vetenskapligt faktum i materiell form. Steven Shapin har visat att den experimentella vetenskapen kom att betrakta fakta som

oförvrängda återspeglings av verkligheten.<sup>52</sup> Enligt honom var publiken väsentlig för implementeringen av experimentell kunskap genom kollektiva presentationer och bevitnande av experiment.

På konferensen 4th International Congress of Phonetic Sciences vid Helsingfors universitet 1961 citerade Haskins Laboratories forskningsdirektör Franklin S. Cooper Fants yttrande om datamängder som ett av de största problemen för talanalys, och konstaterade att forskare kunde ta hjälp av syntetiserare för att koncentrera sig på det väsentliga i datamaterialet – ett sätt att förenkla ”Naturen” och göra den begripligare.<sup>53</sup> Jag påminns om det Bruno Latour och Steve Woolgar skriver om vetenskapligt arbete i sin studie *Laboratory life*: en av de mest centrala vetenskapliga färdigheterna är förmågan att avgöra vilka data man ska behålla och vilka man ska gallra bort. Forskare strävar efter att skapa ordning i kaos genom att sortera fakta från brus. Att formulera ett tydligt argument på grundval av en massa observationer samt formulera en robust och övertygande beskrivning är så krävande att det kan vara en väsentlig bedrift när man lyckas med det.<sup>54</sup> Det speciella med Ove var att den inte bara var en sorterings- och urvals metod i jakten på fakta. Den var även ett påstående i sig, ett vetenskapligt faktum som man kunde höra.

Enligt den nutida talforskaren Brad Story förkroppsligar syntetiserare etablerade fakta om mänskligt tal. I vilken grad vi kan producera övertygande artificiellt tal är ett mått på vår kunskap om hur mänskligt tal produceras.<sup>55</sup> Nästan precis samma definition använde Rolf Carlson och Björn Granström flera årtionden tidigare. Båda forskade om talsyntes som doktorander på Fants laboratorium i början och mitten av 1970-talet, och båda blev professorer i talforskning vid KTH. I en festskrift till Gunnar Fants 60-årsdag 1979 skrev Carlson och Granström om talsyntes: ”Denna metod avslöjar explicit det nuvarande kunskapsläget om talproduktion och taluppfattning.”<sup>56</sup> Just därför var det så viktigt att andra experter i talforskning tillsammans skulle höra Oves tal som cirkulation av teorin som den var en del av.

Ove-maskinerna framträder som kraftfulla argument och bevis för Fants källa-filter-teori. De var inte bara en väsentlig del och ett förkroppsligande av hans teoretiska och analytiska arbete utan även något som kunde höras, som hade en egen röst. De var påtagliga och förnimbara uttryck för vetenskaplig kunskap, för att Fants teori var en ”bra modell av verkligheten”. En viktig del i att skapa representationer av naturen var vilka som bevitnade detta. Steven Shapin betonar att publiken spelar en viktig roll för den experimentella kunskapen: de som bevitnar kollektiva

presentationer och experiment.<sup>57</sup> Det var avgörande att forskarkollegor hörde Oves tal, och det måste ha varit triumfartat och betydelsefullt att låta Ove välkomna seminariegästerna och att spela upp exempel på talsyntes som publiken inte kunde skilja från mänskligt tal, så som Fant gjorde på Acoustical Society of Americas möte i Philadelphia 1961.<sup>58</sup> De syntetiserade rösterna förespråkade förståelsen av talet som hade skapat dem.<sup>59</sup>

Harry Collins har vidareutvecklat Shapins idéer om offentliga experiment genom att skilja mellan experiment och demonstrationer. Medan experiment utförs för att ta reda på något i den naturliga världen, som är oberäknelig och ofta rörig, är demonstrationer inövade presentationer där man visar upp resultaten av sitt vetenskapliga arbete för publik. Forskarens tekniska virtuositet kan göra det svårt för publiken att ifrågasätta de fakta som läggs fram.<sup>60</sup> Före text-till-talsystem hade framställningen av naturligt klingande syntetiserat tal synnerligen starka inslag av teknisk virtuositet. Johan Liljencrants har skämtsamt kallat sig "trollkarlens lärling" i förhållande till Fant.<sup>61</sup> Tillsammans skapade de demonstrationer som inte bara skulle höras och ses, utan även avlyssnas av en akademisk publik. Publikens färdigheter i den vetenskapliga praktiken att lyssna till syntetiserare gjorde det även möjligt att uppfatta den tekniska virtuositeten som en del av föreställningen. De tre tillfällen då Ove och Pat pratade tillsammans kan förstås som kraftfulla demonstrationer.

På konferensen i Helsingfors 1961 sammanfattade Franklin S. Cooper användningen av talsyntetiserare: I kommunikationsbranschen användes de som "sofistikerade telefoner". Inom talforskningen användes de dels som "mångsidiga informanter" som lät forskare manipulera spektret som de ville och höra resultatet omedelbart, dels som artikulatoriska modeller.<sup>62</sup> De användes alltså både i experiment och i demonstrationer. Ove förknippades med imponerande demonstrationer, men figurerade i ännu större utsträckning i experimentell vetenskap och tyst vetenskaplig praktik. Den sistnämnda förmedlades genom vardagliga demonstrationer som en del av utbildningen i fonetik. Därmed blir andra aktörer intressanta utöver teorins talesman Gunnar Fant: de som kan ses som både användare av och publikum för den teoretisk-praktiska apparaten.

## Lyssnande som verktyg för forskning och undervisning

Talsyntetiserare som Ove var måttstocken för kunskapsläget beträffande mänskligt tal och representanter för teorin om hur tal skapas. De användes rutinmässigt för att producera tal i avsikt att studera tal.<sup>63</sup> Dessutom var de praktiska pedagogiska verktyg i undervisningen i fonetik och tal-kommunikation. Collins nämner inte sådana pedagogiska demonstrationer när han skriver om distinktionen mellan demonstrationer och experiment, men den sortens demonstrationer går att knyta till hans begrepp ”interaktionell expertis”: en tredje kunskapsform mellan formella fakta och informell kroppslig kunskap. Interaktionell expertis förmedlar kunskap om det tysta kroppsliga till aktörer som ska lära sig att praktisera vetenskap.<sup>64</sup>

För att granska användandet av Ove-syntesapparaterna som interaktionella verktyg reser vi nu över Östersjön. Fonetiklaboratoriet vid Åbo universitet i Finland var ett av de institut som köpte en Ove-maskin för eget bruk. Kalevi Wiik, assistent på laboratoriet, hade besökt Tal-transmissionslaboratoriet vid KTH 1961 och då använt Ove II för att skapa 158 olika syntetiska ljud som skulle ingå i identifikationstest som han genomförde för sin doktorsavhandling om finska och engelska vokaler.<sup>65</sup> Så fort Wiik blev utnämnd till professor i fonetik vid Åbo universitet 1968 började han utrusta laboratoriet med egna syntetiserare. Innan årtiondets slut hade laboratoriet hunnit skaffa Bell Labs artificiella struphuvud, Lawrences Pat-syntesapparat samt Ove Ib, en uppdaterad version av Ove I som tillverkades av Liljencrants företag Fonema (se bild 9.1).<sup>66</sup> Före dessa inköp hade laboratoriet endast ägt en ljudspektrograf som studenter använde för att analysera formanter och frekvenser.

Den nya utrustningen sköttes av den nyanställda deltidsteknikern Jarkko Metsätähti, som även fick i uppdrag att tillverka ny utrustning. Han började med att tillverka talrörsimulatorn Eeva, som än i dag finns på universitetet. Den hade knappar på frampanelen för att manipulera funktionerna i talapparatsens delar, från struphuvudet till läpparna och näsan (se bild 9.2). Senare byggde Metsätähti en större version, ”stora Eeva”, med en mer avancerad kontrollmekanism för olika variabler. Ända fram till slutet av 1970-talet använde magisterstudenter stora Eeva i arbetet med sina avhandlingar (se bild 9.3 på nästa uppslag).<sup>67</sup>

Sommaren 1970 gjorde laboratoriet sitt största förvärv: en Ove IIIb från Fonema. Metsätähti hade fått lära sig att manövrera maskinen, och för att dra nytta av den nya apparaten erhöll laboratoriet en licens för

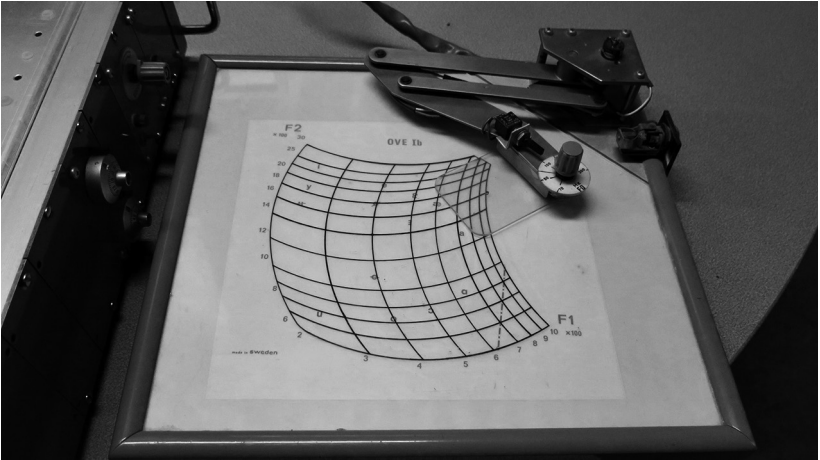


Bild 9.1. Ove Ib på fonetiklaboratoriet i Åbo.  
Foto: Tiina Männistö-Funk.



Bild 9.2. Syntetiseraren Eeva, tillverkad i slutet av 1960-talet på fonetiklaboratoriet i Åbo.  
Foto: Tiina Männistö-Funk.

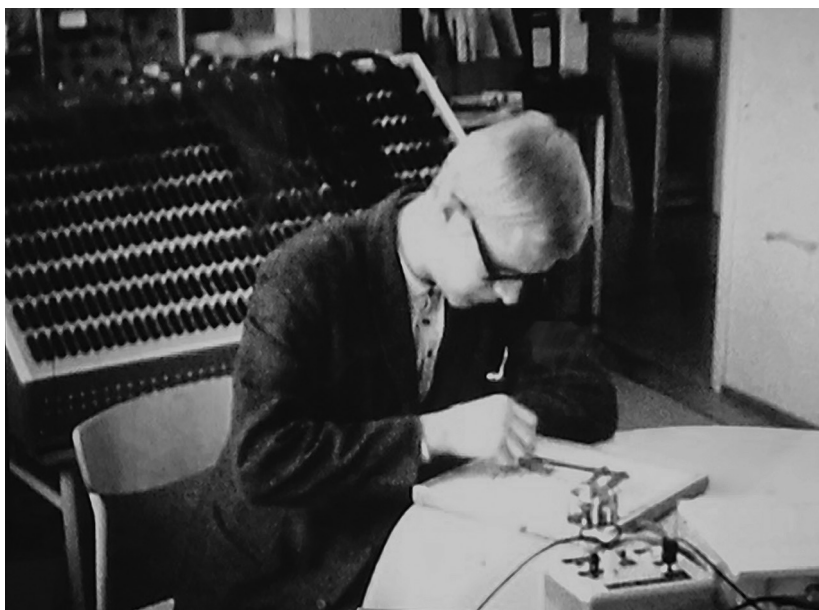


Bild 9.3. Forskaren Pekka Hirvonen använder Ove Ib i slutet av 1960-talet. I bakgrunden syns den större Eeva-syntetiseraren, som sedermera kasserades.

Källa: Stillbild från en super8-film av Jarkko Metsätähti.

att använda minidatorn på universitetets metodologiavdelning. Talsyntes blev nu till ett grundelement vid fonetisk forskning och utbildning vid Åbo universitet. Inom kort skaffade laboratoriet en egen minidator och ett centralminne för att kunna använda Ove IIIb. Personalen kom igång med programmering genom att skapa spektrogram av talsekvenser – den första sekvensen bestod av orden ”hei vain” (”hejsan”) yttrade av lektor Ilkka Raimo. Sedan nedtecknades formanternas numeriska representationer, numren överfördes till hållkort och kortet matades in i minidatorns översättningsprogram. Slutligen blev det dags att programmera och finslipa detaljer. Med denna metod fick de Ove IIIb att yttra meningen ”Saaplari pojat, tää on synteettistä puhetta!” (”Nämen sablar killar, det här är syntetiskt tal!”), som härmade professor Wiiks karakteristiska uttrycks-sätt. I enlighet med traditionen att iscensätta triumfatoriska demonstrationer blev yttrandet höjdpunkten på avdelningens julfest 1970.<sup>68</sup>

Det första text-till-tal-programmet i Finland utvecklades 1973 av forskarna i Åbo med hjälp av laboratoriets minidator och Ove IIIb. Programmet gjorde det möjligt att syntetisera vilka ord och meningar som helst



genom att mata in dem i datorn, och det förevisades på den årliga finska fonetikkonferensen 1974. Det användes och modifierades vidare under cirka arton år, ända tills programmet förstördes i en olycka som raderade datorns minne i början av 1990-talet.<sup>69</sup> Utvecklingen av liknande regelsyntesprogram stod nu till förfogande för talforskare som en kombination av formantsyntesteknologi och ny teknologi i digital databehandling. Carlson och Granström utvecklade ett svenskt text-till-tal-program i Fants laboratorium 1974.<sup>70</sup>

Enligt forskarna i Åbo var målet med text-till-tal-programmet att göra talsyntes så väl lämpat som möjligt för undersökningar om talperception.<sup>71</sup> Intressant nog nämnde de lektor Raimos "absoluta gehör" som ett viktigt element i finslipandet av syntesprogrammet under årens lopp.<sup>72</sup> Det är ett av ytterst få tillfällen i hela mitt litteraturmaterial om talsyntetiserarnas historia då lyssnandepraktiker eller färdigheter i att höra kommenteras direkt, trots att de uppenbart utgjorde en viktig del av arbetet.

Syntetiserarna förändrade fonetikforskningen på ett sätt som knappt syns i publikationerna, men som kan spåras bland instrument och vetenskapliga praktiker. Det var en ny metodologisk praxis där forskarna fick maskinerna att tala och sedan lyssnade till dem.

Man kan urskilja olika aspekter av lyssnande i det vardagliga användandet av syntetiserare i undervisning och forskning. För det första övade studenterna på den fonetiska praktiken att lyssna genom att använda apparaterna. När de väl behärskade apparaterna kunde de producera syntetiska ljud att användas i experiment där testpersoner fick lyssna till ljud med ytterst små skillnader i frekvens och identifiera vad de hörde. På så sätt fick forskarna en uppfattning om tröskelvärden för de ljud som uppfattas som tal. Den högsta nivån av vetenskapligt lyssnande innefattade färdigheter, även virtuositet, som gjorde det möjligt att utveckla och finjustera hård- och mjukvara för syntetisering samt bevittna lyckad talsyntes. Professionellt lyssnande lärdes ut, praktiserades och förevisades med hjälp av Ove. Att förstå vad talforskning var, var att lyssna till Ove.

Tack vare sitt användargränssnitt hade Ove Ib ett särskilt pedagogiskt värde: för att producera ljud förde man ett stift över en yta med markeringar för de två lägsta resonansfrekvenserna, som kunde producera de flesta vokalljuden (se bild 9.1).<sup>73</sup> Studenten kunde se frekvensförändringar och omedelbart höra hur de påverkade ljudet. I Åbo användes maskinerna för talsyntes ofta i undervisningen för att demonstrera olika fonetiska fenomen. På en obligatorisk kurs i talsyntes lärde sig studen-

terna använda syntes vid uppställningen av ett fonetiskt experiment. I magisteravhandlingarna var syntes den mest framträdande metoden. Flertalet jämförde vokaler i olika språk och syntesapparaterna användes för att skapa olika vokalljud som försökspersoner fick lyssna på (bild 9.4 visar resultatet av ett sådant experiment).<sup>74</sup> Maskinerna användes också för studier på högre nivå: Ove IIIb användes till exempel för att producera experimentella stimuli för Olli Aaltonens doktorsavhandling om vokalperception.<sup>75</sup>

I sitt tal om talsyntes på fonetikkonferensen i Helsingfors 1961 beskrev Franklin Cooper en vision om en ”allmänt tillgänglig syntetiserare som möjligtvis kunde fungera som kompletterande verktyg för ljudspektrografen”.<sup>76</sup> I början av 1960-talet fanns det bara några få

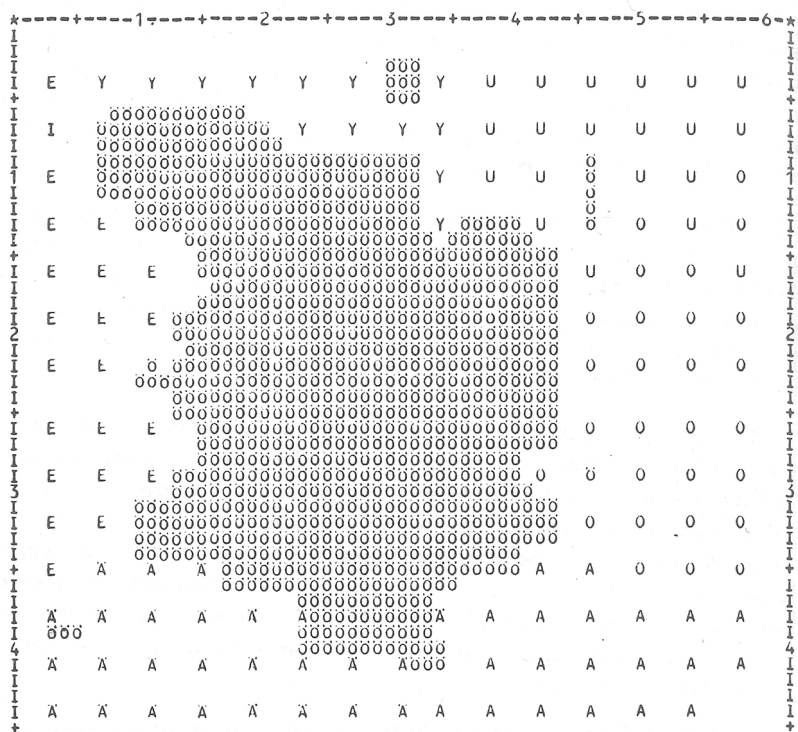


Bild 9.4. Det fonematiska området för vokalen ö i finskan på basis av avlyssningsexperiment med akustiska stimuli genererade av Ove III.

Källa: Publicerad i Olli Aaltonen & Markus Mattila, ”Puhesynteesi fonetiikan opeutuksessa ja tutkimuksessa Turussa”, i *Fonetiikan päivät – Oulu 1979: IX Fonetiikan päivillä Oulun yliopistossa 7.–8.9.1979 pidetyt esitelmät*, red. Kari Suomi (Uleåborg, 1980), 1–19.

forskningsinstitut i världen som var utrustade med syntetiserare och den nödvändiga tekniska personalen, men redan i början av 1970-talet var talsyntetiserare standardutrustning även vid små forskningsenheter som fonetiklaboratorierna vid finska universitet. Tack vare de kommersiellt tillverkade Ove-maskinerna hade Coopers vision blivit verklighet. Förutom Åbo universitet hade även Jyväskylä universitet en egen Ove Ib. Ove IIIb i Åbo användes även av några andra finska universitet när de hade behov av att syntetisera längre sekvenser.<sup>77</sup> Antti Sovijärvi vid Helsingfors universitet, professor i fonetik och banbrytare för den finska vokalforskningen, använde en Ove IIIb.<sup>78</sup> Det internationella forskningsfältet hade gemensamma intressen med avseende på forskningsteman såväl som metodologi, men den geografiska närheten kan ha spelat en roll för Ove-enheternas popularitet i Finland. Inget laboratorium i Finland verkar ha använt till exempel Pattern Playback-syntetiseraren, trots att psykolingvisten Willem Levelt har lyft fram den som standardverktyget för att generera tal före datorsyntesens tid och Fant själv prisat apparatens pedagogiska egenskaper.<sup>79</sup>

Digital databehandling förändrade hur talsyntetiserarna användes och vilka kontrollmetoder man valde, men de underliggande syntesmetoderna förändrades inte så mycket. Ur forskningens perspektiv gjorde digitaliseringen det dessutom lättare och flexiblare att skapa syntetiska talstimuli. Fonetiklaboratoriet i Åbo hade tillsammans med Centret för kognitiv neurovetenskap 1997 skaffat ett datorprogram som innehöll Klatt88-syntetiseraren, och det användes för att skapa stimuli för studentuppgifter och forskning.<sup>80</sup> Klatt88 var en formantsyntetiserare som hade utvecklats på 1970-talet av Dennis Klatt vid MIT. I grund och botten var det en utvecklad version av syntetiserare som Ove och Pat: den kombinerade deras kaskad- och parallellstrukturer i en maskin. Den började användas som modell för flera text-till-tal-syntessystem, så som MITalk, Dectalk, Prose-2000 och Klattalk.<sup>81</sup>

Douglas Robertson har undersökt datorrevolutionen inom naturvetenskap, och han lyfter fram att datorerna medförde nya och utvidgade möjligheter för visualisering.<sup>82</sup> Samtidigt förstärkte de talforskarnas förmåga att höra och lyssna. Men som Jon Agar har påpekat tjänar datorisering oftast bara till att automatisera processer som redan existerar, och så var fallet även inom fonetikfältet.<sup>83</sup> Man kan hävda att datoriseringen urholkade värdet av tyst virtuositet i lyssnandet allteftersom programmering av syntetiskt tal i allt lägre grad kom att bestå i att mödosamt pröva sig fram.

Talsyntesen transformerade fonetikforskningen. Fastän den uppfattades som ett verktyg och en modell, och inte som ett mål i sig, fick den omfattande effekter på vilka ämnen som kunde undersökas och vilka resultat som kunde genereras. Enligt lingvisten Mary Beckman ”handlade det inte bara om att grundforskningen i talvetenskapen tillhandahöll grundläggande kunskap för utvecklingen av teknologier i talsyntes. Själva grundforskningen kunde inte ha genomförts utan utvecklingen inom talsyntes.”<sup>84</sup> Detta är lätt att urskilja i mångfalden av perceptuella studier som bygger på att testpersoner får lyssna till noggrant genererade stimuli.

Datoriseringen medförde även en annorlunda effekt, som Gunnar Fant i efterhand beskrev som en övergång mot mer tillämpningsorienterade aktiviteter inom talforskning.<sup>85</sup> När Fant såg tillbaka på sin tid inom talforskningen från ett 2000-talsperspektiv konstaterade han:

Jag anser att grundforskningen hade en relativt framträdande plats under första halvan av århundradet, medan applikationerna däremot dominerade den andra halvan, till följd av datavetenskapens allt växande roll. Jag har beskrivit detta fenomen som en polarisering av forskningsinsatserna. Datorkraft kan inte ersätta djupgående kunskap om talkoden.<sup>86</sup>

## Vetenskapligt lyssnande mellan kommersiella applikationer

Efter att i flera årtionden främst ha varit ett akademiskt fenomen började talsyntes i slutet av 1970-talet och början av 1980-talet träda fram i det allmänna teknologiska medvetandet i form av kommersiellt tillverkade och marknadsförda tillämpningar. En av de första mer spridda konsumtionsprodukterna baserade på formantsyntes var den pedagogiska leksaken Speak & Spell, som Texas Instruments lanserade 1978. Användaren kunde öva på att stava ord som maskinen läste upp, eller skriva in bokstäver och höra maskinen läsa upp dem. Leksaken blev berömd för att rymdvarulen E.T. använde den för att ringa hem i filmen 1982.<sup>87</sup> Sedan 1970-talet har robotar med begränsad talsyntes använts som attraktioner vid teknikevenemang och produktmässor. När Steve Jobs introducerade den första Macintosh-datorn 1984 var showens höjdpunkt när han lät datorn tala genom att använda text-till-tal-programmet Macintalk.<sup>88</sup> Förutom arkadspel och andra förströelser, högtalarutrop samt talande bilar och andra modeflugor förblev assisterande teknik den främsta konsumenttillämpningen av syntetiserat tal. Dectalk-syntetiseraren var utrustad med

några standardröster, där man även kunde justera vissa parametrar. Den förvalda manliga rösten Perfect Paul blev välkänd som Stephen Hawkings syntetiska röst sedan slutet av 1980-talet.<sup>89</sup>

Hurudant var förhållandet mellan kommersiella applikationer av talsyntes och den akademiska talforskningen? Utforskningen av formanter och deras syntetisering var grunden för tillämplig talsyntes inom olika fält, men forskarna själva ville hålla sig på avstånd från applikationerna. På fonetiklaboratoriet vid Åbo universitet framhöll personalen att talsyntes först och främst var en forskningsmetod för akustisk fonetik, även om försök att vidareutveckla syntetiserare för andra syften bedrevs parallellt med den akademiska forskningen.<sup>90</sup>

Samtidigt som de kommersiella applikationerna av talsyntes började lanseras verkade den aktivaste perioden för akademisk forskning om fonetik och syntes vara förbi. I Åbo miste fonetikdisciplinen sin professur när Wiik pensionerades 1997.<sup>91</sup> Vid många finska universitet tynade fonetik bort som självständig disciplin eller institution, och mellan 1980-talet och början av 2000-talet minskade antalet tjänster till hälften.<sup>92</sup> Samtidigt började maskinerna att försvinna. Ännu i slutet av 1990-talet användes Ove IIIB-syntetiseraren för att generera vokalljud för en magisteravhandling i Åbo.<sup>93</sup> Runt 2005 kasserades maskinen i samband med en flytt, som för att markera slutet på en era.

På KTH:s taltransmissionslaboratorium var utvecklingen inte lika starkt avtagande, men likväl utsattes även Fant och hans forskningsgrupp för påtagliga förändringar. I Fants festskrift 1979 påpekade Liljencrants att gynnsamma förhållanden hade rått under 1960-talet men att det senaste årtiondet hade varit svårare:

Under sjuttioalet har grundforskningen befunnit sig i en rätt krävande ställning i fråga om finansieringen; det har funnits en stark betoning på att man förväntas ta fram praktiska tillämpningar. Gunnar har alltid varit ytterst envis med att grundforskning måste finnas i bakgrunden till praktiska tillämpningar, även om den måste bedrivas på bakgården ibland.<sup>94</sup>

Långt senare, 2005, skrev Fant själv: ”Redan från början har jag blivit van att sälja grundforskning under täckmanteln av potentiella applikationer, vilket nuförtiden är den rådande situationen överallt.”<sup>95</sup> Eftersom laboratoriet var inhyst i en teknisk högskola och dess finansiering till stor del kom från Styrelsen för teknisk utveckling (STU) fanns stor efterfrågan på praktiska tillämpningar. Assisterande teknik för personer med variationer i hörsel- och talförmågan var ett område som laboratoriet fokuserade på,

till exempel genom att 1978 utveckla en talmaskin för en svensk tonåring med cp i ett experimentellt projekt.

Carlson och Granström föreslog en tydligare fördelning av syntesarbetet mellan applikationer och talforskning, eftersom målen och kraven för talproteser och allmänna syntesstrategier skilde sig så mycket åt.<sup>96</sup> Ännu på 1970-talet användes dock syntes som ett forskningsverktyg. Exempelvis tillverkades 1977 en sjungande syntetiserare som ett avhandlingsprojekt, ”närmast för vetenskapligt bruk”, även om utvecklaren ansåg att den möjligtvis också kunde användas som ett musikinstrument.<sup>97</sup>

Fant betonade själv gång på gång vikten av att hålla igång grundforskningen, också för att kunna stödja möjliga tillämpningar. I sina senare texter, i början av 2000-talet, poängterade han att kunskapen om talet fortfarande var bristfällig, varför han efterlyste en robustare teoretisk grund som även kunde höja talteknologiernas kvalitet.<sup>98</sup> Enligt honom hade talsyntesens tekniska framgångar gett upphov till vanföreställningar om att man hade uppnått djupgående insikt (*profound insight*) i mänskligt tal.<sup>99</sup> Syntesen, ett hörbart belägg för vetenskapliga fakta, tycktes ha vänt sig mot sina utvecklare i och med att problemet framstod som tekniskt övervunnet. Tack vare talforskarnas vetenskapliga arbete hade syntetiserat tal blivit tillräckligt lätt att producera och lät tillräckligt bra för att kunna lösgöra sig från sina akademiska kopplingar. Att lyssna till syntetiskt tal var inte längre enbart en vetenskaplig kunskapspraktik.

Under den tidsperiod som detta kapitel omfattar framträdde lyssnandet till syntetiserat tal som en vetenskaplig praktik, om än besläktad med och kopplad till de teknologier som tillämpade talöverföring i konsumentprodukter. Lyssnandet som vetenskaplig praktik var frikopplat från omedelbara kommersiella och tekniska syften. Det hade ett vetenskapligt syfte: att bättre förstå mänskligt tal. Inom forskningen var det en del av den ständiga metodologiska växlingen mellan hörande och seende, vilken tjänade det komplicerade uppdraget att behandla och analysera tal så att det till sist kunde skrivas ut på papper. Tack vare att de som producerade och bevittnade det syntetiserade talet var virtuosa lyssnare blev praktiken till ett vetenskapligt faktum som kunde tala för sig självt. Ur lyssnande till telefoner föddes en epok av aktiv och dynamisk talforskning som var sammanvävd med talsyntes, tills den återgick till tekniska frågor som gällde datorsyntesens kommersiella tillämpningar.

Lyssnande var en väsentlig praktik inom talforskningen, och olika slags lyssnande förekom i olika sammanhang: det pedagogiska lyssnandet förmedlade tyst kunskap till blivande forskare, det experimentella lyssnandet

var en komponent i produktionen av vetenskapliga fakta med syntes som verktyg, det bevittnande lyssnandet av experter bekräftade vetenskapliga teorier som syntetiserarna manifesterade. I forskningspublikationer förblir alla dessa praktiker dolda i tystnaden mellan raderna, men syntetiserarna själva gestaltar dem på ett närvarande och materiellt sätt. Med sina röster bär de högljudd vittnesbörd om dem som visste att lyssna.

## Noter

1. Även det skrivna ordet är en transduktion mellan två sinnen: hörsel och syn. Hur ord ser ut styr hur vi uppfattar dem. Akronymerna i texten är ett exempel på det: "OVE", skrivet med versaler och uttalat "o ve e", står för vetenskaplighet. Men samtidigt som forskarna gärna använde "OVE" som en vetenskaplig symbol talade de om "Ove", skrivet med gemener och uttalat som ett vanligt mansnamn, som en viktig arbetskollega. Efter att i det inledande citatet ha skrivit akronymerna med versaler övergår vi nu till gemener, vilket illustrerar denna dynamik.
2. Inspelning och transkription 83,8 i arkivet för Smithsonian Speech Synthesis History Project vid National Museum of American History.
3. Gunnar Fant, "Historical notes", *Speech, Music and Hearing – Quarterly Progress and Status Report*, vol. 47, nr 1 (2005).
4. Dennis H. Klatt, "Review of text-to-speech conversion for English", *Journal of the Acoustical Society of America*, vol. 82, nr 3 (1987), ljudillustrationer på lp-skiva.
5. Steven Shapin & Simon Schaffer, *Leviathan and the air-pump. Hobbes, Boyle, and the experimental life* (New Jersey, 1985), s. 78–79.
6. Se t.ex. Brad H. Story, "History of speech synthesis", i *The Routledge handbook of phonetics*, red. William F. Katz & Peter F. Assmann (London, 2019).
7. Se Fabian Brackhane & J. Trouvain, "Stationen mechanischer Sprachsynthese vom 18. bis zum 20. Jahrhundert", *Studientexte zur Sprachkommunikation*, vol. 68 (2013).
8. Se Jonathan Sterne, *The audible past. Cultural origins of sound reproduction* (Durham, 2003), s. 79.
9. Se Joseph Auner, "'Sing it for me': Posthuman ventriloquism in recent popular music", *Journal of the Royal Musical Association*, vol. 128, nr 1 (2003); Jacob Smith, "Tearing speech to pieces: Voice technologies of the 1940s", *Music, Sound, and the Moving Image*, vol. 2, nr 2 (2008), s. 183–206.
10. J.L. Flanagan, *Speech analysis, synthesis and perception* (Heidelberg, 1972), s. 211; Klatt, "Review of text-to-speech"; Juergen Schroeter, "Section introduction. Articulatory synthesis and visual speech", i *Progress in speech synthesis*, red. J.P.H. van Santen, J.P. Olive, R.W. Sproat & J. Hirschberg (New York, 1997), s. 179; Roberto Pieraccini, *The voice in the machine. Building computers that understand speech* (Cambridge, 2012), s. 194.
11. Smith, "Tearing speech to pieces", s. 184; Mara Mills, "Media and prosthesis: The vocoder, the artificial larynx, and the history of signal processing", *Qui Parle: Critical Humanities and Social Sciences*, vol. 21, nr 1 (2012), s. 109.
12. Mills, "Media and prosthesis", s. 110–111.
13. Story, "History of speech synthesis", s. 15; Mills, "Media and prosthesis", s. 134; Friedrich Kittler, *Grammophon Film Typeuriter* (Berlin, 1986), s. 77–78.
14. Se Smith, "Tearing speech to pieces", s. 189.
15. David A. Mindell, *Between human and machine. Feedback, control, and computing before cybernetics* (Baltimore, 2002), s. 132–133.
16. Se Jon Agar, *Science in the twentieth century and beyond* (Cambridge, 2012), s. 373–374; Kittler, *Grammophon*, s. 374.



17. Norbert Wiener, *The human use of human beings* (1950; London, 1989), s. 79. Norbert Wiener, *Materia, maskiner, människor* (Stockholm, 1964), s. 87.
18. Willem J.M. Levelt, *A history of psycholinguistics. The pre-Chomskyan era* (Oxford, 2013), s. 550–551.
19. Norbert Wiener, "Speech, language, and learning", *The Journal of the Acoustical Society of America*, vol. 22 (1950), s. 697.
20. Thomas L. Hankins & Robert J. Silverman, *Instruments and the imagination* (Princeton, 1995), s. 219; Frank Rives Millikan, "Joseph Henry and the telephone" i *Smithsonian Institution Archives* (Washington, 2007); Flanagan, *Speech analysis*, s. 206.
21. Mills, "Media and prosthesis", s. 135.
22. Homer Dudley & T.H. Tarnoczy, "The speaking machine of Wolfgang von Kempelen", *The Journal of the Acoustical Society of America*, vol. 22, nr 2 (1950), s. 151–166. Därefter framträder Kempelens maskin som en etablerad startpunkt i syntestens historia, se t.ex. Flanagan, *Speech analysis*, s. 205–206.
23. Klatt, "Review of text-to-speech".
24. Gunnar Fant, "Half a century in phonetics and speech research", i *Fonetik 2000. Proceedings*, red. A. Botinis & N. Torstensson (Skövde, 2000); Gunnar Fant, *Speech acoustics and phonetics. Selected writings* (Dordrecht, 2004a), s. ix.
25. Gunnar Lidén & Gunnar Fant, "Swedish word material for speech audiometry and articulation tests", *Acta Oto-Laryngologica*, vol. 43, nr 116 (1954).
26. Fant, *Speech acoustics and phonetics*, s. 216.
27. Gunnar Fant, "Speech research in a historical perspective", i *From sound to sense. Fifty years of discoveries in speech communication*, red. Janet Slifka, Sharon Manuel & M. Matthies (Cambridge, 2004b).
28. Fant, "Half a century", s. 2.
29. Rolf Carlson & Björn Granström, "Some notes about Gunnar, Ove, and speech synthesis", i *Gunnar Fant 60 years. Speech, Music and Hearing – Quarterly Progress and Status Report*, vol. 20, nr 2 (1979); Fant, *Speech acoustics and phonetics*, s. 1; "Half a century".
30. Story, "History of speech synthesis", s. 18; Levelt, *History of psycholinguistics*, s. 552.
31. Fant, "Half a century", s. 2.
32. Fant, "Half a century"; *Speech acoustics and phonetics*.
33. Fant, "Historical notes", s. 2; Carlson & Granström 1979; Gunnar Fant & Janos Mártony, "Speech synthesis: Instrumentation for parametric synthesis (OVE II)", *Speech Transmission Laboratory – Quarterly Progress and Status Report*, vol. 2 (1962).
34. Johan Liljencrants, "The conjurer's apprentice", i *Gunnar Fant 60 years. Speech, Music and Hearing – Quarterly Progress and Status Report*, vol. 20, nr 2 (1979).
35. Rolf Carlson & Björn Granström, "In memoriam: Johan Liljencrants (1936–2012)", *Journal of the International Phonetic Association*, vol. 42, nr 2 (2012).
36. Fant, "Half a century"; Johan Liljencrants, "The OVE III speech synthesizer", *IEEE Trans.Audio Electroac.* vol. 16, nr 1 (1968); Carlson & Granström, "Johan Liljencrants (1936–2012)".

37. Liljencrants, "The conjurer's apprentice", s. 14.
38. Dudley & Tarnoczy, "Speaking machine", s. 151.
39. Story, "History of speech synthesis", s. 17; Gunnar Fant, *Modern instruments and methods for acoustic studies of speech* (Stockholm, 1958), s. 6.
40. Se Flanagan, *Speech analysis*, s. 150.
41. Fant, "Half a century".
42. Gunnar Fant, "Sound spectrography", i *Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki (1961)*, red. Antti Sovijärvi, Pentti Aalto (Haag, 1962), s. 21; Fant, "Historical notes".
43. Se till exempel Schroeter, "Articulatory synthesis and visual speech", s. 179; Björn Lindblom, Johan Sundberg, Peter Branderud, Hassan Djamshidpey & Svante Granqvist, "The Gunnar Fant legacy in the study of vocal acoustics", *10ème Congrès Français d'Acoustique, Apr 2010* (Lyon, 2010).
44. Gunnar Fant, *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations* (Haag, 1960), s. 20.
45. Fant, *Acoustic theory of speech production*, s. 93; "Speech research in a historical perspective".
46. Se Fant, *Acoustic theory of speech production*, s. 113. Som exempel på sådan forskning nämner Fant den finske fonetikern Antti Sovijärvis doktorsavhandling från 1938.
47. Se Lindblom m.fl., "The Gunnar Fant legacy".
48. Rolf Carlson, & Björn Granström, "Rule-based speech synthesis", i *Springer handbook of speech processing*, red. Jacob Benesty, M. M. Sondhi & Yiteng Huang (Heidelberg, 2008).
49. F. S. Cooper, "Speech synthesizers", i *Proceedings of the 4th International Congress of Phonetic Sciences (Helsinki, 1961)* red. Antti Sovijärvi, Pentti Aalto (Haag, 1962), s. 8–10; Story, "History of speech synthesis", s. 19–20.
50. Story, "History of speech synthesis", s. 19.
51. Liljencrants, "The conjurer's apprentice", s. 15.
52. Shapin & Schaffer, *Leviathan and the air-pump*, s. 78–79.
53. Cooper, "Speech synthesizers", s. 4.
54. Bruno Latour & Steve Woolgar, *Laboratory life* (Princeton, 1986), s. 33–39.
55. Story, "History of speech synthesis", s. 27.
56. Carlson & Granström, "Gunnar, Ove, and speech synthesis", s. 16.
57. Steven Shapin, "Pump and circumstance: Robert Boyle's literary technology", *Social Studies of Science*, vol. 14, nr 4 (1984).
58. Cooper, "Speech synthesizers", s. 10.
59. Jfr Lindblom m.fl., "The Gunnar Fant legacy".
60. Harry Collins, "Public experiments and displays of virtuosity: The core-set revisited", *Social Studies of Science*, vol. 18, nr 4 (1988).
61. Liljencrants, "The conjurer's apprentice", s. 13.
62. Cooper, "Speech synthesizers", s. 5.

63. Ibid., s. 11.
64. Harry Collins, "Interactional expertise as a third kind of knowledge", *Phenomenology and the Cognitive Sciences*, vol. 3, nr 2 (2004).
65. Kalevi Wiik, "Phoneme boundaries of Finnish vowels", *Proceedings of the 4th International Congress of Phonetic Sciences, (Helsinki, 1961)* red. Antti Sovijärvi, Pentti Aalto (Haag, 1962), s. 793.
66. Markus Mattila, "From Ove to Klatt: Speech synthesis at Turku phonetics laboratory", i *Studies in speech communication. Dedicated to Ilkka Raimo on his 65<sup>th</sup> birthday*, red. Maija S. Peltola & Jyrki Tuomainen (Åbo 2004), s. 2.
67. Jarkko Metsätähti, intervju 27 januari 2020; Mattila, "From Ove to Klatt", s. 2. Eva baserades på samma funktionslogik som Fants tidiga syntesapparat Lea (Line Electrical Analog), byggd i slutet av 1950-talet, och andra motsvarande simulationer av talapparaten. Se Fant, *Methods for acoustic studies of speech*, s. 74–76.
68. Mattila, "From Ove to Klatt", s. 5.
69. Pertti Hurme & Ilkka Raimo, "Suomen kielen sääntösynteesiä: Ensimmäinen versio", i *Fonetiikan päivät Tampere 1974* (Tammerfors, 1975); Mattila, "From Ove to Klatt", s. 7.
70. Carlson & Granström, "Gunnar, Ove, and speech synthesis", s. 19.
71. Olli Aaltonen & Markus Mattila, "Puhesynteesi fonetiikan opetuksessa ja tutkimuksessa Turussa", i *Fonetiikan päivät – Oulu 1979: IX Fonetiikan päivillä Oulun yliopistossa 7.–8.9.1979 pidetyt esitelmät*, red. Kari Suomi (Uleåborg, 1980); Hurme & Raimo, "Suomen kielen sääntösynteesiä".
72. Mattila, "From Ove to Klatt", s. 7.
73. Story, "History of speech synthesis", s. 19.
74. Aaltonen & Mattila, "Puhesynteesi fonetiikan".
75. Aaltonen, Olli, *Vowel perception. Behavioural and psychophysiological experiments* (Åbo 1997)
76. Cooper, "Speech synthesizers", s. 12.
77. Se <https://www.jyu.fi/tdk/museo/esineet/vokaalisyntetisaattori.html>.
78. Antti Iivonen, "Antti Sovijärvi", *Phonetica*, vol. 53, nr 1–2 (1996). Sovijärvi utvecklade syntes för finska och ungerska. I sin tidigare forskning hade han använt röntgenundersökning, i likhet med Fant.
79. Levelt, *History of psycholinguistics*, s. 552; Fant, "Historical notes", s. 14.
80. Mattila "From Ove to Klatt", s. 9.
81. Robert Donovan, *Trainable speech synthesis* (Cambridge, 1996); Story, "History of speech synthesis", s. 25.
82. Douglas S. Robertson, *Phase change. The computer revolution in science and mathematics* (Oxford, 2003).
83. Agar, *Science in the twentieth century*, s. 379.
84. Mary E. Beckman, "Speech models and speech synthesis", i *Progress in Speech Synthesis*, red. J.P.H. van Santen, J.P. Olive, R.W. Sproat & J. Hirschberg (New York, 1997), s. 186–187.

85. Fant, "Speech research in a historical perspective".
86. Fant, *Speech acoustics and phonetics*, s. 1.
87. Debbie A. Rowe, "From wood to bits to silicon chips: A history of computer synthesized speech", i *Computer synthesized speech technologies. Tools for aiding impairment*, red. John Mullenix & Steven Stern (Hershey, 2010).
88. Macintalk var inte en understödd del av produkten och det dröjde nästan tjugo år, till 2001, innan en text-till-tal-applikation blev standard i Macdatorerna. Rowe 2010, s. 14. Se en film från lanseringen på <https://www.youtube.com/watch?v=2B-XwPjn9YY>.
89. Donovan 1996; Story, "History of speech synthesis", s. 25. De andra rösterna var Beautiful Betty, Huge Harry, Frail Frank, Kit the Kid, Rough Rita, Uppity Ursula, Doctor Dennis och Whispering Wendy.
90. Aaltonen & Mattila, "Puhesynteesi fonetiikan".
91. Mattila, "From Ove to Klatt", s. 14.
92. Olli Aaltonen, "Puhe kommunikaatiomuotona ja tutkimuskohteena", *Puhe ja kieli*, vol. 28, nr 2 (2008).
93. Mattila, "From Ove to Klatt".
94. Liljencrants, "The conjurer's apprentice", s. 15.
95. Fant, "Historical notes", s. 10.
96. Carlson & Granström, "Gunnar, Ove, and speech synthesis", s. 19–20.
97. Björn Larsson, "Music and singing synthesis equipment (MUSSE)", *Department for Speech, Music and Hearing – Quarterly Progress and Status Report*, vol. 18, nr 1 (1977), s. 40.
98. Fant, "Half a century", s. 7.
99. Fant, "Speech research in a historical perspective", s. 11.