

Uppgift 1. I Tallinn i Estland finns ett unikt sångarstadion, Lauluvaljak.



Den gigantiska scenen är 73 meter bred, 32 meter djup, och rymmer femton tusen sångare som mest. Publiken sitter i en grässlätt mitt emot, som rymmer upp till trehundra tusen personer. Den jättestora halvkupolen är akustiskt sinnrik, och gör att man hör ganska bra utan aktiv förstärkning.

- (a) Löptiderna blir stora. Ungefär hur många millisekunders osynk upplever dirigenten i mitten som mest, om alla femtontusen sjunger precis på vad de ser av hennes slag? (2 p)
 $c = 340 \text{ m/s}$; $\Delta l \approx 32 \text{ m}$ ger **storleksordningen 100 ms.**
- (b) Ute på publikplats, där avstånden till alla sångarna är ungefär lika stort (näja), gäller att ju fler sångare på scenen, desto starkare blir det. Varje fördubbling av antalet okorrelerade men lika starka ljudkällor ger då en fördubbling av intensiteten. Om man ökar antalet sångare från 500 till 2500, hur många dB starkare blir då ljudet som når publiken? (2 p)
 $\Delta L = 10 \cdot \log(2500/500) = 10 \cdot \log(5) \approx +7,0 \text{ dB}$

Om man står inne i själva jättekören är dock avstånden till de andra sångarna väldigt olika. Vi ska nu härleda hur ljudintensiteten J i sångarklungans mitt varierar som funktion av klungans radie r , alltså som funktion av hur stor kören är. Vi antar för enkelhetens skull att antalet sångare per kvadratmeter σ är konstant över hela scenen, och att alla sångarna producerar samma akustiska effekt w vardera. Märk att det inte är värdet på J som efterfrågas, utan dess beroende av r . Det innebär att värdena för σ och w ej behöver vara kända för att erhålla lösningarna i (c) och (d). Kupolens eventuella inverkan försummas.

- (c) Tänk först att du omges av en enda ring av sångare på ett led, som alla står på avståndet r från dig. Antalet sångare per meter, tangentiellt längs ringens omkrets, är konstant: ju större r , desto flera sångare. Härled hur den ljudintensitet J som du nu utsätts för beror av r . Rita gärna. Det är bara *proportionaliteten* som är intressant, dvs om J är proportionell mot r , r^2 , $1/r$, $1/r^2$ eller någon annan funktion av r . (2 p)

J är proportionell mot antalet ljudkällor N som i sin tur är proportionellt mot r . J är också omvänt proportionell mot avståndet i kvadrat, $1/r^2$. Alltså är $J \propto r/r^2 = 1/r$.

- (d) Tänk sedan att ytterligare koncentriska ringar av sångare läggs till, med samma antal sångare per meter som tidigare, även radiellt. Visa hur ljudintensiteten i mitten ökar med allt fler ringar vars r ökar. *Ledning:* minns att integralen av $1/x$ är $\ln(x)$, för $x > 0$. (2 p)

Vi ska summera intensitetsbidragen över alla ringar, vilket är detsamma som att integrera svaret i (c) över dr :

$$J = \int_{r>0}^r \frac{a}{r} dr = a \cdot \ln(r) + C$$

Konstanten a kan införas här som en skalfaktor som inrymmer σ och w , men de beror inte av r , och det gör inte heller C . Svaret är alltså att $J \propto \ln(r)$. Ju större r blir, desto långsammare ökar intensiteten med r .

- (e) Skulle du som sångare i mitten alls höra skillnad, om antalet sångare N ökade från 500 till 2500? Antag att σ är 4 personer per m^2 , och beräkna den resulterande skillnaden i intensitetsnivå. (2 p)

Med ett σ på 4 personer per m^2 i en rund klunga får vi $r_{500} = 6,3$ m och $r_{2500} = 14,1$ m. Även om vi inte känner w så ges ändringen vi söker av

$$\Delta L_J = 10 \log\left(\frac{\ln(r_{2500})}{\ln(r_{500})}\right) \approx 10 \log\left(\frac{\ln(14,1)}{\ln(6,3)}\right) \approx 1,4 \text{ [dB]}. \text{ Det är en knappt märkbar skillnad.}$$

Uppgift 2. (10 p)

Ett cylindriskt rör som är stängt i båda ändar kan kallas en halvvågspipa. Det uppvisar resonans när ett helt antal halva våglängder är lika med rörets längd. Ett rör som är öppet i ena änden och slutet i andra änden kallas för en kvartsvågspipa. Det uppvisar resonans när ett udda antal fjärdedels våglängder är lika med rörets längd.

- a) Vilken typ av pipa är hörselgången respektive ansatsröret? Motivera. (3 p)

Hörselgången är en kvartsvågspipa. Av texten mitt på sidan 9-4 framgår att även ansatsröret nominellt kan betraktas som en kvartsvågspipa, åtminstone på öppna vokalljud. Det är så gott som stängt i änden vid stämbanden.

- b) Beräkna hörselgångens ungefärliga längd. (3 p)

Lägsta resonansen är 3 kHz vilket ger en våglängd om knappt 12 cm vid 37° C. En kvarts våglängd är då knappt 3 cm.

- c) Beräkna längden av den kvartsvågspipa som skulle ger samma lägsta resonansfrekvens som ansatsröret ger till vokalen "a", som i "far". (2 p)

Lägsta resonansen F_1 är ca 600 Hz vilket ger en våglängd om knappt 60 cm vid 37° C. En kvarts våglängd är då knappt 15 cm.

- d) Beräkna det relativa fel man får i (b) och (c) om man förutsätter normal rumstemperatur istället för människans kroppstemperatur. (2 p)

37° C resp 20° C insatta i formeln på sid 1-2 ger en skillnad på 2,9%.

Uppgift 3. Rumsakustik (10 p)

Bygg-Berit renoverar ett tomt rektangulärt vardagsrum med måtten $4,1 \times 8,0 \times 2,7$ m³. Golvet är av trä, taket är av betong, väggarna är putsade, och det finns totalt 6 m² glasfönster, som är stängda. Rummets α_m vid 500 Hz är 0,045. (α för fönsterglas beror egentligen av tjocklek m.m., men vi använder värdet för glas i kompendiets tabell.)

- a) Beräkna rummets efterklangstid. (3 p) $A = S \cdot \alpha_m = 5,89$ m², $V = 88,56$ m³ ger $T = 2,4$ s. De olika ytornas egenskaper kan man strunta i.

- b) När Bygg-Berits slipmaskin är igång, utstrålar den en akustisk effekt om 100 mW. Beräkna ljudnivån som då råder i större delen av rummet. Avrunda till hela decibel. (3 p)

"I större delen av rummet" innebär att vi först söker den diffusa intensiteten $J = 4W/A$; $L_J = 10 \log(J/J_{ref})$ ger sedan 108 dB. Ett felaktigt värde på A från (a) godtas som indata om det bara är rimligt.

- c) Bygg-Berit försöker sänka ljudnivån genom att öppna alla fönstren. ”Absorptionen” blir då 100% i fönsteröppningarna. Beräkna hur mycket ljudnivån då sjunker. Avrunda till hela decibel. (4 p)
 Fönstrens absorptionsyta ökar från 5 % till 100% av 6 m^2 . Rummets A blir alltså $0,95 \cdot 6 \text{ m}^2$ större än i (a). En upprepning av beräkningen i (b) ger då en skillnad på -3 dB . Det är rimligt eftersom A i stort sett har fördubblats, och då ska den diffusa intensiteten ha halverats.

Uppgift 4. Allmän audioteknik (10 p)

Ett, kryss eller tvåa? Markera det alternativ som ger **störst upplevd dynamik**, om övriga systemegenskaper och yttre betingelser hålls konstanta. Med upplevd dynamik menas här *skillnaden i hörnivå* mellan starkaste och svagaste nyttoljud som apparaten eller systemet förmedlar med bibehållen kvalitet. Skriv av rätt rad på inlämningsbladet.

Några val är lätta och några är inte så självklara. Om du tycker att det blir alltför förenklat att svara med 1-X-2 (”det beror ju på...”), skriv också en kommentar eller motivering till ditt val.

(1 p per delfråga. Några kommentarer eller motiveringar behövs inte för full poäng, om tipsraden är rätt, men de kan hjälpa om den inte är det.)

	<i>Egenskap eller betingelse</i>	<i>1</i>	<i>X</i>	<i>2</i>	<i>Svar</i>
a	Samplingsfrekvens	48000 Hz	spelar ingen roll	96000 Hz	X
b	Bitar per sampel	8	spelar ingen roll	16	2
c	Ditherbrus	av	spelar ingen roll	på	1
d	Slutstegets uteffekt	10W	spelar ingen roll	100W	2
e	Mikrofonens riktverkan i stökig inspelningsmiljö	kula	spelar ingen roll	kardioid	2
f	Loudness-funktion	av	spelar ingen roll	på	1
g	Antalet uppdragna mixerreglar, om endast en kanal får signal	1	spelar ingen roll	8	X
h	Fläktbuller i lokalen	på	spelar ingen roll	av	2
i	CD-spelarens objektiv	smutsigt	spelar ingen roll	rent	X
j	Kompressor	på	spelar ingen roll	av	2

Noter:

- a) egentligen kan man få litet bättre dynamik vid 96 kHz men det har vi inte talat om i kursen.
- b)
- c)
- d)
- e)
- f) När loudness är av, hörs basen minska snabbare mot lägre volym, skillnaden i hörnivå ökar

g) Uppgift 5. Talteknologi.

Läs tidningsartikeln nedan (ur *New Scientist*, 28 okt 2006) om ett nytt taligenkänningsystem.

Technology

It's the next best thing to a Babel fish

IMAGINE mouthing a phrase in English, only for the words to come out in Spanish. That is the promise of a device that will make anyone appear bilingual, by translating unvoiced words into synthetic speech in another language.

The device uses electrodes attached to the face and neck to detect and interpret the unique patterns of electrical signals sent to facial muscles and the tongue as the person mouths words. The effect is like the real-life equivalent of watching a television show that has been dubbed into a foreign language, says speech researcher Tanja Schultz of Carnegie Mellon University in Pittsburgh, Pennsylvania.

Existing translation systems based on automatic speech-recognition software require the user to speak the phrase out loud. This makes conversation difficult,



You'll speak her language

as the speaker must speak and then push a button to play the translation. The new system allows for a more natural exchange. "The ultimate goal is to be in a position where you can just have a conversation," says CMU speech researcher Alan Black.

In October 2005 Schultz and her colleague Alex Waibel demonstrated the first automatic translator that could pick up electrical signals from face and throat muscles and convert them into text or synthesised speech – a technique called sub-vocal speech recognition. This ran on a laptop and translated Mandarin Chinese

to English or Spanish, but it could only translate around 100 words, each of which had first to be spoken into the system by the user, to "train" it on their voice.

Now the team has developed a system that can recognise a potentially limitless lexicon. Their secret is to detect not just words but also the phonemes that form the building blocks of words. The system then uses these to reconstruct the word. To translate

"The secret is to detect not just words but also the building blocks of words"

from English to another language, the user only has to train the system on the 45 phonemes used in spoken English.

The researchers use software that has been taught to recognise which phonemes are most likely to appear next to each other and in what order. When it encounters a string of phonemes it is unfamiliar with or has only partially heard, it uses this knowledge to come up with a range of sequences that make sense given the surrounding phonemes and words, assigns a probability to each one, and then picks the one with the highest probability.

The system still has some way to go. Faced with a sequence of words it has never heard before, it picks the right phoneme sequence only 62 per cent of the time. This nevertheless ranks as "a very significant achievement" according to Chuck Jorgensen, who is working on using sub-vocal speech recognition to control robots at NASA's Ames Research Center in Moffett Field, California. "This is showing that the technology is really within reach."

Schultz's team plan to attach the phoneme recognition software to their prototype Spanish or German translators, once they have improved its accuracy. **Celeste Biever** ●

Besvara dessa frågor. Alla svar ska motiveras.

- (a) I kompendiet står det att "Den akustisk-parametriska omvandlingen av talsignalen är ett grundläggande steg vid all taligenkänning." Stämmer det även för detta nya system? (3 p)
Nej, systemet tar inte in ljud alls utan istället kroppens elektriska signaler.
- (b) Är det nya systemet rent mönsterigenkännande, eller även kunskapsbaserat? (3 p)
Det är inte bara mönsterigenkännande, eftersom det identifierar enskilda fonem och relaterar dem till kontexten. Det förutsätter alltså kunskap om det aktuella språket, till exempel.
- (c) Journalisten skriver "The secret is to detect not just words but also the building blocks of words" som om det vore en ny idé. Kommentera hur innovativt detta egentligen är, med ledning av kompendiets text. (4 p)
Av diskussionen i kompendiets avsnitt 10.2.5 framgår att det inte är någon ny idé.