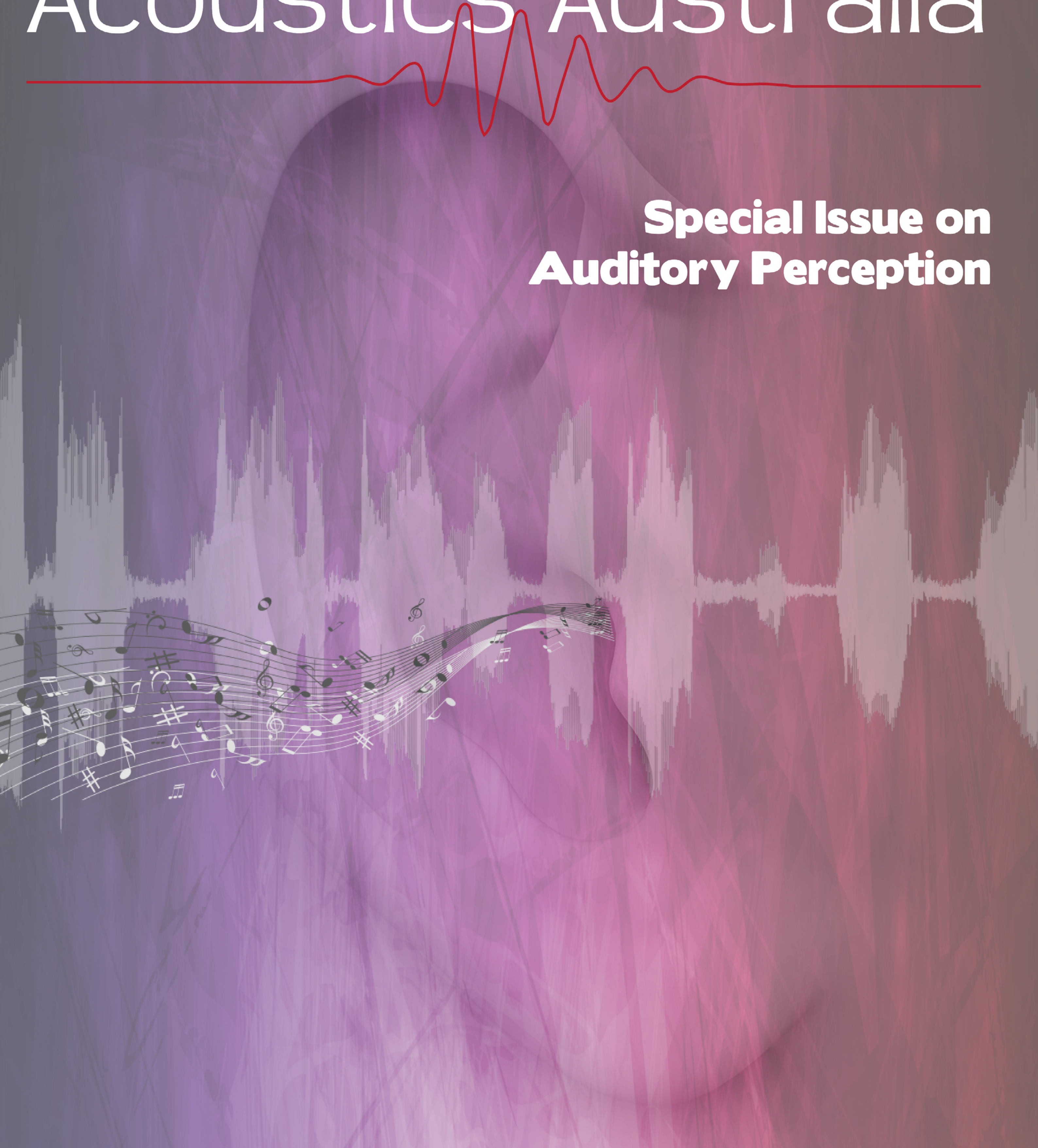
Acoustics Australia



Special Issue on Auditory Perception

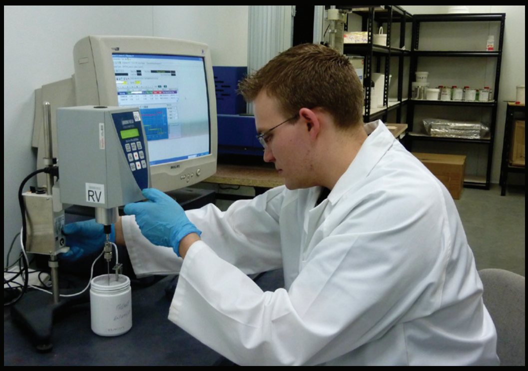
Australian Acoustical Society Volume 42 Number 2 August 2014

not
ROCKET SCIENCE
but
REAL SCIENCE

Pyrotek Noise Control recognises that the effective specification of a product needs reliable data. As such we are working with some of the world's leading testing and certification organisations to help support our materials with clear unbiased test data giving a specifier the facts about our products' performance.

With an ongoing research and testing budget we have some interesting data to share.

To keep up to date with all the additions to testing we are making, simply contact us for a personal visit or review our website www.pyroteknc.com and sign up for our 'PRODUCT UPDATE' emails and we will keep you up to date with our developments.



www.pyroteknc.com

Pyrotek
noise control

manufacturing quietness

you can hear a pin drop

Acoustics Australia

Acoustics Australia
All Editorial Matters
(articles, reports, news,
book reviews, new products, etc)
The Editor, Acoustics Australia
c/o Marion Burgess
AcousticsAustralia@acoustics.asn.au
www.acoustics.asn.au

General Business
(subscriptions, extra copies,
back issues, advertising, etc.)
Mrs Leigh Wallbank
P O Box 70
OYSTER BAY NSW 2225
Tel (02) 9528 4362
Fax (02) 9589 0547
wallbank@zipworld.com.au

Australian Acoustical Society
Enquiries see Page 77

Acoustics Australia is published by the Australian Acoustical Society
(A.B.N. 28 000 712 658)
ISSN 0814-6039

Printed by
Cliff Lewis Printing
91-93 Parraweena Rd,
CARINGBAH NSW 2229
Tel (02) 9525 6588
Fax (02) 9524 8712
email: matt@clp.com.au
ISSN 0814-6039

Vol. 42, No. 2 August 2014

From the Guest Editor . . . 75
From the President . . . 77
From the Chief Editor . . . 78

PAPERS

Special Issue: AUDITORY PERCEPTION

Native and Non-Native Speech Perception
Daniel Williams and Paola Escudero . . . 79

In Thrall to the Vocabulary
Anne Cutler . . . 84

Active Listening: Speech Intelligibility in Noisy Environments
Simon Carlile . . . 90

Auditory Grammar
Yoshitaka Nakajima, Takayuki Sasaki, Kazuo Ueda and Gerard B. Remijn . . . 97

Low Frequency Spatialization in Electro-Acoustic Music and Performance: Composition Meets Perception
Roger T. Dean . . . 102

Neuroscientific Investigations of Musical Rhythm
Daniel J. Cameron and Jessica A. Grahn . . . 111

Music Training: Lifelong Investment to Protect the Brain From Aging and Hearing Loss
Nina Kraus and Travis White-Schwoch . . . 117

Pitch Processing in Music and Speech
Barbara Tillmann . . . 124

Cochlear Implants Can Talk But Cannot Sing In Tune
Jeremy Marozeau, Ninia Simon and Hamish Innes-Brown . . . 131

ACOUSTICS FORUM

Introducing the Noise Database
Elizabeth Beach . . . 136

LETTERS

Infrasound Sensitivity and Harmonic Distortion
Henri Boutin, John Smith and Joe Wolfe . . . 137

REGULAR ITEMS

News . . . 138
Workplace Health and Safety News . . . 140
AAS News . . . 141
New Products . . . 143
Future Conferences . . . 146
Sustaining Members . . . 148
Diary . . . 150
Advertiser Index . . . 150

Cover design: Helena Brusic

FROM THE GUEST EDITOR

Auditory perception is a cornerstone of human communication. Nine articles collected for this special issue of "Acoustics Australia" review contemporary research with an emphasis on communication through speech perception and music perception. With immersion in a particular linguistic environment during infancy and early childhood, the acquisition of speech perception occurs seemingly effortlessly. But what processes enable children and adults' speech perception? To what degree are the statistical properties of a particular speech environment and perception of speech in that environment related? Are perceptual resources specific to speech or are they "domain-general"? These are just some of the research questions concerning speech perception that will be addressed.

Music, as far as we know, has been a part of every human culture and there has been lively debate about its possible evolutionary significance – adaption, exaption, epiphenomenon? Largely non-referential, distinguishing music from language, it is expressive, communicative, and personally, socially and culturally meaningful. In the present context of auditory perception, and especially when passed on in an "oral tradition", music can demonstrate constraints on auditory perception and memory.

The special issue reflects current work conducted in research laboratories in Australia, Canada, England, France, Japan and the USA. Two of the articles are concerned with speech perception (Cutler; Williams & Escudero), three with principles of auditory perception that apply to speech and music – grouping (Nakajima et al.), informational masking (Carlile) and localization (Dean), two on pitch processing (Tillmann; Marozeau et al.), and two on music perception (Kraus; Cameron & Grahn). A little more detail about each now follows.

Anne Cutler depicts speech recognition as a rapid competition between candidate words and reviews the way acoustics cues – phonemes, stress, and embedding patterns – differ across languages. The development of speech perception in infancy is described and theorized by Daniel Williams and Paola Escudero. Languages of the world and speakers of those languages provide a natural laboratory for research and Williams and Escudero bring speech perception processes into relief comparing native and non-native language listeners. In a similar way, Barbara Tillmann summarizes research wherein specialist populations have been sampled – musicians, speakers of tone languages, or people with congenital amusia (tone deafness) – to investigate the perception of pitch in music and in speech. Also focusing on pitch, Marozeau, Simon and Innes-Brown sketch the principles underlying cochlear implants (CIs) and explain the way CIs convey only two out of three main pitch cues.

Yoshitaka Nakajima and colleagues research auditory grouping and stream segregation. Their paper reports an auditory grammar and its application to phenomena such as auditory continuity and melody perception. This is the first report in English taken from their book published in Japanese in 2014. Roger Dean reviews research into localization of low frequency sound. While localization research has tended to concentrate on mid to high frequencies, Dean argues that as the spatialization of low frequencies is an artistic and creative device in electro-acoustic composition and performance it warrants experimentation. Informational masking is the topic of Simon Carlile's review. Prescient comments by Colin Cherry concerning the "cocktail party problem" 60 years ago are revisited and the significance and complexity of attention in perceiving speech in noisy environments documented.

In the context of aging populations and older adults' communication difficulties, including perceiving speech in noise, Nina Kraus and Travis White-Schwoch review research that demonstrates potential benefits and compensation from music training. Daniel Cameron and Jessica Grahn discuss a fundamental dimension of music and its power to move us – rhythm. The review outlines recent studies of rhythm perception and synchronization using functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG) and Transcranial Magnetic Stimulation (TMS) methods. The review concludes that motor system excitability is modulated by listening to musical rhythm.

For the opportunity to edit this special issue and for her guidance and sage advice throughout the process, I am grateful to Chief Editor, Marion Burgess. Heartfelt thanks to the authors for their thoughtful contributions, Editorial Assistant Truda King, for the cover Helena Brusic with assistance from Kate Falkenburg, and for their time, incisive comments and fast turn-around of manuscripts, our 18 reviewers: Jorg Buchholz, Densil Cabrera, Harvey Dillon, Neville Fletcher, Manon Grube, Julene Johnson, Ghada Katthab, Stephen McAdams, Elizabeth Margulis, Robert Mayr, Sylvie Nozaradan, Kirk Olsen, Aniruddh Patel, Brett Swanson, Bill Thompson, Leon van Noorden, Joe Wolfe and Ivan Zavada. I hope all readers enjoy this special issue.

Catherine (Kate) Stevens,

MARCS Institute and School of Social Sciences & Psychology, University of Western Sydney

kj.stevens@uws.edu.au
http://marcs.uws.edu.au

MUNISENSE

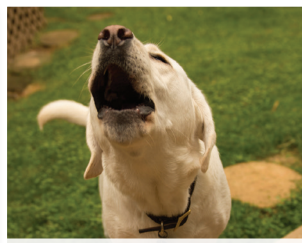
Wireless networked noise monitoring systems

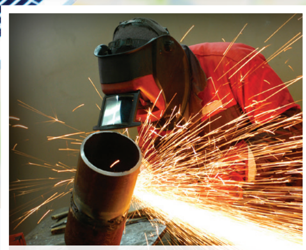
SAVTEK SALES+ RENTALS


Construction noise


Mining noise


Aircraft noise

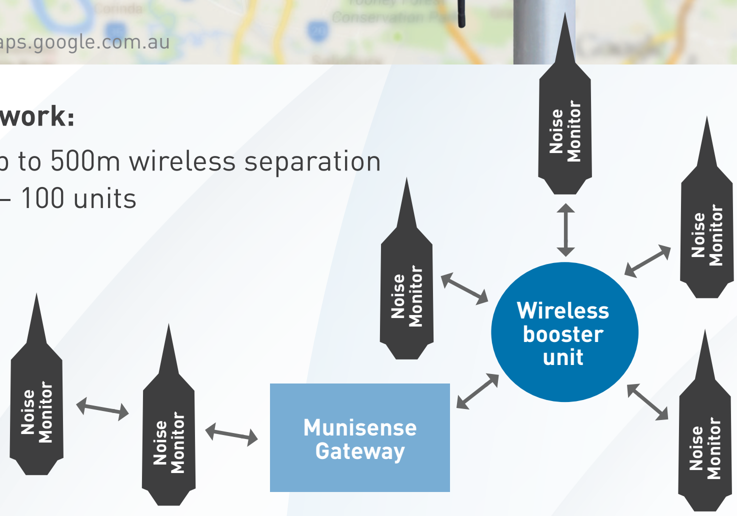

Community noise


Industrial noise


Rail & traffic noise

Network:

- Up to 500m wireless separation
- 1 – 100 units



Features:

- Type 1 or Type 2
- Solar powered
- Outdoor weather-proof microphone

Options:

- Streaming audio
- Directional noise monitor
- Synchronised weather data

FROM THE PRESIDENT

Well folks, as I come to the end of my present term as President of the AAS, I can state that we are looking good.

As you all know, Internoise 2014 is the big event for this year and things are shaping up for this conference to be a huge success. We have over 1000 abstracts and a huge exhibition with many international visitors heading our way to enjoy Melbourne and our conference. We also have in the order of 80 papers by AAS members who could be eligible to win the Gerald Riley Prize for the best AAS paper at Internoise. This latter Award is for $2500 and is sponsored by the Victorian Division.

With respect to other matters of note, a major achievement is that Acoustics Australia has negotiated a contract with Springer Publishing to be the publisher of the journal, effective from January 2015. The rights for AAS members will remain the same with free electronic access to the journal via a web link plus a low resolution pdf available for each member. Each issue will still have the sections on news, notes, advertising etc prepared by the AAS. The Chief Editor plus the editorial panel, appointed by AAS, will have the primary responsibility regarding the papers and technical notes, however, the processing tasks leading to the final publication will be undertaken by Springer Publishing. The first noticeable change will be an online submission and paper management system which should be operational late 2014. As well as fully indexing all the articles and technical notes, Springer will be promoting the journal widely. These will lead to an enhanced reputation for our journal.

We have done a lot of work behind the scenes to upgrade our website. You would have noticed some of the improvements, particularly with respect to renewal of subscriptions. The payment gateway however does seem to be working well and we hope that the upgraded website will be at final release by the time you read this message. Please make use of the expanded capability of the new site.

Another exciting development is the finalisation of the terms and conditions for our proposed Research Grants. The first call for applications is expected to be in August and we look forward to supporting some good research that will be of interest to all members. We hope to be able to make an announcement regarding a Grant at Internoise in November.

Finally, the Code of Ethics is up for review and other administrative reviews are being conducted, including the use of the Society's logo.

Tracy Gowen will be your next President and I hope that you will support Tracy in her new role. I want to thank Peter Heinze as outgoing VP and Federal Council for their support over the last two years.

I look forward to seeing many of you in Melbourne this November. Regards for now,

Norm Broner

AUSTRALIAN ACOUSTICAL SOCIETY ENQUIRIES

NATIONAL MATTERS

* Notification of change of address
* Payment of annual subscription
* Proceedings of annual conferences

Richard Booker - General Secretary
Australian Acoustical Society
PO Box 1843 Toowong DC QLD 4066
Tel: (07) 3122 2605
GeneralSecretary@acoustics.asn.au
www.acoustics.asn.au

SOCIETY SUBSCRIPTION RATES

For 2014/15 Financial Year (Inc. GST):

Fellow and Member	$150
Graduates	$115 - $150
Associate and Subscriber	$115
Retired and Maternity	$45
Student	No Charge

DIVISIONAL MATTERS

Enquiries regarding membership and sustaining membership should be directed to the appropriate State Division Secretary

AAS - NSW Division
Laura Allison
c/- AECOM
Level 21, 420 George Street
Sydney, NSW 2000
Tel: (02) 8934 0035
Fax: (02) 8934 0001
Laura.Allison@aecom.com

AAS - Queensland Division
PO Box 760
Spring Hill Qld 4004
Sec: Richard Devereux
Tel: (07) 3217 0055
Fax: (07) 3217 0066
rdevereux@acran.com.au

AAS - SA Division
AECOM,
Level 28, 91 King William St
ADELAIDE S.A. 5005
Sec: Darren Jurevicius
Tel: (08) 7100 6400
Fax: (08) 7100 6499
darren.jurevicius@aecom.com

AAS - Victoria Division
C/- JACOBS
PO Box 312 Flinders Lane
Melbourne VIC 8009
Sec: Peter Pirozek
Tel: (03) 8668 6356
ppirozek@globalskm.com

AAS–WA Division
Australian Acoustical Society
Western Australia Division
Unit 3/2 Hardy Street
South Perth WESTERN
AUSTRALIA 6151
Sec: Kingsley Hearne
Tel: (08) 9474 5966
kingsley@gabriels.net.au

FROM THE CHIEF EDITOR

The suggestion for a special issue on "Auditory Perception" came from Neville Fletcher and I was delighted when Professor Kate Stevens, from the MARCS Institute and School of Social Sciences & Psychology at the University of Western Sydney, took up the challenge. She has used her networking capacity to bring together a varied range of papers on different aspects of Auditory Perception. This issue of nine papers, with a few in line for the December supplement, is a testament to her enthusiasm and diligence.

This issue of the journal marks the first of a number of changes in the journal as it is the first to be produced with the primary mode of distribution as email attachment and not as hard copy delivered by post. Thus we are no longer focussed on optimising the use of every hard copy page and so you will note some little changes in the layout.

It is also with pleasure that we announce another change in that a contract has been signed with Springer Publishing to produce the journal from the commencement of 2015. The AAS maintains the editorial responsibilities for the content and Springer will take over the processing tasks leading to the publication. We expect the online paper submission system to be ready in late 2014.

We have retained the rights of the AAS members to have free access to the journal. Also each issue will include as "front matter" all the familiar sections like news, new products etc.

The Springer paper management system and publishing will replace the current inefficient manual processing. The reputation that goes with Springer publishing will enhance the status of the journal. There will be benefits to authors in that, once processed through the review and accepted, a pre-publication version will be immediately available online. As well, the formal indexing and promotion of the content via Springer should in due course increase the citation rate and the impact factor.

Although the journal is going through changes, the editorial policy will be maintained, namely that Acoustics Australia does not aspire to be primary research journal but to publish articles that have some relevance to the AAS membership. However the content depends upon what is received so a reminder to all members to consider submitting to Acoustics Australia.

In conclusion, I am very pleased to advise that in early 2015 we intend to have an issue on Room Acoustics with Sir Harold Marshall as the special issue editor. In addition to the invited papers for this issue, we welcome contributed papers and technical notes on this topic.

Marion Burgess
Chief Editor

ACOUSTICS AUSTRALIA INFORMATION

GENERAL BUSINESS
Advertising Subscriptions

Mrs Leigh Wallbank
PO Box 70, OYSTER BAY 2225
Tel (02) 9528 4362
Fax (02) 9589 0547
wallbank@zipworld.com.au

PRINTING, ARTWORK
Cliff Lewis Printing
91-93 Parraweena Road
CARINGBAH NSW 222
Tel (02) 9525 6588 Fax (02) 9524 8712
email: matthew@clp.com.au

ARTICLES & REPORTS NEWS, BOOK REVIEWS NEW PRODUCTS

The Editor, Acoustics Australia
c/o Marion Burgess
acousticsaustralia@acoustics.asn.au

SUBSCRIPTION RATES

1 year - A$77.20

Rates include GST.

All subscriptions will be provided in pdf
Hard copy available on application
Extra charge $120 per year plus postage
Agents rates are discounted.

ADVERTISING RATES
Casual Rates - Single Issue

	Non-member	Sus Mem
Full Page	862.40	768.90
Half Page	559.90	503.80
1/3 Page	431.20	388.30
1/4 Page	359.70	323.40

COLUMN RATE $29.70
WEB ADS $467.50

All rates are AUD and include GST

All enquiries to:
Mrs Leigh Wallbank
Tel (02) 9528 4362 Fax (02) 9589 0547
wallbank@zipworld.com.au

NATIVE AND NON-NATIVE SPEECH PERCEPTION

Daniel Williams[1] and Paola Escudero[2]

[1]Institute of Humanities & Creative Arts, University of Worcester, Worcester, WR2 6AJ, United Kingdom

[2]MARCS Institute, University of Western Sydney, Sydney, Australia

d_williams@worc.ac.uk

This review examines research on speech perception by both native and non-native listeners. The development of speech perception in infancy is first considered and a theoretical model that accounts for this is introduced. A brief overview then follows of several research areas under the umbrella of non-native speech perception, namely cross-dialect, cross-language and second-language speech perception. It is shown that non-native and native speech perception is critically shaped by the specific ways in which speakers use acoustic cues in speech production.

INTRODUCTION

A central goal of research on speech perception has been to uncover the auditory dimensions that listeners use to derive linguistically meaningful utterances from the speech signal. In order for listeners to successfully decode an utterance, they must be able to reliably distinguish the speech sounds of a particular language.

The speech signal presents listeners with a multitude of acoustic information, along different auditory dimensions (e.g., formants or voice onset time), that lies within the limits of human hearing. Perceiving speech is not simply a task of attending to these auditory dimensions equally; of critical importance is how listeners integrate the multiple dimensions to successfully map them onto particular speech sound categories. This crucial task is exemplified by the difficulties non-native listeners may face in discriminating non-native speech sound contrasts that make use of auditory dimensions in different ways from those in their native languages.

Consider the well-known example of the discrimination of the English /l-r/ contrast (in words such as ‘lead’ and ‘read’) by Japanese learners of English. While both native English listeners and Japanese learners attend to the same auditory dimensions, e.g., the second and third formants (F2 and F3) of /l/ and /r/, they do so in different ways. Iverson *et al*. [1] show that Japanese listeners are most sensitive to variation in F2, but not to variation in F3 which is a more reliable cue for successfully separating the two sounds. English listeners, on the other hand, are most sensitive to variation in F3 and consequently exhibit far greater discrimination accuracy.

The relative preference for certain cues over others, referred to as cue weighting, differs between native and non-native listeners. The sensitivity to these cues develops early in life and is related to the acoustic dimensions found in infants’ ambient language. Unsurprisingly, Japanese does not have an alveolar sound contrast that is differentiated by F3 as in English. Though notoriously difficult, Japanese learners of English can begin to change their relative use of auditory dimensions, or attend to new dimensions, to improve their discrimination accuracy [2].

This review first provides a brief examination of how speech perception develops early in life and offers a theoretical model to account for this. The implications of this early experience are then reviewed in three related areas of non-native speech perception, namely cross-dialect, cross-language and second-language speech perception.

DEVELOPMENT OF SPEECH PERCEPTION

Theoretical accounts of native and non-native speech perception usually have in common that an individual’s early experience with language shapes non-native speech perception and/or second-language learning in adulthood. Some well-known accounts include Kuhl *et al*.’s [3] Native Language Magnet model, Best’s [4] Perceptual Assimilation Model and Flege’s [5] Speech Learning Model.

Speech sounds are produced with great variability, yet individuals learn to identify each instance as belonging to one of a finite group of speech sound categories. Escudero’s [6-8] Linguistic Perception (LP) model advocates that these categories emerge from the mapping of auditory dimensions according to how they are used and integrated in a listener’s language or language variety (e.g., dialect). That is, native listeners of a particular language prefer those auditory dimensions that reliably differentiate the sounds in their speech production for that language, which is referred to as the ‘optimal perception hypothesis’. Therefore listeners with different early experiences of language, and consequently divergent linguistic knowledge, will differ in how they perceive the same auditory events [6-8], as will be described below.

How might infants learn to map auditory dimensions onto speech sound categories? Evidence has repeatedly shown that infants younger than six months can discriminate most speech sounds in any language [9]. As adults, this apparent ability declines and the discrimination of non-native speech sounds becomes more difficult [10, 11, 1] (depending on the particular contrast [12]), while the discrimination of native speech sounds becomes more accurate.

Early theories on phonetic learning posited that infants possess innate phonetic capacities to distinguish all speech sounds and that the apparent decline is due to fine-tuning according to the acoustic dimensions that are relevant for sounds present in their ambient language [13, 14]. However, evidence suggests that early phonetic abilities may not be innate as some animals are capable of discrimination resembling that of humans [15] and infants are also able to discriminate non-speech sounds [16].

Infants' apparent early phonetic abilities may thus be a reflection of general auditory perception and the decline in discrimination performance in adulthood is a result of the development of speech perception, as proposed by the LP model. This perceptual development is formalised within the model as a 'perception grammar' (see [6-8] for fuller explanations of the workings of the model).

Boersma, Escudero and Hayes [17] and Escudero [8] propose that this initial perceptual development is auditory-driven and a potential mechanism is distributional learning. Instances of the same speech sound are produced with great variability on several acoustic dimensions by native speakers, but the most commonly occurring instances are those around the edges rather than the middle of an acoustic continuum. Infants have been shown to be remarkably sensitive to frequency distributions along such continua [18], ultimately leading to successful discrimination of speech sound contrasts.

The development of speech perception occurs due to changes in infants' perception grammar to cope with the distributional properties found in speech in their environment, resulting in several auditory inputs being mapped onto the most frequently perceived categories. In other words, auditory values, e.g., duration, voice onset time, first (F1) and second (F2) formants and so on, will be mapped onto a finite number of phonetic categories. Further perceptual development takes place when a lexicon develops, which within the model is referred to as lexicon-driven perceptual learning.

It is in this way that the LP model accounts for the early experiences with language in shaping speech perception in adulthood. That is, adults' perception grammars have developed for the optimal perception of instances of speech sounds as encountered in their speech environments. Hence adults may encounter difficulty in accurately perceiving non-native speech sounds that use auditory dimensions differently from those to which optimal perception is geared, as in the case of Japanese listeners' discrimination of the English /r-l/ contrast described above. Adults are nevertheless able to learn to discriminate such contrasts more accurately by shifting their relative use of auditory dimensions that are important for distinguishing the two speech sounds [2, 6, 8].

NON-NATIVE SPEECH PERCEPTION

The study of non-native perception typically features listeners who differ with respect to their experience with the non-native language. *Naïve* or *inexperienced (non-native) listeners* are individuals for whom the linguistic variety of the speech signal is unfamiliar, i.e., listeners who have no or very limited experience with it, while *second-language (L2) listeners* or *learners* are individuals who are actively involved in learning a language.

Under the umbrella of non-native speech perception, two related areas of research have emerged, *cross-language* and *second-language (L2) speech perception*. The former generally refers to the processing of the non-native speech signal in terms of one's native language, which has long been held to reveal the auditory dimensions that are (ir)relevant for native listeners as well as changes in infants' early abilities [19]. Studies in the latter area, by contrast, typically investigate adult learners with varying degrees of experience with the L2 [20, 21], including the many factors that can affect learning, e.g., formal language instruction, motivation, length of residence in an L2-speaking country among others [22]. Additionally, cross-language perception with naïve adult listeners can reveal how beginning learners will perceive the sounds of an L2, referred to as the 'initial state' [7, 8].

A further area of non-native speech perception, though perhaps not conventionally grouped within it, is *cross-dialect speech perception*. Below we will see that findings from this area of research are also very relevant to cross-language and L2 speech perception.

Cross-dialect speech perception

The early phonetic ability of infants is apparent in cross-dialect speech perception as infants are able to discriminate between the dialect around them and unfamiliar dialects of the same language [23]. However, younger toddlers find it more difficult to recognise words spoken in an unfamiliar non-native dialect than older toddlers [24], indicating that adaptation to non-native dialects occurs with phonological development at the onset of word learning. An apparent 'bias' toward listeners' native dialect can extend into adulthood. For instance, listeners within and between English-speaking countries may use acoustic cues, such as the F1, F2 and duration of vowels, in slightly different ways to identify the same phonological categories [6, 25]. This is also the case for Spanish and Portuguese speakers from Latin American and Europe [26].

Listeners are able to adapt to unfamiliar dialects, even after limited exposure [27], though phonetic similarity between listeners' and speakers' dialects facilitates adaptation [28, 29]. However, some sound contrasts may be persistently problematic for non-native dialect listeners, especially when a phonologically equivalent contrast does not exist in listeners' native dialect. [30-35].

One example is the lack of the English vowel /ʌ/ in Northern British English dialects where words such as 'book' and 'buck', which contain the phonetically and phonologically distinct /ʊ/ and /ʌ/ vowels in Southern British English dialects, are realised as [ʊ]. This lack of separation between the two phonological categories is mirrored in speech perception: Northern listeners' 'exemplars' of Southern /ʌ/ (based on duration, F1 and F2 values) are very unlike how Southern speakers produce this vowel and resemble /ʊ/ [36].

Northern listeners who have lived in the South of England for an extended period of time are able to shift their exemplar locations so that Southern /ʌ/ exhibits higher F1 frequencies than /ʊ/, indicating a phonologically distinct vowel. Nevertheless, experienced Northern listeners' exemplar locations do not accurately match those of Southern listeners or how the vowel is produced by Southern speakers, meaning Southern /ʌ/ is still problematic for even experienced Northern listeners.

Cross-language speech perception

Recent research in cross-language speech perception has also examined perception of sounds in different varieties of an unfamiliar language. Escudero and Chládková [37] show that Spanish listeners are sensitive to differences in F1 and F2 values of American English and Southern British English vowels, e.g., Southern British English /æ/ is perceived to be more similar to Spanish /a/, whereas American English /æ/, which exhibits lower F1 values than in Southern British English, is perceived to be more like Spanish /e/. This suggests different initial states for Spanish learners of the two English dialects.

Listeners whose native vowel system is much smaller than that of the non-native language are more likely to perceive some non-native vowels as instances of the same native category, often leading to poor discrimination accuracy. However, this also depends on the specific acoustic properties of native vowels. For example, Salento Italian and Peruvian Spanish both exhibit the same five-vowel system, but listeners perceive some Standard Southern British English vowels differently. Escudero *et al*. [38] show that Standard Southern British English /ɒ-ɔː/ are mapped onto a single category /o/ by Salento Italian listeners but onto both /o/ and /u/ by Peruvian Spanish listeners. This suggests greater discrimination accuracy than Salento Italian listeners and different initial states for both groups of listeners [38].

Different initial states may be observed due to listeners' different native dialects. For example, a major difference between the Bohemian and Moravian Czech dialects is the /ɪ-iː/ contrast: Bohemian Czech /iː/ has a lower F1 and is longer than /ɪ/, while only durational differences contrast these two vowels in Moravian Czech. Chládková and Podlipský [39] show that naïve Bohemian Czech listeners perceive the Dutch vowels /i/ and /ɪ/ to be most similar to their Czech /ɪ/ and /iː/ categories, respectively, while Moravian Czech listeners perceive the two Dutch vowels to be most similar mainly to their Czech /ɪ/ category. It is predicted therefore that Moravian Czech listeners' discrimination of the Dutch contrast will be poorer and therefore more difficult for them to learn than for Moravian Czech individuals.

Second-language speech perception

Predictions based on the initial states from cross-language research appear to be borne out in second-language learners.

Firstly, L2 learners' perceptual development depends on how speech sounds are contrasted in their L2 environment, as proposed in the LP model. In speech production, the English /iː-ɪ/ contrast is realised by F1 differences in Scottish English and by F1 and durational differences in Southern British English. Escudero and Boersma [6] presented Spanish learners of English with synthetic stimuli that varied in equal auditory steps along duration, F1 and F2 dimensions (covering the ranges of F1 and F2 of naturally produced Scottish English /iː/ and /ɪ/) and instructed learners to select the English vowel they heard by clicking on a picture representing /ɪ/ or /iː/. Spanish learners who were learning English in the South of England made greater use of duration to perceive the contrast, while those learning English in Scotland tended to use F1 and F2. While Spanish learners of Scottish English used auditory dimensions also relevant for perceiving vowels in Spanish, those learning Southern English made use of a new auditory dimension not used in Spanish, namely duration, demonstrating different learning strategies depending on the dialect being learned.

Secondly, the native dialects of learners may affect how speech sounds and contrasts are learned in a L2. Escudero *et al*. [40] found that Flemish Dutch and North Holland Dutch learners of English exhibit different levels of errors identifying the English vowels /ɛ/ and /æ/ due different confusion patterns arising from different native dialects. Likewise, Escudero and Williams [41] show that native dialect affects Peruvian and Iberian Spanish learners of Dutch on several vowels and contrasts. Indeed, learners' native dialect was generally a better predictor of L2 discrimination accuracy than measures of L2 proficiency. Both of these studies demonstrate that differential early experiences with their native language influenced subsequent L2 learning.

Finally, L2 learners are able to achieve discrimination accuracy comparable to that of native listeners, but the ways in which auditory dimensions are integrated in speech perception may not resemble that of native listeners. For instance, Escudero *et al*. [42] demonstrate that Spanish learners of Dutch can successfully categorise Dutch /aː-ɑ/ tokens as accurately as native Dutch listeners, but their cue-weighting is very different. While both native Dutch listeners and Spanish learners rely on duration, F1 and F2, Spanish learners exploit duration more heavily than spectrum whereas Dutch listeners use spectrum more heavily than duration. Nevertheless, Spanish listeners' category boundary of Dutch /aː-ɑ/ is less clearly defined, suggesting some uncertainty.

Furthermore, cue-weighting in a manner similar to Dutch listeners does not guarantee accurate categorization as Escudero *et al*. [42] report on a group of naïve German listeners who also performed the same task. Like Dutch listeners, they weighted spectrum more heavily than duration, which suggests German individuals will learn the Dutch contrast in a different way from Spanish learners. However, German listeners were less accurate at categorizing Dutch /aː-ɑ/ tokens than native Dutch listeners, suggesting experience with the Dutch language is of course necessary for more accurate discrimination.

CONCLUSIONS

The present review has highlighted the development of native speech perception from infancy into adulthood and its influences on non-native speech perception. Escudero's [7, 8] Linguistic Perception model formalises this development as a perception grammar in which individuals map relevant auditory dimensions onto speech sounds in accordance with the acoustic dimensions used in speech production in their speech environments (referred to as 'optimal perception'). Individuals' differing experiences with language, including varieties of the same language, thus influence which auditory dimensions are used and how these are used in the perception of speech sounds in unfamiliar non-native languages, different dialects and second languages.

REFERENCES

[1] P. Iverson, P. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Ketterman and C. Siebert, "A perceptual interference account of acquisition difficulties for non-native phonemes", *Cognition* **87**, B47-B57 (2003)

[2] P. Iverson, V. Hazan and K. Bannister, "Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults", *Journal of the Acoustical Society of America* **118**, 3267-3278 (2005)

[3] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)", *Philosophical Transactions of the Royal Society B* **363**, 979-1000 (2008)

[4] C. T. Best, "A direct realist view of cross-language speech perception", in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium MD, 1995, pp. 171-204

[5] J. Flege, "Second language speech learning: Theory, findings and problems", in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* York Press, Timonium MD, 1995, pp. 229-273

[6] P. Escudero and P. Boersma, "Bridging the gap between L2 speech perception research and phonological theory", *Studies in Second Language Acquisition* **26**, 551-585 (2004)

[7] P. Escudero, *Linguistic perception and second-language acquisition: Explaining the attainment of optimal phonological categorization*, LOT Dissertation Series 113, Utrecht University 2005.

[8] P. Escudero, "The linguistic perception of SIMILAR L2 sounds", in P. Boersma and S. Hamann (eds.), *Phonology in Perception*, Mouton de Gruyter, Berlin, 2009, pp. 152-190

[9] P. Escudero, "Phoneme discrimination", in P. J. Brooks and V. Kempe (eds.), *Encyclopedia of Language Development*, Sage, Thousand Oaks CA, 2014, pp. 444-447

[10] H. Goto, "Auditory perception by normal Japanese adults of the sounds 'L' and 'R'", *Neuropsychologia* **9**, 317-323 (1971)

[11] J. Werker and R. Tees, "Phonemic and phonetic factors in adult cross-language speech perception", *Journal of the Acoustical Society of America* **75**, 1866-1878 (1984)

[12] C. Best and G. McRoberts, "Infant Perception of non-native contrasts that adults assimilate in different ways", *Language and Speech* **46**, 183-216 (2003)

[13] P. Eimas, "Auditory and phonetic coding of the cues for speech: discrimination of the /r-l/ distinction by young infants", *Perception and Psychophysics* **18**, 341-347 (1975)

[14] A. Liberman and I. Mattingly, "The motor theory of speech perception revised", *Cognition* **21**, 1-36 (1985)

[15] P. Kuhl, "Discrimination of speech by nonhuman animals: basic auditory sensitivities conductive to the perception of speech-sound categories", *Journal of the Acoustical Society of America* **70**, 340-349 (1981)

[16] P. Jusczyk, D. Pisoni, A. Walley and J. Murray, "Discrimination of relative onset time of two-component tones by infants", *Journal of the Acoustical Society of America* **67**, 262-270 (1980)

[17] P. Boersma, P. Escudero and R. Hayes, "Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories", *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 3-9 August 2003, pp. 1013-1016.

[18] J. Maye, J. F. Werker and L. A. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination", *Cognition* **82**, B101-B111 (2002)

[19] J. Werker and C. Lalonde, "Cross-language speech perception: initial capabilities and developmental change", *Developmental Psychology* **24**, 672-683 (1988).

[20] I. MacKay, J. Flege, T. Piske and C. Schirru, "Category restructuring during second-language speech acquisition", *Journal of the Acoustical Society of America* **110**, 516-528 (2001)

[21] J. Flege and I. MacKay, "Perceiving vowels in a second language", *Studies in Second Language Acquisition* **26**, 1-34 (2004)

[22] C. Best and M. Tyler, "Nonnative and second-language speech perception: Commonalities and complementaries", in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*, John Benjamins, Amsterdam, 2007, pp. 13-24

[23] J. Butler, C. Floccia and G. Goslin, "Infants' discrimination of familiar and unfamiliar accents in speech", *Infancy* **16**, 392-417 (2011)

[24] C. Best, M. Tyler, T. Gooding, C. Orlando and C. Quann, "Development of phonological constancy: toddlers' perception of native and Jamaican-accented words", *Psychological Science* **20**, 539- 542 (2009)

[25] A. Ghonim, J. Lim, J. Smith and J. Wolfe, "The division of the perceptual vowel plane for different accents of English and the characteristic separation required to distinguish vowels", *Acoustics Australia* **41**, 160-164 (2013)

[26] K. Chládková and P. Escudero, "Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects", *Journal of the Acoustical Society of America* **131**, EL119-EL125 (2012)

[27] J. Maye, R. N. Aslin and M. K. Tanenhaus, "The weckud wetch of the wast: lexical adaptation to a novel accent", *Cognitive Science* **32**, 543-562 (2008)

[28] M. Sumner and A. Samuel, "The effect of experience on the perception and representation of dialect variants", *Journal of Memory and Language* **60**, 487-501 (2009)

[29] J. Le, C. Best, M. Tyler and C. Kroos, "Effects of non-native dialects on spoken word recognition", *Proceedings of Interspeech 2007*, Antwerp, Belgium, 27-31 August 2007, pp. 1589-1592

[30] C. Clopper, "Effects of naturally-occurring segmental ambiguity on cross-modal priming of minimal pairs", *Journal of the Acoustical Society of America* **129**, 2659 (2011)

[31] S. Dufour, N. Nguyen and U. H. Frauenfelder, "The perception of phonemic contrasts in a non-native dialect", *Journal of the Acoustical Society of America* **121**, EL131-EL136 (2007)

[32] S. Dufour, N. Nguyen and U. H. Frauenfelder, "Does training on a phonemic contrast absent in the listener's dialect influence word recognition?", *Journal of the Acoustical Society of America* **128**, EL43-EL48 (2010)

[33] A. Brunellière, S. Dufour, N. Nguyen and U. H. Frauenfelder, "Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception", *Cognition* **111**, 390-396 (2009)

[34] A. Brunellière, S. Dufour and N. Nguyen, "Regional differences in listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study", *Brain and Language* **117**, 45-51 (2011)

[35] B. Conrey, G. F. Potts and N. A. Niedzielski, "Effects of dialect on merger perception: ERP and behavioural correlates", *Brain and Language* **95**, 435-449 (2005)

[36] B. Evans and P. Iverson, "Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences", *Journal of the Acoustical Society of America* **115**, 352-361 (2004)

[37] P. Escudero and K. Chládková, "Spanish listeners' perception of American and Southern British English vowels: Different initial stages for L2 development", *Journal of the Acoustical Society of America* **128**, EL254-EL260 (2010)

[38] P. Escudero, B. Sisinni and M. Grimaldi, "The effect of vowel inventory and acoustic properties in Salento learners of Southern British English vowels", *Journal of the Acoustical Society of America* **135**, 1577-1584 (2014)

[39] K. Chládková and V. Podlipský, "Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners", *Journal of the Acoustical Society of America* **130**, EL186-EL192 (2011)

[40] P. Escudero, E. Simon and H. Mitterer, "The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception", *Journal of Phonetics* **40**, 280-288 (2012)

[41] P. Escudero and D. Williams, "Native dialect influences second language vowel perception: Peruvian versus Iberian learners of Dutch", *Journal of the Acoustical Society of America* **131**, EL406-EL412 (2012)

[42] P. Escudero, T. Benders and S. Lipski "Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners", *Journal of Phonetics* **37**, 452-465 (2009)

IN THRALL TO THE VOCABULARY

Anne Cutler
The MARCS Institute, University of Western Sydney, Sydney, Australia
a.cutler@uws.edu.au

Vocabularies contain hundreds of thousands of words built from only a handful of phonemes; longer words inevitably tend to contain shorter ones. Recognising speech thus requires distinguishing intended words from accidentally present ones. Acoustic information in speech is used wherever it contributes significantly to this process; but as this review shows, its contribution differs across languages, with the consequences of this including: identical and equivalently present information distinguishing the same phonemes being used in Polish but not in German, or in English but not in Italian; identical stress cues being used in Dutch but not in English; expectations about likely embedding patterns differing across English, French, Japanese.

THE *MESS-IN-THE-MESSAGE* PROBLEM

At St John's Church in Darlinghurst, Sydney, a sign reads "Only God can turn a MESS into a message, a TEST into a testimony, a TRIAL into a triumph, a VICTIM into a victory". Full marks to the writer for effective use of the resources that English offers! But the proposition of exclusivity is, in fact, inaccurate, because turning a *mess* into a *message*, and the like, is what every speaker of English has to do every day to understand speech. The vocabulary gives us no choice in this matter. By the standards of the world's languages, English has many phonemes (42 in the Australian variety [1], where the world cross-language mean is 31 and the mode 25 [2]; Figure 1). If distinctiveness were a priority, 42 is still a tiny number from which to construct the hundreds of thousands of words of the English vocabulary. All vocabularies are like this: a huge number of words, constructed from a trivially small number of contrastive speech sounds. The inevitable result is that there is very little distinctiveness. Words closely resemble other words (*word: bird, curd, herd, weed, wide, wade, work, whirl, worse*...). Further, short words occur accidentally in longer ones (*test* in *testimony*, etc.); and many longer words, not at all related, begin in the same way (*trial* and *triumph*, etc.).

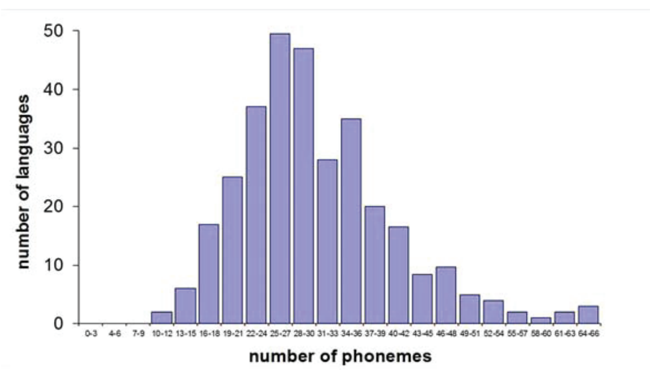


Figure 1. Phoneme inventory size across languages. In a representative sample of world languages, the count (vertical axis) of languages by size of phoneme inventory (in increments of three, the smallest set being 10-12 phonemes, the largest 64-66). The mode is 25; Australian English (42 phonemes) is in the upper tail of the distribution.

Spoken language thus presents more word recognition options than is desirable from the listener's perspective; not only the string of words intended by a speaker, but accidental embedded words within some of those intended words, and beyond that, words embedded across the intended words as well (*victim* in *evict immediately*, *worse* in *were stopping*, etc.). Of course many of these options will completely mismatch the context and can easily be rejected if recognised; but it is rare for listeners to become at all aware of the multiple alternative options, because the efficiency of word recognition is such that the intended string is usually settled upon rapidly and the unintended other options are efficiently discarded. Though many possibilities are briefly available, and indeed compete among one another for recognition, any option that the unfolding speech input mismatches can be immediately discarded.

With experimental methods from the psycholinguistic laboratory it is possible to discern traces of the fleeting presence of rejected competitor words. One such method (eyetracking) offers visual representations of word options. For example, listeners who hear *Now pick up the sandal*.... while looking at a display showing a sandal, a sandwich, and two other objects, typically look to each of the objects with names that begin *sand-* until the moment at which the speech makes clear that it is one and not the other. The point here is that no one waits until a whole word has been presented; the input is assessed continuously for evidence of what it might be. Interestingly, if one of the other objects is, say, a candle (a name differing from the target word only in initial sound), that too attracts looks – not as many as the sandwich, but more than some object with a dissimilar name [3]. (There is more on why this might be useful in section 4 "Test in Testimony Versus Detest" below.)

In another method (cross-modal fragment priming) listeners decide whether a written string of letters is a real word or not, as they listen to some speech. The critical question is how the speech input affects availability of word options, with the availability revealed by how fast a string can be acknowledged as indeed a real word. Typically, response to the same word is compared in three situations: (a) when the spoken input is a completely different word or fragment (this is a baseline for how easy the word is to recognise); (b) when the input is all or part of the same word (this produces priming – the word is heard and seen at the same time, so this should make it highly available); and (c) when the speech partly overlaps with, but mismatches the written word [4].

For example, listeners might respond to DELIVER, while hearing a word fragment ending a neutral sentence such as *The password for this week is...* The baseline control fragment could be *supplem-*, while the matching fragment is *deliv-*. Positive lexical decision responses to DELIVER will be quite a lot faster in the latter case. Then suppose the case (c) fragment is *delish-* (from the word *delicious*). What we see then is that responses to DELIVER are <u>slower</u> than after the baseline fragment (*supplem-*), i.e., the word form *deliver* is <u>less</u> available than even in that neutral situation. Its availability has actually been inhibited. The first few phonemes of the target word made both *deliver* and *delicious* available, but on arrival of the [ʃ] that matched *delicious* but mismatched *deliver*, the mismatched word was rejected. The inhibition of responses to written DELIVER after spoken *delish* reveals auditory *deliver's* temporary, ultimately unsuccessful, presence.

Exactly the same happens when we hear *trial* or *triumph*, *victim* or *victory*; at some point, the input forces us to choose one option over the other. Embedded words such as *mess* in *message* may be fleetingly the strongest available candidate, but they too are eventually overridden by stronger evidence for a longer candidate word. And given the structure of all vocabularies, this type of efficient continuous evaluation of the incoming acoustic signal entails constant resolution of *mess-in-the-message* issues, whenever we hear speech.

SELECTIVE LISTENING: PHONEMES

Speech signals are multidimensional, and listeners can call on several information sources to make effectively the same decision. Decisions involve choice between word options, and hence between speech sounds that crucially distinguish them. Deciding that one is hearing *victim* or *victory* becomes possible when the sixth phoneme turns out to be [m] or [r]. Phonemes are famously not separated entities in speech signals, but overlap, forming the "speech code" [5]. Some sounds can not be reliably identified without knowing what the following sound is, and upcoming sounds may be cued by signals that are present in the speech before the primary speech gestures for the upcoming sound have actually been made. Both types of cue, direct and contextual, are available for listeners to use, and use them they do. The fifth phoneme in *victim* and *victory* is a reduced vowel that will effectively signal already whether an [m] or [r] follows, since both a following consonant's place and manner of articulation (e.g., bilabial and nasal in [m], alveolar and approximant in [r]) have detectable effect in a preceding vowel (similarly, a vowel influences a preceding consonant too).

In some cases, however, available information is just ignored. Whether or not coarticulatory cues are used can be tested by cross-splicing speech signals. Using this method, Harris [6] discovered that listeners use the transitional information from [f] to an immediately following vowel to decide that they are hearing [f]; [f] from *fi* spliced to [i] from *si*, *shi* or *thi* was hard to recognise. This showed that the transitional information was being used to inform decisions. But the results for [s] were quite different: [s] from *si* was as easy to recognize whether the vowel came from *si*, *shi*, *thi* or *fi*. Harris' interpretation invoked the acoustic signal, which gives clearer information for the [s] of *sea* than for the [f] of *fee*. But it turned out (many years later) that the underlying reason was not acoustics, but the vocabulary! The acoustics of [s] and [f] are the same in Dutch, German, and Italian as in English, but Harris' result does not replicate in those languages: Dutch, German, and Italian listeners all ignore transitional information for both [f] and [s] [7, 8]. In contrast, listeners from Madrid (speakers of Castilian Spanish) exhibit the same asymmetry as English-speakers, using transition information for [f] but not for [s].

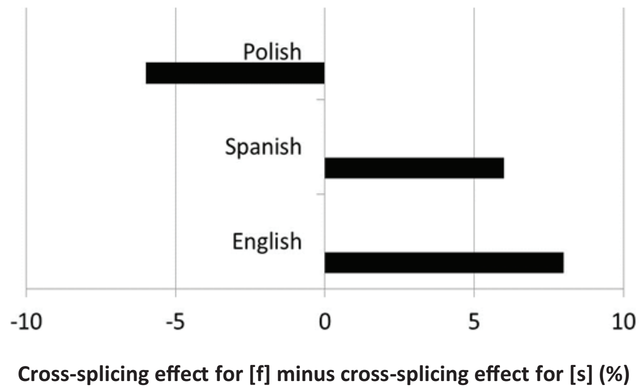


Figure 2. The cross-splicing effect for [f] (how much longer it takes to identify a token of [f] that has been spliced onto a vowel originally uttered after a different sound, vs. after an [f]) minus the same cross-splicing effect for [s], in English and Spanish (showing a positive difference, i.e., a greater effect for [f]) versus Polish (negative, i.e., greater for [s]) [7]. In German and Dutch, the same experiment showed virtually no effect for either sound.

What do speakers of English and of Castilian Spanish have that speakers of Dutch, German and Italian don't? Wagner [7, 8] found the answer: both English and Castilian Spanish have the sound [θ], the dental fricative as in *thick* and *thin* and Castilian *gracias*. This sound is notoriously difficult to distinguish from [f] [9]; so if your vocabulary's phoneme set contains both sounds, it pays to attend to all the information there is in identifying them. Dutch, German and Italian have no spectrally similar fricatives at all, and English and Spanish have no sibilants that could be easily confused with [s]. That is why the attention to transitions occurs, in these experiments, only for [f] and only in English and Spanish. Suppose, though, that your language did have spectrally similar sibilants? Then it would pay to attend to the transitional information for [s]. Polish is such a language, and Polish listeners in Wagner's study [7] indeed showed the reverse pattern to English and Spanish listeners – for them, cross-spliced [s] was harder than cross-spliced [f]. They used the transitional information to identify [s] but not [f] (see Figure 2). This was the crucial sign that the results were due to the language-specific phoneme sets from which the vocabularies were built.

Sorting out which words are really present in a speech stream, and which word forms are only accidentally or partially present, requires listeners to identify as rapidly as possible the speech sounds (phonemes) being uttered. In some languages, pairs of quite similar phonemes are best distinguished by attending to their transition into and out of abutting phonemes. Such transitional information is always there in the acoustic signal, but in some other languages there is no need to take notice of it.

SELECTIVE LISTENING: STRESS

Not only segments make distinctions between words. Some languages call on an extra dimension for this task. In tone languages (e.g., Mandarin), pitch movements in syllables distinguish words. And in lexical-stress languages (e.g., English or Spanish), words can differ in stress alone: *insight* is stressed on the first syllable, *incite* on the second, though the phonemes in each word are the same.

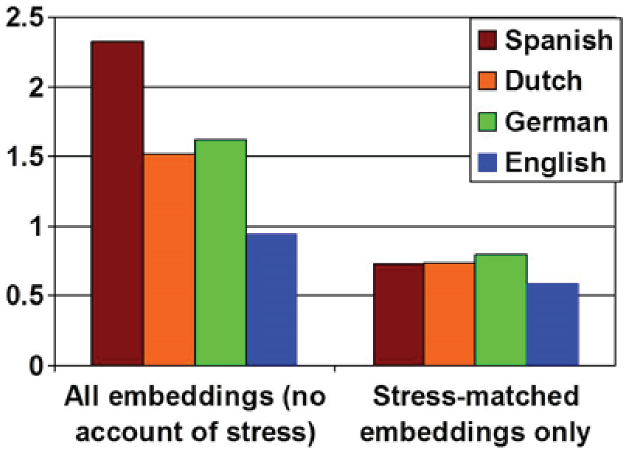


Figure 3. The effect of taking stress into account in computing embedding frequency across the vocabulary, for four languages. If stress is not considered, English *enterprise* contains *enter* and *prize*, and *settee* has *set* and *tea*. If primary stress must match, *enterprise* has only *enter*, and *settee* only *tea*. This stress reduction removes about two-thirds of embedded words from the calculation for Spanish, and about half for Dutch or German. Per word of speech each language then ends up with, on average, less than one embedding. English averaged less than one embedding without considering stress.

Such an extra dimension is particularly useful with smaller phoneme repertoires – e.g., Mandarin (for tone) or Spanish (for stress), each of which have 25 phonemes (world-wide the most common number). The Spanish *mess-in-message* problem (the average number of embeddings in real speech[1]) is reduced by 68.5% if it is computed taking stress pattern as well as segments into consideration [10]; on average 2.32 competitor words per really uttered word if stress pattern is not considered, but only 0.73 competitors per real word once stress is controlled. Obviously, it rewards Spanish listeners to take account of stress as they listen to speech. Indeed they do so. In cross-modal priming experiments in Spanish [4], a mismatching spoken fragment such as *prinCI-* (from *prinCIpio* 'principle'; the second syllable has primary stress) paired with written PRINCIPE (*PRINcipe*, 'prince') inhibited responses in just the same way as a mismatching segment did (e.g., *histe-* from *histeria* 'hysteria' with HISTORIA 'history').

Neither the competitor reduction in Spanish, nor the Spanish cross-modal fragment priming findings, replicate however in English [10, 11]. This is due not only to English's many phonemes, but also to the way English deals with stress. One of the phonemes that Spanish does not have is the reduced and central vowel schwa [ə]. English has it, though, in a majority of unstressed syllables. So, minimal stress pairs (e.g., *insight-incite*) are rare in English. Most English word pairs that are spelled the same way but differ in stress have spectrally different vowels: e.g., the *REcord* vs. to *reCORD*. The stressed first syllable always has a full vowel, but the unstressed first syllable contains schwa.

All listeners must process segments, of course. For listeners to English, just attending to segments gets most stress differences too. There is little more yield from attending also to suprasegmental cues to stress – the differences of duration, amplitude and pitch that Spanish listeners use to distinguish the initial syllables of *principe* versus *principio*. Note that these differences are indeed there in English speech whenever we compare English words with the same syllable contrasting primary versus secondary stress – e.g., the initial syllables of *MUsic/muSEum*, or *ADmiral/admiRAtion*. Table 1 shows acoustic measures from a set of such pairs.

Table 1. Mean acoustic measures (of duration, F0, rms amplitude and spectral tilt) across a female speaker of Australian English's utterance of 21 word pairs with the same first syllable, differently stressed (e.g. *mu-* with primary stress from *MUsic*, vs. *mu-* with secondary stress from *muSEum*) [14].

	Primary	Secondary
duration	381 ms	350 ms
mean F0	208 Hz	180 Hz
max F0	224 Hz	202 Hz
sd F0	12.9 Hz	11.3 Hz
mean rms	641	511
sd rms	229	174
spectral tilt	.909	.202

However, pairs like that are rare in English. So, calculation of competitor reduction with versus without stress information reveals a much smaller reduction in English than in Spanish. Importantly, the average number of competitors per real word is already below one (0.94) even without considering stress [10] (see Figure 3). One would not be surprised to find English listeners failing to attend to the suprasegmental information, and indeed that is exactly what cross-modal studies revealed: there is no significant inhibition for stress-mismatching fragments in English (e.g. [11]; these studies used matched-vowel fragments such as the first two syllables of *ADmiral/admiRAtion*). Figure 4 shows the relative amount of inhibition across languages.

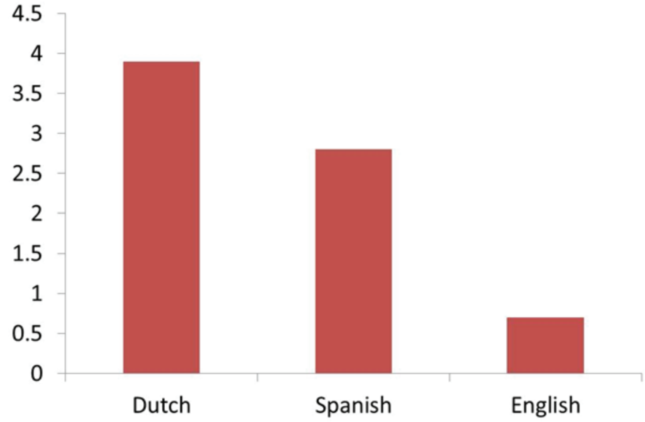


Figure 4. How much inhibition does stress mismatch cause? Across three languages, responses to a (written) word were compared after a stress-mismatching spoken fragment versus after an unrelated control fragment. A mismatch example in Dutch is DOMINANT 'dominant' after hearing *domi*- from *DOminee* 'pastor'; in Spanish, COMEDOR 'dining- room' after *come*- from *coMEdia* 'comedy'; in English, ENTERTAIN after *enter*- from *ENterprise*. In each of these cases the written target word would be stressed on the third syllable (N.B. the effect is equivalent wherever a stress mismatch occurs). The percentage to which responses are slower after stress mismatch than after a control (e.g. *pano*-) is highly significant in Dutch and Spanish, but not in English (where there is not even one percentage point difference).

Stress languages such as Dutch and German, though they are closely related to English with a very similar phonology of stress, differ from English in that they have a much lesser tendency to use the vowel schwa in unstressed syllables. In consequence, considering stress in computing competitor numbers has a larger effect in each of these languages [12], so that listeners benefit (more than English listeners do) from computing stress in recognising words. Dutch listeners, indeed, show the significant inhibition from stress mismatch in the crossmodal task [13], inter alia with primary versus secondary-stress contrasts as in *DOminee* 'pastor' vs. *domiNANT*, 'dominant'.

A useful side effect for listeners of these languages occurs, then, when they use English as a second language. As Table 1 shows, the suprasegmental cues to stress are fully there in English speech, even if English listeners ignore them when recognising words like *admiral* and *admiration*; but Dutch listeners, given English words, do not ignore this information, and actually outdo native English-speakers in correctly assigning syllables differing in stress [11, 14, 15]. Of course, this does not yield them a great benefit in competition reduction, but it may compensate for other ways in which listening to a second language is harder than listening to the native language.

With stress distinctions between words, as with segmental distinctions, acoustic information that is present in speech may or may not be exploited in the recognition of spoken words; the vocabulary dictates how much use is made of it.

TEST IN TESTIMONY VERSUS DETEST

Short words can be embedded anywhere in a longer carrier word (there is a *mess* in *domestic* as well as in *message*, and a *test* not only in *testimony* but also in *intestine* and *detest*). These competitors do not all carry equal weight in the recognition process. One of the first findings to appear in studies of embedding was that there is more embedding at the beginning of carrier words than at the end [16] – and English and Spanish both showed the same pattern in this respect [10]. This asymmetry is important, because a speech signal remains ambiguous as long as words overlap from their onsets. Thus, as noted, onset competitors such as the sandwich (given spoken *sandal*) attract more looks in eyetracking studies than offset competitors such as the candle [3]. Interestingly, though, this pattern appears to be be under listener control; in a little background noise – even when the target word itself is not masked by the noise at all – offset competitors are no longer so disadvantaged [17]. It seems that listeners allow for the possibility that their perception might be impaired in difficult listening conditions, and adjust their expectations about word similarity to deal with this, considering other words ending in the same way as the apparent input as well as words beginning in the same way.

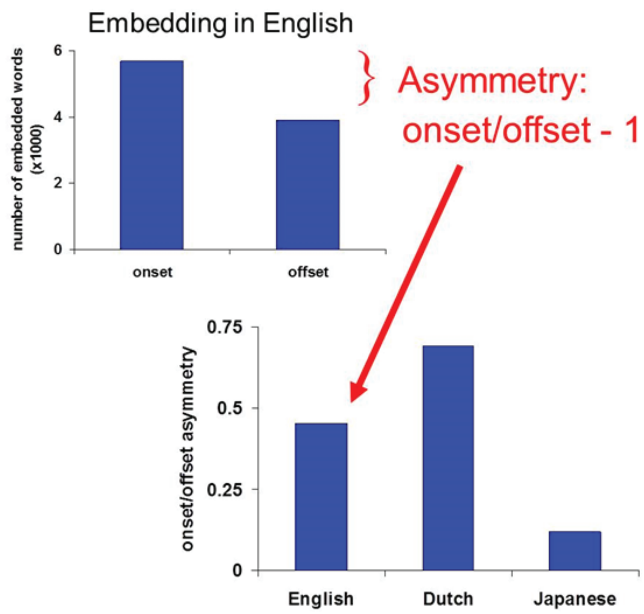


Figure 5. In English, a 1.45:1 ratio holds for the number of embedded words at a carrier word onset (e.g., *enter* in *enterprise*) versus at offset (e.g., *prize* in *enterprise*). The asymmetry (.45) is plotted in the lower graph, where English is compared with Dutch (more asymmetry) and Japanese (virtually no asymmetry).

Because of the importance of competition in understanding speech recognition, asymmetry in the distribution of competitors in the vocabulary will impact upon listening. But once more, not for all listeners equivalently – because this pattern too varies across vocabularies. It is found in English, indeed in all the Germanic languages, and in Spanish; but it is not found in the vocabulary of Japanese. There, due to Japanese word structure, the amount of embedding at carrier word onset and offset is almost the same, as Figure 5 shows [18].

What really causes such asymmetry? As noted, the Germanic languages all have lexical stress, as does Spanish. Does this create asymmetry in embedding patterns? The Germanic languages in particular all exhibit a strong tendency towards initial stress [19]. Combine this with the tendency (strongest of all in English) for unstressed syllables to be weak, and it is clear that syllables at the ends of words are less likely to happen to correspond to other stand-alone words in the vocabulary. (The asymmetry is most marked in German, as German tends to add an unstressed syllable, pronounced as schwa, to words with mono- or disyllabic cognates in English or Dutch: *cat*, *kat*, *Katze* in English, Dutch, German, respectively; *pill*, *pil*, *Pille*; *cigar*, *sigaar*, *Zigarre*; *guitar*, *gitaar*, *Gitarre*; etc.).

Japanese has neither lexical stress nor the vowel schwa. But Japanese also does not have suffixing morphology, which is another prime candidate for source of the embedding asymmetry. Suffixes, either inflections (*talking*, *spotted*, *boxes*) or derivations (*business*, *vibration*, *government*) also tend to be weak syllables that cannot stand alone. All of the stress languages with the embedding asymmetry also have suffixes.

Either stress or suffixes or both could underlie the observed asymmetry. The best way to sort out this question is to find a language which has one of these structural features, but not the other. French is such a language: It has suffixes, but it does not have lexical stress. If only stress is necessary, and sufficient, for an embedding asymmetry, French will pattern like Japanese and show no asymmetry; if suffixing is necessary and sufficient, then French will also show an asymmetry – it should pattern like English. If both factors contribute, we might expect it to fall roughly halfway between the effect found in English and in Japanese respectively.

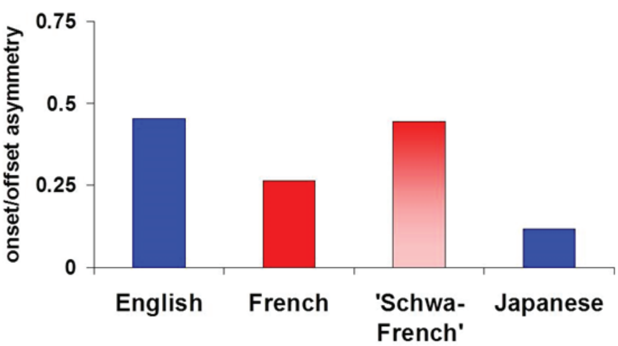


Figure 6. The same onset/offset asymmetry as shown for English and Japanese in Figure 5, now also for French (like English, with suffixes, but like Japanese, without weak final syllables) and for an imaginary variety of French, “schwa- French”, in which weak final syllables containing schwa are added as well.

Using lexical resources for French [20], and computing embedding patterns in its vocabulary as in the other languages, produces a pattern that is indeed halfway between those of English and of Japanese [21]: see Figure 6. Suffixes thus encourage embedding asymmetry. Stress may add further to this effect. The role of stress, we argued, is played out mainly in the placement of weak syllables, i.e., those with schwa. French, though it does not have lexical stress, does have the vowel schwa in its phoneme repertoire. Moreover, it is phonologically legal to expand the role of schwa in a way that would make French more Germanic-like; in certain French dialects, words such as *ville*, *petite*, *bonne*, spelled with a final letter *e* which in standard French is silent, have an extra syllable pronounced as schwa [22]. If all such words in the lexicon of standard French are assumed to be indeed given this extra final schwa in their pronunciation, and the embedding patterns across the vocabulary are then recalculated in the same way, the embedding asymmetry in this artificial variety of French becomes almost exactly that of English [21]. This is again shown in Figure 6.

Thus, both a phonological difference between stressed and unstressed syllables, and the presence of morphological affixes, affect the distribution of embedded words in the vocabulary and hence the patterns of competition affecting speech presented to listeners. English, with both stress and suffixes, has many more words with initial embeddings (*testimony*) than final embeddings (*detest*). This causes patterns of listening that differ from those in languages such as Japanese, which has neither of these precipitating factors, and in consequence, no such embedding asymmetry.

CONCLUSION

Speech acoustics presents a vast array of useful information, but some of it, sometimes, in some languages, gets ignored (even though exactly the same information is put to effective use in other languages). Information is only used where it makes a measurable difference in distinguishing one phoneme from another and hence one word from another, as well as really uttered words from words that are only accidentally present in speech.

Possibly it would be useful to master the exploitation of additional speech cues, for instance when we try to learn, as adults, a language in which those cues are used (Spanish, Dutch or Polish, to use some of the above examples). Research has not yet addressed whether appropriate training could bring this about. Indeed, we do not yet even know when child learners begin to use cues appropriately, nor whether early-bilingual users of two languages that encourage different cue use strategies switch such strategies as they switch language perception. But in general the picture for most of us is that there is nothing much we can do about which cues we use. Our native vocabulary has us in its thrall.

FOOTNOTE

[1] This is computed by calculating the embedded words in every word in the vocabulary (whereby syllable boundary match is required – so, *can* is counted in *candle* but not in *scandal* or *cant*, etc.) and then multiplying every carrier word’s total of embeddings by the carrier word’s frequency in a relevant corpus. This procedure thus accounts for a frequent word (e.g., *dinner*, with embedded *din*) contributing more to average listeners’ experience of lexical competition than an infrequent word (*redintegrate*, with embedded *din*, *great* etc.).

REFERENCES

[1] F. Cox and S. Palethorpe, "Australian English", *Journal of the International Phonetic Association* **37**, 341-350 (2007)

[2] I. Maddieson, I. *Patterns of sounds*, Cambridge University Press, Cambridge 1984

[3] P. Allopenna, J. Magnuson and M. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping model", *Journal of Memory and Language* **38**, 419–439 (1998)

[4] S. Soto-Faraco, N. Sebastián-Gallés and A. Cutler, "Segmental and suprasegmental mismatch in lexical access", *Journal of Memory and Language* **45**, 412–432 (2001)

[5] A.M. Liberman, F.S. Cooper, D.P. Shankweiler and M. Studdert-Kennedy, "Perception of the speech code", *Psychological Review* **74**, 431–461 (1967)

[6] K.S. Harris, "Cues for the discrimination of American English fricatives in spoken syllables", *Language and Speech* **1**, 1–17 (1958)

[7] A. Wagner, M. Ernestus and A. Cutler, "Formant transitions in fricative identification: The role of native fricative inventory", *Journal of the Acoustical Society of America* **120**, 2267–2277 (2006)

[8] A. Wagner, "Cross-language similarities and differences in the uptake of place information", *Journal of the Acoustical Society of America* **133**, 4256 (2013)

[9] G.A. Miller and P. Nicely, "An analysis of perceptual confusions among some English consonants", *Journal of the Acoustical Society of America* **27**, 338-352 (1955)

[10] A. Cutler, D. Norris and N. Sebastián-Gallés, "Phonemic repertoire and similarity within the vocabulary", *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju, Korea, 2004, pp. 65–68

[11] N. Cooper, A. Cutler and R. Wales, "Constraints of lexical stress on lexical access in English: Evidence from native and nonnative listeners", *Language and Speech* **45**, 207–228 (2002)

[12] A. Cutler and D. Pasveer, "Explaining crosslinguistic differences in effects of lexical stress on spoken-word recognition", *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, Germany, 2006, pp. 237–240

[13] W. van Donselaar, M. Koster and A. Cutler, "Exploring the role of lexical stress in lexical recognition", *Quarterly Journal of Experimental Psychology* **58A**, 251–273 (2005)

[14] A. Cutler, R. Wales, N. Cooper and J. Janssen, "Dutch listeners' use of suprasegmental cues to English stress", *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, pp. 1913–1916

[15] A. Cutler, "Greater sensitivity to prosodic goodness in non-native than in native listeners", *Journal of the Acoustical Society of America* **125**, 3522–3525 (2009)

[16] J.M. McQueen, A. Cutler, T. Briscoe and D. Norris, "Models of continuous speech recognition and the contents of the vocabulary", *Language and Cognitive Processes* **10**, 309–331 (1995)

[17] J.M. McQueen and F. Huettig, "Changing only the probability that spoken words will be distorted changes how they are recognized", *Journal of the Acoustical Society of America* **131**, 509-517 (2012)

[18] A. Cutler, T. Otake and L. Bruggeman, "Phonologically determined asymmetries in vocabulary structure across languages", *Journal of the Acoustical Society of America* **132**, EL155-EL160 (2012)

[19] M. Trommelen and W. Zonneveld "Word stress in the West Germanic languages", in *Word prosodic systems in the languages of Europe* ed. H. van der Hulst, (Mouton de Gruyter, Berlin, 1999) pp. 478-515

[20] B. New, C. Pallier, M. Brysbaert and L. Ferrand, "Lexique 2: A new French lexical database", *Behavior Research Methods* **36**, 516-524 (2004)

[21] A. Cutler and L. Bruggeman, "Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry", *Proceedings of INTERSPEECH 2013*, Lyon, France, 2013, pp. 2812-2816

[22] J. Durand, C. Slater and H. Wise, "Observations on schwa in southern French," *Linguistics* **25**, 983-1004 (1987)

ACTIVE LISTENING: SPEECH INTELLIGIBILITY IN NOISY ENVIRONMENTS

Simon Carlile
School of Medical Sciences and The Bosch Institute, University of Sydney, Sydney, Australia
simonc@physiol.usyd.edu.au

Attention plays a central role in the problem of informational masking, a key element of the cocktail party problem, itself described more than 60 years ago. This review considers recent research that has illuminated how attention operates, not only on the auditory objects of perception, but on the processes of grouping and streaming that give rise to those objects. Competition between endogenous and exogenous attention, the acoustic and informational separability of the objects making up an auditory scene and their interaction with the task requirement of the listener all paint a picture of a complex heterarchy of functions.

INTRODUCTION

Here we briefly review some experiments that have contributed to our understanding about listening in noisy environments: The so-called "cocktail party problem" (CPP). Concurrent talkers at a cocktail party will mask each other not only in terms of their energetic overlap but in the way in which a listener can extract meaning from the ensemble of sounds. Central to this process is how the bottom-up mechanisms of grouping segregate the acoustic elements associated with each sound and then how streaming these groups over time contribute to the formation of the auditory objects of our perception. Over the last decade, research has increasingly pointed to the important role of attention in overcoming informational masking. We will consider some of the evidence that attention not only acts on auditory objects but can modulate some of the "primitive" processes of grouping and stream formation. These advances have significantly shifted the focus from "hearing in noise" to listening as an active cognitive process and complement the development of ideas more generally in speech recognition and semantic processing [1].

ENERGETIC AND INFORMATIONAL MASKING IN THE COCKTAIL PARTY PROBLEM

When sounds are occurring concurrently (or even in close temporal proximity) the perception of any one sound can be interfered with by the other sounds. In general, this is referred to as masking and its study has a long history in hearing research. Over the last couple of decades or so, masking has come to be classified as energetic or informational masking. Energetic masking is probably the most well understood (although this understanding is incomplete on a number of levels [2]): When one sound is sufficiently loud that it dominates the output of a processing channel, it will mask a second quieter sound as it is unable to influence the output of the channel. Often termed peripheral masking, this could be conceived of as the motion of the basilar membrane being dominated by a high intensity sound so that the target sound makes no appreciable impact on the output of the cochlea. Psychoacoustically, this sort of phenomenon has been modelled as the output of a critical band energy detector and the ability to predict the presence of a target [3, 4].

Informational masking is most often described as the component of masking that cannot be accounted for by energetic masking. On the one hand this is a simple and parsimonious explanation but on the other, it is not very helpful in understanding the sources of such masking. What has become clearer over the last decade or so is that informational masking can involve interactions at many stages of processing. These include the segregation of spectral components associated with a particular sound, the perceptual grouping and streaming of those components to form an auditory object, spatial and non-spatial attentional control, working memory and other aspects of executive and cognitive functions. The study of informational masking goes back to the mid-1970s although hints as to its effects can be seen in the analysis and discussions of many papers leading up to that time. A splendid and detailed review of the history and early work on informational masking can be found in [5]. That review also considers in detail the work involving multi-tone complexes and the respective roles of target and masker uncertainty in generating informational masking. In this short review we will be more concerned with informational masking as it applies to speech masking and its application to understanding the cocktail party problem.

It has long been recognised that the segregation of a talker of interest from other background talkers is a challenging task. Colin Cherry coined the term the "cocktail party problem" in his seminal paper in 1953 [6]. In a break with the dominant, signal detection based research themes of the time, his paper was focussed on the roles of selective attention in speech understanding, the "statistics of language", voice characteristics, the effects of temporal binaural delays and the costs and time course of switching attention. He makes a very clear distinction between the sorts of perception that are studied using simple stimuli used to study energetic masking and the "acts of recognition and discrimination" that underlie understanding speech at the cocktail party. In this most prescient of papers, Cherry foreshadows much of the work that has now come to dominate research into informational masking and auditory scene analysis as it applies to speech intelligibility in noisy environments. Despite these penetrating insights, most of the work over the last half of the 20th Century continued to be dominated by bottom-up approaches focussed more on energetic masking effects and binaural processes resulting in masking release (see [7, 8] for excellent reviews of much of this work). Notably though, Bronkhorst describes how others had noted that speech interference of speech understanding seemed to amount to more than the algebraic sum of the spectral energy. Indeed, as early as 1969, Carhart and colleagues had referred to this as "perceptual masking" or "cognitive interference" [9].

Right at the turn of the century, Richard Freyman and colleagues reported an experiment that demonstrated that differences in the perceived locations of a target and maskers (as opposed to actual physical differences in location) produced significant unmasking for speech but not for noise [10]. Such a result was not amenable to a simple bottom-up explanation of energetic masking – Freyman appropriated the term "informational masking" and this work led to a large number of studies which have systematically looked at what was driving this speech on speech masking. When the target and the masker both originated from the front of the listener the masking was higher when the masker was speech than when it was noise, particularly at unfavourable signal-to-noise ratios (SNRs) [10]. This indicated that the masker talker was contributing a level of informational masking over and above the energetic masking associated with its SNR. Informational masking was reported to be its highest with two competing talkers and with further increases in the number of the talkers the mixture becomes increasingly more dominated by energetic masking [11]. The exact number of talkers at which masking is maximised probably relates to the speech material and the nature of the talkers but does suggest a relatively small limit on the number of competing streams in informational masking. Similarities between the target voice and the competing talkers was also shown to markedly increase informational masking [12, 13] but again this effect did not increase from 2 to 3 competing maskers when corrected for SNR. Interestingly, listening monaurally to three talkers of the same gender as the target talker (i.e. high similarity) produced less masking than if one of the talkers was a different gender. This "odd sex" distractor effect indicated that the informational masking is not simply mediated by the extent of the similarity between the target and the maskers - a point we will return to later.

Varying the actual locations of target and maskers will result in changes in the relative levels of the targets and maskers in each ear. These level changes result from differences in the interactions of the sounds from each source with the head and pinna of the outer ear. Presumably, an advantage in hearing out the target of interest could simply result from attending to the ear with the most favourable SNR (see Figure 1): so called "better-ear" listening.

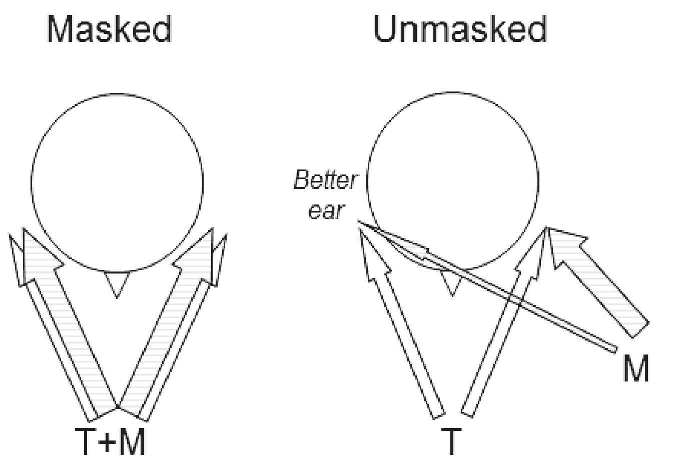


Figure 1. Spatial release from masking can be demonstrated by comparing the speech reception threshold (SRT) obtained with the target and the masker co-located (Masked - usually in front) with the SRT obtained when the masker[s] are moved to another location (Unmasked). The SNR increase at the "better ear" is indicative of how much of the masking release can be attributed to energetic unmasking.

To examine the effects of actual difference in location between target and maskers, Kidd and colleagues [14, 15] compared the speech reception thresholds (SRT) for a target and a masker collocated in front to the SRT obtained with the masker at 90º to the right. They used interleaved frequency channels of modulated noise band speech or band filtered noise to manipulate the energetic interactions between the target and the maskers. In summary, they found that the "better ear" effect could account for around 7 dB of unmasking when the masker was a noise but an average of 18 dB unmasking was found when the masker was another talker. This suggests that the spatial separation of targets and masker provided a much greater advantage for informational compared to energetic masking. In this experiment the modulated noise band speech would have produced target and masker voices that sound quite similar, producing quite high levels of informational masking.

Another strategy employed to "hear out" a talker of interest, particularly against a background of other talkers, is to take advantage of the amplitude modulation of the maskers to "glimpse" the target talker during the intervals where the SNR is favourable. Consonant recognition in modulated noise was found to be well predicted by the proportion of the target "glimpsed" over a -2 dB to +4 dB "local" or instantaneous SNR range [16]. In a clever binaural experiment, Brungart and Iyer [17] presented diotically over headphones the better ear glimpses available with a target in front and symmetrically placed maskers. Such a diotic paradigm maximised the SNR but eliminated the perception of difference in location of the target and maskers. They found that the glimpses, even though they only appeared transiently in one or the other ear, provided a significant unmasking. In that experiment the gender of the maskers and target were different so that the informational masking was relatively low (i.e. the listeners could already easily tell the talkers apart). By contrast, Glyde and colleagues [18] used the speech materials from the LiSN-N speech test [19] where the amount of informational masking could be varied. They found that speech intelligibility was significantly worse in the diotic glimpsing condition when compared with natural binaural listening and the magnitude of the difference was larger when there was more informational masking. Together these results suggest that the perception of the difference in location was adding significantly to the unmasking and that better ear glimpsing was effective mainly for energetic rather than informational masking.

Consistent with the above, a number of experiments have shown that informational masking is not about audibility. An analysis of the sorts of speech identification errors for speech masked by other talkers shows that, more often than not, the errors relate to a word spoken by a masker rather than a guessing error (e.g. [12, 14]). This shows that not only are the maskers audible but they are intelligible and it is their attribution to the target "stream" that is compromised [see also below]. Familiarity with the target talker (i.e. knowing who to listen for) provides an advantage [11, 12] as does knowing where to listen [20] or when to listen [21] (see also [22]) although the same does not appear to be the case for the maskers [23]. Both auditory, and visual cues about "where" and "when" to listen can be very effective, even in the absence of information about the target content [24].

Maskers in a foreign language also produce informational masking, although less so compared to maskers from the listener's native language, or second language in the case of bilinguals [25-27] (but see [28]). Again, this is masking that is over and above that produced by the energetic interactions between the sounds. Informational masking is also still present but somewhat reduced if the speech from the masker talker is reversed in time [28] but this may be complicated by the increased forward masking because of the phonetic structure of the speech used [25]. A most important point however, is that regardless of the extent of masking produced by these maskers, it appears that intelligibility by itself is not a requirement to produce some level of informational masking. This might suggest a quite different process compared to that discussed above where incorrect but intelligible words are attributed to the target talker.

A ROLE OF AUDITORY ATTENTION

Cueing "where" or "what/who" to listen for reduces informational masking indicating an important role for a top-down focus of attention towards the source or voice of interest. In fact, the word confusions discussed above, suggest that informational masking might be due to a failure of attention towards the voice of interest.

In general, attention is thought of as a process of biased competition involving (i) bottom-up (exogenous) attention driven by such qualities as salience and novelty and (ii) top-down (endogenous) attention driven by cognitive steering based on specific task requirements [29]. The ability to focus and report the talker in one ear in Cherry's experiment is a good example of endogenous attention while noticing the change in gender in the unattended ear represents brief exogenous control. The odd-sex distractor effect found by Brungart and colleagues [13] could represent a bottom-up driven change in the focus of attention that then manifests itself as a task related error in reporting the target conversation.

Returning to the original work of Cherry, listening dichotically with a different talker in each ear, largely eliminates energetic masking, or at least the energetic masking attributable to interactions on the basilar membrane. This would provide two highly segregated channels through which the listener can attend. Cherry (1953) reported that while attending to one ear and reporting on the information presented at that ear, the information in the other ear was completely forgotten to the extent that listeners were even unaware that the language of the talker had been changed or that the speech was reversed in time. Interestingly, some statistical properties of the masker talker were sufficiently salient as to be noticed, such as a change in the gender of the talker or the replacement of the talker with a 400 Hz tone. In experiments conducted nearly 50 years later, the listening task was made harder by introducing a masker talker in the same ear as the target [30]. Under these conditions, almost any masker in the other ear was able to produce substantial masking of the target talker. Furthermore, the strength of the contralateral interferer was related to the level of the speech like fluctuations in the spectral envelope - an effect that was thought to engage some form of "preattentive central auditory processing mechanism … that interferes with a listeners ability to segregate speech signals presented in the opposite ear" ([31], p301). There are also a range of listening conditions where the characteristics of the talkers in the unattended ear can direct attention to a target in the attended ear (see for e.g. [32, 33]). This suggests that some level of lexical processing is carried out, even in the absence of attention. These sorts of mechanisms might also explain the masking effects of reversed speech. Such a masker will have the same fluctuations as forward speech but is otherwise unintelligible. This is also consistent with Cherry's observation that significant changes in the unattended speech did come to the attention of the listener, presumably because of some inherent salience. We will return to the issue of what makes speech "speechy" in the context of informational masking later.

GROUPING AND STREAMING SOUND ELEMENTS INTO OBJECTS

Although our understanding of auditory attention is not as mature as say with visual attention, it has been argued that, in line with what is known about visual attention, attention is applied to a perceptual object [34, 35]. For instance, if we perceive an orange rolling across the table, the perceptual elements will include the edges and contours of the shape, the colours, the textures, the motion etc. While these are all encoded and to some extent processed separately, this collection of features are then bound together to form the percept of the object - the orange. In common parlance, an auditory object might be considered to be a particular source of sound - a talker of interest, an instrument in an ensemble or a specific environmental source such as a car. Object formation likely involves an interaction of processes that include the segregation and encoding of the many separate features that we use to distinguish between sounds, as well as an analysis of those features that enables a categorical association between a sound and some meaning [36].

So what are the relevant features for the auditory system? As a sound of interest usually occurs on a background of other sounds, at any point in time, the pattern of stimulation of the inner ear is a multiplexed representation of the sum total of the sounds. Most naturally occurring sounds are spectrally sparse which means that, unless there are very many sounds competing, despite their concurrency, a significant proportion of each sound is on average, not masked by other sounds. So the first challenge for the auditory system is identifying which elements in the pattern of stimulation relate to which sounds - this is the basic problem of auditory scene analysis (see [37] for a foundation work in this area and [38, 39] for relatively recent and quite accessible reviews).

In summary, the auditory system employs mechanisms that exploit the acoustic characteristics of physically sounding bodies to parse out the elements related to different concurrent sounds. For instance, all the acoustic elements that turn on or off at the same time are likely to come from a common source, as are those that are harmonically related or that modulate synchronously in amplitude. These are referred to as acoustic grouping cues, which work on relatively small time frames to segregate and then group the sonic elements of the different sounds [40].

The ability to link together or “stream” these small segments over longer time frames also relies on similar principles of plausibility. For instance sequential groups which come from the same spatial location, have the same or similar fundamental frequencies or spectral profiles or that are changing progressively are all likely to have come from the same source. Such groups are then perceptually linked together into a stream that becomes associated with the auditory object (e.g. [41]). Continuity of voice [12], location [42, 43], prosody and talker characteristics [41], amongst other things, facilitate the streaming of one talker against multiple background talkers. Moreover, over time, continuity also enhances spatial selectivity for a particular target stream [42] indicating that the effects of selective attention appear to build up over seconds. Reverberation has been shown to reduce speech intelligibility and is associated with a degradation in the temporal coding of the fine structure and, to a lesser extent, the envelope interaural time difference (ITD) cues to spatial location [44; 45, 46]. Indeed, the individual differences seen in such listening conditions with normally hearing listeners appear to be related to the fidelity with which the auditory brainstem encodes the periodic temporal structure of sounds [45, 46]. Assuming that the differences in the locations of the target and maskers are the important cues in maintaining the integrity of the target stream, then the fidelity of the location cues must be playing a role in maintaining the spatial continuity supporting attentional selectivity and streaming. For the spatial continuity to be effective, however, the spatial cues must also be encoded and transmitted with sufficient fidelity within the auditory nervous system. This may also provide some clues to the nature of the problems underlying the failure of the hearing impaired listener to solve the cocktail problem where the encoding of fine temporal structure is also compromised.

In the context of the cocktail party scenario, attention also needs to be switched from one talker to another in the course of conversational turn-taking: i.e. there is an intentional break in continuity in order to follow what the next talker is saying. This requires a switch in both non-spatial and spatial attention to a new voice at a new location. Using a dichotic listening paradigm Koch and colleagues [47] found a substantial cost of intentional switching of attention between ears - particularly in the context of reaction time and accuracy of a simple cognitive task applied to the information provided by the target talker. By varying the time between cueing and stimulus, their data also suggests that there is a substantial "inertia" in the auditory attention switching which does not seem to take advantage of early visual cueing for preparation. A recent study [48] examined a group of school age children with significant hearing difficulties in the classroom but no negative audiological findings, auditory processing disorder (APD) diagnosis or other generalised attentional disorder. Using a speeded syllable identification in a variant of a probe-signal task [49], these children were found to have a deficit in attentional reorientation in time. In trials where the target did not occur at the expected time, sensitivity to the target took several seconds to recover, several fold longer than matched controls. This increased inertia in attentional control would have made it very difficult for this group of children to follow the normal turn-taking in the conversation and is consistent with the observations of Koch et al (2011).

When speech streaming breaks down, either as a result of perturbation of the continuity cue or as a result of intentional switching, listeners are likely to attribute words spoken by the masker talkers to the target talker - the classic finding of informational masking. Again, the segregation of the acoustic elements, their perceptual grouping and recognition of the words is not the problem. It is their incorrect attribution to the target talker. Through this lens it is easy to understand how the similarity between concurrent talkers can have such a profound effect on the amount of informational masking. In the early studies discussed above, it was found that spatial separation, or even the perception of a difference in the locations of the target and the maskers, significantly decreased confusion errors (presumably as a result of improved streaming) and thereby produced a significant reduction in informational masking.

The idea that attention is applied to a perceptual object has the important consequence that processing of the object as a whole is in some way enhanced and not just a single feature or features. The advantage of spatial separation was seen even when attention was directed to a non-spatial feature like the timbre of the voice rather than location [50]. In another study [51] subjects were asked to attend to one of two competing speech streams based on their location or pitch. The continuity of the task-irrelevant (non-attended) feature was shown to still influence performance in an obligatory manner.

Recent work has also suggested that object and stream formation is not a simple hierarchical process that provides the objects for selection by attention. The demands of the task performed also have an effect on the formation of objects, particularly where the grouping between various acoustic elements is ambiguous. This is seen particularly in the interactions between the identification of "what" and "where" attributes of an object. A sound's location can be determined unambiguously by the acoustic cues at ears if those cues cover a wide range of frequencies. Grouping should determine which acoustic elements are associated and therefore contribute to the calculation of location and in turn, location is then used to stream the information from a particular source (see [52]). While this idea implies a simple hierarchical processing (grouping then localisation then streaming), experiments using ambiguous sound mixtures suggest more complex interactions. If an acoustic element could contribute to more than one object, the sound mixture is then ambiguous. The contribution a particular element makes to the spectro-temporal content of an object can depend on the focus of attention and the strength of the relative grouping and streaming cues. More surprisingly, however, if an element is not allocated to the attended (foreground) object, it is not necessarily allocated to the background object, that is it gets 'lost' [53]. Likewise, the relative contribution of an ambiguous sound element to the determination of "what" or "where" varies according to the task of the listener (judge "what" or judge "where") and demonstrates considerable individual differences [54]. On one hand, the locations of two sound sources whose components were spectro-temporally intermixed could be reliably estimated on differences in interaural time difference cues to location - that is they can be segregated and localised. On the other hand, spatial segregation and localisation was unable to support identification in the absence of other grouping cues [55].

While many of the experiments discussed above demonstrate the relative strength of the different grouping and streaming cues, in difficult listening situations, the segregation of a target talker is still very challenging. In summary, informational masking could result from (i) ambiguity in the sound mixture and a failure to properly segregate and group the spectral components associated with the target; (ii) disruption of continuity that supports successful sequential streaming of the grouped elements associated with the target of interest; (iii) an error in selecting the appropriate object or stream resulting from high levels of similarity between the cues available for continuity or to (iv) sustain selective attention on the appropriate stream i.e. where saliency in the masker drives the focus of attention away from the target. Of course there are many other factors that also come into play here such as semantic context [27, 56], working memory (e.g. [57-59]), visual cues such as lip reading (e.g. [60, 61]) etc.

The combination of well controlled psychophysical paradigms together with vital imaging (particularly MEG) have driven some spectacular advances in understanding the neural basis of the formation of auditory objects and streams (review: [62]; e.g. [63-66]) and the potential role of temporal coherence in the binding of features (review [67]). Likewise, great strides have been made in understanding the recruitment of auditory spatial and non-spatial attention systems (e.g. [68-70]) and the attentional modulation of activity at different cortical levels [66]. Unfortunately space limits more discussion of these fascinating issues although the interested reader could start with the selection of recent reviews and references above.

CONCLUDING REMARKS

Hearing science and audiology owes much to the fundamental and pioneering work of Harvey Fletcher and colleagues in the early decades of the 20th century. The development of the articulation index (AI) and later the speech transmission index (SII) are founded on the basic assumptions that the intelligibility of speech is related to the instantaneous signal-to-noise ratio within the critical bands up to 6 kHz to 8 kHz. This "bottom-up" approach to understanding speech in noise was very successful in predicting the effectiveness of the telecommunications systems for which it was originally intended. As the attention of researchers turned to more complex maskers such as competing talkers, these energetic masking explanations became less adequate in explaining the extent of masking interactions or the masking release that was afforded by differences between the target and maskers. This informational masking pointed to more complex and cognitive levels of interference that went far beyond the spectro-temporal interactions of the sound energy associated with multiple sources.

This shifted the focus from a single channel problem such as understanding a voice on a telephone line to one of auditory scene analysis – itself a very ill posed problem in a mathematical sense. Research over the last few decades has revealed how the auditory system exploits the physical characteristics of naturally sounding bodies to parse the multiple concurrent sources that most often comprise the listening environment. More critically, this is not just a passive or automatic process but can be influenced by endogenous or "top-down" attentional control. Grouping of the features associated with a sound provides the perceptual object that becomes the focus of attention which in turn is modulated by task requirements and the executive intentions of the listener.

Thus hearing becomes listening – an active process involving a heterarchy of functions feeding forward and feeding backward, weighing the evidence on its saliency, reliability and relevance. Now more than 60 years on, we may finally be within striking distance of Cherry's original goals of understanding the "acts of recognition and discrimination" that enable those critical interactions at the cocktail party.

ACKNOWLEDGEMENTS

Virginia Best, Gavin Lin and Anthony Wakilicz are warmly acknowledged for their careful review and helpful comments on an earlier version of this manuscript.

REFERENCES

[1] Heald S. L. M. & Nusbaum H. C., Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience* **8**:35-35 (2014)

[2] Durlach N. I., et al., Note on informational masking (L). *J Acoust Soc Am* **113**:2984-2987 (2003)

[3] Fletcher H., Auditory patterns. *Reviews of Modern Physics* **12**:0047-0065 (1940)

[4] Green D. M. & Swets J. A., *Signal detection theory and psychophysics* (John Wiley and Sons, New York) (1966)

[5] Kidd G., Jr., Mason C., Richards V., Gallun F., & Durlach N., Informational masking. Auditory perception of sound sources, *Springer Handbook of Auditory Research*, eds Yost W, Popper A, & Fay R (Springer US), Vol 29, pp 143-189 (2008)

[6] Cherry E. C., Some experiments on the recognition of speech with one and two ears. *J Acoust Soc Am* **25**:975-979 (1953)

[7] Bronkhorst A. W., The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica* **86**:117-128 (2000)

[8] Ebata M., Spatial unmasking and attention reated to the cocktail party problem. *Acoustical Science and Technology* **24**:208-219 (2003)

[9] Carhart R., Tillman T. W., & Greetis E. S., Perceptual masking in multiple sound background. *J Acoust Soc Am* **45**:694-& (1969)

[10] Freyman R. L., Helfer K. S., McCall D. D., & Clifton R. K., The role of perceived spatial separation in the unmasking of speech. *J Acoust Soc Am* **106**:3578-3588 (1999)

[11] Freyman R. L., Balakrishnan U., & Helfer K. S., Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am* **115**:2246-2256 (2004)

[12] Brungart D., Simpson B. D., Ericson M., & Scott K., Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am* **110**:2527-2538 (2001)

[13] Brungart D. S., Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am* **109**:1101-1109 (2001)

[14] Arbogast T. L., Mason C. R., & Kidd G.,Jr., The effect of spatial separation on informational and energetic masking of speech. *J Acoust Soc Am* **112**:2086-2098 (2002)

[15] Arbogast T. L., Mason C. R., & Kidd G., Jr., The effect of spatial separation on informational masking of speech in normal-hearing and hearing impaired listeners. *J Acoust Soc Am* **117**:2169-2180 (2005)

[16] Cooke M. P., A glimpsing model of speech perception in noise. *J Acoust Soc Am* **119**:1562-1573 (2006)

[17] Brungart D. S. & Iyer N., Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *J Acoust Soc Am* **132**:2545-2556 (2012)

[18] Glyde H., et al., The effects of better ear glimpsing on spatial release from masking. *J Acoust Soc Am* **134**:2937-2945 (2013)

[19] Cameron S., Dillon H., & Newall P., Development and evaluation of the listening in spatialized noise test. *Ear and Hearing* **27**:30-42 (2006)

[20] Kidd G., Jr., Arbogast T. L., Mason C. R., & Gallun F. J., The advantage of knowing where to listen. *J Acoust Soc Am* **118**:3804-3815 (2005)

[21] Varghese L. A., Ozmeral E. J., Best V., & Shinn-Cunningham B. G., How visual cues for when to listen aid selective auditory attention. *JARO* **13**:359-368 (2012)

[22] Kitterick P. T., Bailey P. J., & Summerfield A. Q., Benefits of knowing who, where, and when in multi-talker listening. *J Acoust Soc Am* **127**:2498-2508 (2010)

[23] Freyman R. L., Helfer K. S., & Balakrishnan U., Variability and uncertainty in masking by competing speech. *J Acoust Soc Am* **121**:1040-1046 (2007)

[24] Best V., Ozmeral E. J., & Shinn-Cunningham B. G., Visually-guided attention enhances target identification in a complex auditory scene. *JARO* **8**:294-304 (2007)

[25] Rhebergen K. S., Versfeld N. J., & Dreschler W. A., Release from informational masking in time reversal of native and non-native interfering speech (L). *J Acoust Soc Am* **118**:1274-1277 (2005)

[26] Cooke M., Lecumberri M. L. G., & Barker J., The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am* **123**:414-427 (2008)

[27] Brouwer S., Van Engen K. J., Calandruccio L., & Bradlow A. R., Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *J Acoust Soc Am* **131**:1449-1464 (2012)

[28] Freyman R. L., Balakrishnan U., & Helfer K. S., Spatial release from informational masking in speech recognition. *J Acoust Soc Am* **109**:2112-2122 (2001)

[29] Knudsen E. I., Fundamental components of attention. *Ann Rev Neurosci* **30**:57-78 (2007)

[30] Brungart D. S. & Simpson B. D., Within-ear and across-ear interference in a cocktail party listening task. *J Acoust Soc Am* **112**:2985-2995 (2002)

[31] Brungart D. S., Simpson B. D., Darwin C. J., Arbogast T. L., & Kidd G., Jr., Across-ear interference from parametrically degraded synthetic speech signals in a dochotic cocktail-party listening task. *J Acoust Soc Am* **117**:292-304 (2005)

[32] Rivenez M., Darwin C. J., & Guillaume A., Processing unattended speech. *J Acoust Soc Am* **119**:4027-4040 (2006)

[33] Rivenez M., Guillaume A., Bourgeon L., & Darwin C. J., Effect of voice characteristics on the attended and unattended processing of two concurrent messages. *Euro J Cog Psych* **20**:967-993 (2008)

[34] Alain C. & Arnott S. R., Selectively attending to auditory objects. *Frontiers in Bioscience* **5**:D202-D212 (2000)

[35] Shinn-Cunningham B. G., Object-based auditory and visual attention. *Trends in Cognitive Sciences* **12**:182-186 (2008)

[36] Griffiths T. D. & Warren J. D., What is an auditory object? *Nat Rev Neurosci* **5**:887-892 (2004)

[37] Bregman A. S., *Auditory scene analysis : the perceptual organization of sound* (MIT Press, Cambridge, Mass) (1990)

[38] Darwin C. J., Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**:1011-1021 (2008)

[39] McDermott J. M., The Cocktail Party Problem. *Current Biology* **19**:R1024-1027 (2009)

[40] Darwin C. J. & Carlyon R. P., Auditory Grouping. *Handbook of Perception and Cognition, Volume 6: Hearing*, ed Moore B (Academic Press, Orlando, Florida), pp 387-424 (1995)

[41] Darwin C. J. & Hukin R. W., Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *J Acoust Soc Am* **107**:970-977 (2000)

[42] Best V., Ozmeral E. J., Kopco N., & Shinn-Cunningham B. G., Object continuity enhances selective auditory attention. *PNAS* **105**:13174-13178 (2008)

[43] Best V., Shinn-Cunningham B. G., Ozmeral E. J., & Kopco N., Exploring the benefit of auditory spatial continuity. *J Acoust Soc Am* **127**:EL258-EL264 (2010)

[44] Ruggles D. & Shinn-Cunningham B., Spatial Selective Auditory Attention in the Presence of Reverberant Energy: Individual Differences in Normal-Hearing Listeners. *JARO* **12**:395-405 (2011)

[45] Ruggles D., Bharadwaj H., & Shinn-Cunningham B. G., Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *PNAS* **108**:15516-15521 (2011)

[46] Ruggles D., Bharadwaj H., & Shinn-Cunningham B. G., Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology CB* **22**:1417-1422 (2012)

[47] Koch I., Lawo V., Fels J., & Vorlaender M., Switching in the cocktail party: Exploring intentional control of auditory selective attention. *J Exp Psych-Human Percept Perf* **37**:1140-1147 (2011)

[48] Dhamani I., Leung J., Carlile S., & Sharma M., Switch attention to listen. *Nature: Sci Reports* **3**:1-8 (2013)

[49] Wright B. A. & Fitzgerald M. B., The time course of attention in a simple auditory detection task. *Percep Psychophys* **66**:508-516 (2004)

[50] Ihlefeld A. & Shinn-Cunningham B. G., Disentangling the effects of spatial cues on selection and formation of auditory objects. *J Acoust Soc Am* **124**:2224-2235 (2008)

[51] Maddox R. K. & Shinn-Cunningham B. G., Influence of Task-Relevant and Task-Irrelevant Feature Continuity on Selective Auditory Attention. *JARO* **13**:119-129 (2012)

[52] Best V., Gallun F. J., Carlile S., & Shinn-Cunningham B. G., Binaural interference and auditory grouping. *J Acoust Soc Am* **121**:1070-1076 (2007)

[53] Shinn-Cunningham B. G., Lee A. K. C., & Oxenham A. J., A sound element gets lost in perceptual competition. *PNAS* **104**:12223-12227 (2007)

[54] Schwartz A. H. & Shinn-Cunningham B. G., Dissociation of perceptual judgments of "what" and "where" in an ambiguous auditory scene. *J Acoust Soc Am* **128**:3041-3051 (2010)

[55] Schwartz A., McDermott J. H., & Shinn-Cunningham B., Spatial cues alone produce inaccurate sound segregation: The effect of interaural time differences. *J Acoust Soc Am* **132**:357-368 (2012)

[56] Iyer N., Brungart D., & Simpson B. D., Effects of target-masker contextual simillarity on the multi-masker penalty in a three talker diotic listening task. *J Acoust Soc Am* **128**:2998-3010 (2010)

[57] Rudner M., Ronnberg J., & Lunner T., Working memory supports listening in noise for persons with hearing impairment. *J Am Acad Audiol* **22**:156-167 (2011)

[58] Rudner M., Lunner T., Behrens T., Thoren E. S., & Ronnberg J., Working memory capacity may influence perceived effort during aided Speech Recognition in Noise. *J Am Acad Audiol* **23**:577-589 (2012)

[59] Ronnberg N., Rudner M., Lunner T., & Stenfelt S., Assessing listening effort by measuring short-term memory storage and processing of speech in noise. *Speech, Language and Hearing In Press* (2014)

[60] Helfer K. S. & Freyman R. L., The role of visual speech cues in reducing energetic and informational masking. *J Acoust Soc Am* **117**:842-849 (2005)

[61] Mishra S., Lunner T., Stenfelt S., Rönnberg J., & Rudner M., Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers in Systems Neuroscience* **7** (2013)

[62] Shamma S. A. & Micheyl C., Behind the scenes of auditory perception. *Current Opinion in Neurobiology* **20**:361-366 (2010)

[63] Du Y., et al., Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cerebral Cortex* **21**:698-707 (2011)

[64] Ding N. & Simon J. Z., Emergence of neural encoding of auditory objects while listening to competing speakers. *PNAS* **109**:11854-11859 (2012)

[65] Ding N. & Simon J. Z., Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *J. Neurosci.* **33**:5728-5735 (2013)

[66] Golumbic E. M. Z., et al., Mechanisms underlying selective neuronal tracking of attended speech at a "Cocktail Party". *Neuron* **77**:980-991 (2013)

[67] Shamma S. A., Elhilali M., & Micheyl C., Temporal coherence and attention in auditory scene analysis. *Trends Neurosci* **34**:114-123 (2011)

[68] Lee A. K. C., et al., Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience* **6**:190-190 (2012)

[69] Larson E. & Lee A. K. C., The cortical dynamics underlying effective switching of auditory spatial attention. *Neuroimage* **64**:365-370 (2013)

[70] Larson E. & Lee A. K. C., Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *Neuroimage* **84**:681-687 (2014)

AUDITORY GRAMMAR

Yoshitaka Nakajima[1], Takayuki Sasaki[2], Kazuo Ueda[1], and Gerard B. Remijn[1]
[1]Department of Human Science/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka 815-8540, Japan
[2]Department of Psychological and Behavioral Science, Miyagi Gakuin Women's University, Sendai 981-8557, Japan
nakajima@design.kyushu-u.ac.jp

Auditory streams are considered basic units of auditory percepts, and an auditory stream is a concatenation of auditory events and silences. In our recent book, we proposed a theoretical framework in which auditory units equal to or smaller than auditory events, i.e., auditory subevents, are integrated linearly to form auditory streams. A simple grammar, Auditory Grammar, was introduced to avoid nonsense chains of subevents, e.g., a silence succeeded immediately by an offset (a termination); a silence represents a state without a sound, and to put an offset, i.e., the end of a sound, immediately after that should be prohibited as ungrammatical. By assuming a few gestalt principles including the proximity principle and this grammar, we are able to interpret or reinterpret some auditory phenomena from a unified viewpoint, such as the gap transfer illusion, the split-off phenomenon, the auditory continuity effect, and perceptual extraction of a melody in a very reverberant room.

INTRODUCTION

If we try to record and write down what someone says in an everyday conversation, we are very likely to be embarrassed by the fact that the speech includes a fair amount of doubtful parts in terms of grammar. Probably, in our mind we often correct what we hear in an everyday situation according to a grammatical framework, which needs to be shared with the social group to which we belong. Because acoustic information disappears immediately after it is released, often in a noisy environment, the auditory system simply needs a robust framework to connect given pieces of information in a coherent manner. If so, however, the auditory system may need such a framework in order to organize any auditory percept in our everyday life—for example, to hear out footsteps, approaching cars, cats' meows, winds, sound signals of electronic devices, and so on. It is indeed an astonishing capacity of the human auditory system to separate each sound perceptually when mixtures of many different sounds are given to both ears simply as temporal changes of sound pressure [1].

Our research over the past years proceeded from the hypothesis that our auditory system utilizes a kind of grammatical system which is innate to humans, and that this system is a basis of all specific grammars of human languages. So far, the hypothesis is still very primitive, but it helped us to understand and discover new auditory illusions. It seems to have a path to be connected to the phonologies of English, Japanese, or Chinese, and seems to explain partially how notes in Western music are perceived. Some neurophysiological phenomena can be related to this human innate grammar. Thus we called this the *Auditory Grammar*, abbreviated as AG from here on, and wrote a book in Japanese to sum up what is known in relation to this paradigm [2]. An outline of this book is described in this article.

The concept of gestalt quality appeared at the end of the 19th century to explain the fact that one can perceive the same melody in different keys, e.g., in C major and in F# major, even if no common notes are used in two different presentations, e.g., "C D E C | C D E C" and "F#G#A#F# | F#G#A#F#" [3] (Figure 1). Something that cannot be reduced to the natures of individual tones should be there, and this was called the gestalt quality. This was an immediate precursor of gestalt psychology, which appeared as a rather quiet scientific revolution claiming that the whole is not the sum of its parts (e.g., Koffka [4]). The contemporary leading researchers in auditory psychology were not interested in this idea, and rather established a theoretical framework in which auditory phenomena were interpreted as if they had been phenomena observed in an electric circuit [5]. To be fair, this paradigm worked very well for decades to make auditory research a very rigorous and precise field [6].

However, a few related fields could not afford neglecting gestalt psychology. In the field of speech perception, two phenomena, i.e., the cocktail party effect [7] and the auditory continuity effect [8] were reported. The former is a common phenomenon in our everyday life. When two or more people speak different things simultaneously, we are able to perceive that more than one speaker utters different things, and follow one of the speakers to grasp the spoken content. The latter is now a well-known auditory illusion: a speech or music signal, a tone, or a band noise of which a short portion, typically a small fraction of a second, is replaced with an intervening noise can be perceived as continuous, although that portion is missing. In order for this illusion to occur, the noise to fill the missing part should basically cover the frequency range of the original signal with a surpassing intensity. Another important phenomenon related to gestalt psychology has been reported with some interest in music [9]. If two pure tones of 100 ms alternate between 1000 and 1050 Hz, we are likely to hear a single pitch-fluctuating tone as if we hear a trill in music. If the tone frequencies are 1000 and 2100 Hz instead, we are likely to hear two separate streams of different pitches. The latter phenomenon is called *auditory stream segregation* today [1]. Auditory stream segregation is often understood employing a gestalt concept called the *proximity principle*: objects or events that are close to each other tend to be integrated perceptually. Deutsch [10] systematically indicated that those gestalt principles established to understand visual organization in the first half of the 20th century, to which the *similarity principle* and the *common fate principle* are also included, work well to understand auditory organisation especially in music perception.

Figure 1. An example showing the concept of gestalt quality [3]. Even the transposition to the remotest key does not prevent the listener from hearing the same melody, although none of the notes are shared between the two tone sequences.

AUDITORY UNITS

In visual perception, figures and a ground are often formed perceptually to let our visual world make sense. For example, the letters on this page are figures, and they are supported by a ground throughout the page, which includes the parts covered by the letters. The ground does not have a clear shape. The figure-ground idea is often applied to auditory organization—for example, a melody and an accompaniment are sometimes considered as a figure and a ground. We do not take this view because both the melody and the accompaniment have clear shapes. Basically, no parts of the accompaniment are covered and hidden by the melody, and we can even pay attention only to the accompaniment for a long time. We rather assume that the auditory world consists of auditory streams that are concatenations of auditory events and silences. This is not a revolutionary way of thinking, but we just formalized what leading researchers assumed on auditory organization [1, 11, 12]. Auditory events are what we often call sounds in our everyday life: footsteps, hand claps, music notes, or speech syllables, of which we can count the number. An auditory stream is what we hear coherently, often as belonging to the same source, in time, which is a string of auditory events and silences. Auditory events and auditory streams are perceptual units comprising the auditory world.

In order to formalize AG, we took one further step to assume auditory elements that can be smaller than auditory events, called *auditory subevents*. AG describes how such elements are concatenated to form auditory events and streams. Besides gestalt principles, our ideas about auditory subevents seem to have taken shape along with approaches in neuroscientific research on how the human brain deals with incoming sound. It has long been described in various neuroscientific studies that sound edges, i.e. a sound's onset and offset, are signaled at very early stages of cortical processing by cells that only respond to sound edges. These edge-specific neurons are typically different from those that signal the sustained parts of sound, indicating that the brain considers sound edges and sustained parts as different subevents. Current research addresses not only how sound edges are represented in the brain, but also how the information of sound edges and the information of sustained parts of sound are combined and expressed at the level of cortical responses over time. For example, the auditory continuity effect as introduced above and described in more detail below, has been studied specifically to investigate such neural-response integration [13, 14]. Conceptually, AG also proceeds from the integration of sound parts that together constitute auditory events and auditory streams. This enabled us to study the integration of auditory subevents at the behavioral (psychophysical) level by means of creating new auditory phenomena, such as those described below. In the future, similar to the auditory continuity effect, these new sound stimuli can hopefully be subjected to and contribute to neuroscientific research as well.

GAP TRANSFER ILLUSION

An auditory illusion was the starting point to construct AG (Figure 2). Suppose that a frequency glide component of 2500 ms moving from 420.4 to 2378.4 Hz and another glide component of 500 ms moving from 1189.2 to 840.9 Hz cross each other while sharing their temporal middle. This pattern is typically perceived as a long ascending glide and a short descending glide crossing each other—just as how the stimulus pattern was made. These glides physically cross at 1000 Hz, but "crossing" in the present context means that there are ascending and descending glides which share the same pitch region. If a short temporal gap of about 100 ms was introduced onto the middle of the short glide, we hear what we presented—a long ascending glide and two successive short tones. However, if a short temporal gap was introduced onto the middle of the long instead of the short glide, we still hear a long continuously ascending glide and two successive short tones. This is the *gap transfer illusion*, which gave us a chance to think about AG [15].

The long glide with a temporal gap is described by the following scheme:

```
abcdefghijklmnop
<  =  > <  =  >/,
```

where the alphabetic letters indicate temporal positions roughly, and "<" means an onset, "=" a filling, ">" an offset (a termination), and "/" a silence. The short glide is added as follows:

```
abcdefghijklmnop
<  =  > <  =  >/
    <  =  >/     .
```

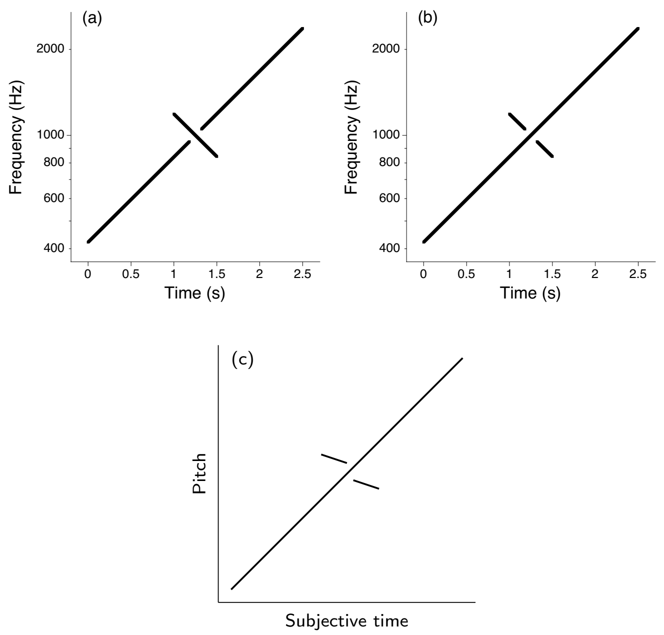


Figure 2. The gap transfer illusion. (a) A typical stimulus pattern inducing the illusion, (b) a stimulus pattern perceived as it is, and (c) the common percept.

Because the onset of the short glide (at the letter "e") and the offset of the first part of the long glide (at "g") are close to each other in time and frequency, the proximity principle works to integrate them perceptually to make an auditory event. The onset of the second part of the long glide (at "i") and the offset of the short glide (at "k") are also close to each other, and again the proximity principle works to integrate them. Thus, we obtain:

```
abcdefghijklmnop
    <=> <=>     ,
```

and two auditory streams are formed as follows:

```
abcdefghijklmnop
<       =      >/
    <=>/<=>/    .
```

It is also possible that the silences (/) in the second line are detected at the beginning, but this does not change the final results. The potentially separate pieces of fillings are integrated as a single filling in the above stream, and this also helps to make the whole pattern grammatical. This shows how a grammatical form of an auditory percept appears from acoustic cues, which are not always grammatical.

AUDITORY GRAMMAR

AG is a grammar indicating how auditory subevents, i.e., onsets (<), offsets (>), fillings (=), and silences (/), are concatenated to form an auditory stream. We first assumed that an auditory stream always begins with an onset, and ends with a silence, and, then formalized a grammar as follows:

1. An onset is followed by a filling or a silence (<= or </).
2. An offset is followed by a silence (>/).
3. A filling is followed by an offset or an onset (=> or =<).
4. A silence is followed by an onset, or ends a stream (/< or /).

This set of rules may be insufficient for future research, but we first summarized what is known empirically, and did not include unnecessary rules for this purpose. New rules may be included in the future.

Three different types of auditory events appear as follows:
1. An event that begins and ends immediately as a single hand clap (<)—followed by a silence (/).
2. An event that begins, continues for a while, and ends as a train whistle (<=>)—followed by a silence (/).
3. An event that begins, continues for a while (<=), and is replaced by another event—starting with an onset (<)—as a music note in a melody.

The gap transfer illusion as described above is considered an auditory phenomenon to construct auditory events of the second type (<=>). If a filling and an offset appear *without* a preceding onset (=>), then AG requires an onset to be restored (<=>; Sasaki et al. [16]).

SPLIT-OFF PHENOMENON

A new illusion was discovered from this theoretical framework. Suppose that a long glide of 1200 ms moves from 420.4 to 965.9 Hz—the first part of the above long glide interrupted by a gap. The beginning part of the second glide is the same as that of the above short glide, but the glide is lengthened—another glide of 1500 ms moves from 1189.2 to 420.4 Hz. These two glides are presented successively with an overlap of 200 ms (Figure 3). This is a new example for this article, but the basic idea was from our previous research [15, 17]. We can hear a long continuous glide going up and down. At about the temporal middle of this ascending-descending glide, we hear a short tone. The acoustic cues of this pattern are:

```
abcdefghijkl
<  =  >
    <  =  >/.
```

The proximity principle works between the onset of the second glide (at "e") and the offset of the first glide (at "g"), and the following streams are obtained:

```
abcdefghijkl
<    =    >/
    <=>/    .
```

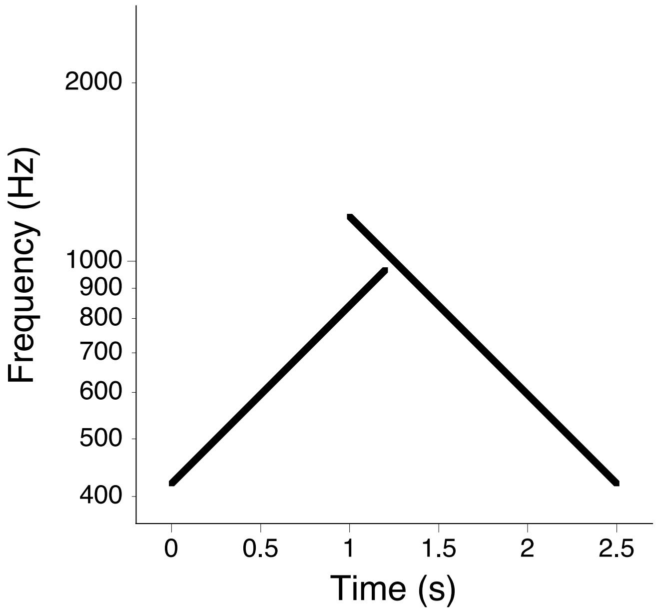


Figure 3. A stimulus pattern inducing the split-off phenomenon. An illusory short tone is perceived around the temporal middle of the pattern.

This is what we call the *split-off phenomenon*. Although we invoked one of the gestalt principles, the proximity principle, our explanation of this phenomenon also revealed a problem of gestalt psychology. The acoustic cues as indicated at first seem to take a very simple shape—two glides with a short overlap. It is difficult from a gestalt-psychological viewpoint to understand why the perceptual system should reconstruct this configuration. There is no reason to cause the split-off phenomenon in order to meet the *Prägnanz* law, which indicates a tendency of our perceptual system to seek for simplicity and regularity [4]. This led us to assume that the proximity principle and AG should be the basis to understand this phenomenon. The proximity between auditory subevents may have high priority in the process of auditory organization, and the proximity principle should not be chained to the classic version of gestalt psychology. AG, though still primitive, is justified by the fact that it gives us opportunities to discover new auditory phenomena. The split-off phenomenon can be observed in a very simple situation which could have been realized in the middle of the 20th century, but it seems that previous researchers did not have an occasion to generate a pattern leading to this illusion.

AUDITORY CONTINUITY EFFECT

If a long pure tone of 2000 ms and 2000 Hz has a temporal gap of 100 ms in the middle, and if the gap is filled with a narrow-band noise around 2000 Hz sufficiently more intense than the pure tone, then we often hear the pure tone not with a temporal gap but as continuous (Figure 4). This is an example of the *auditory continuity effect* [1, 8, 12]. This illusion is often explained by the peripheral behavior of the auditory system, but it seems worthwhile to indicate that this illusion can also be explained within our framework—AG should be always counted as one of the possible explanations.

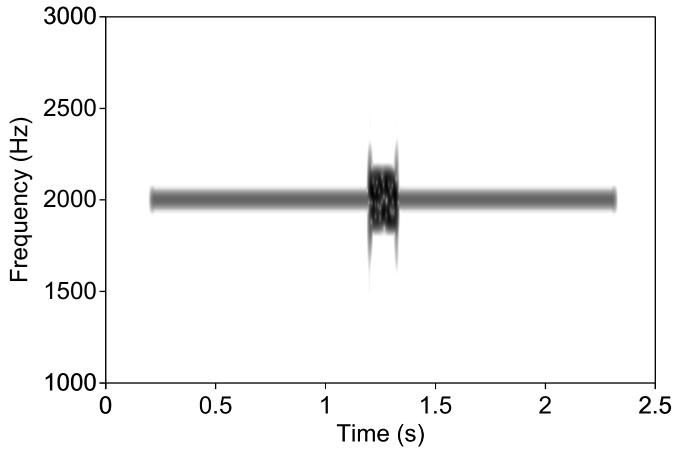


Figure 4. A spectrogram of a stimulus pattern in which the auditory continuity effect can be observed. A long continuous pure tone is perceived, even when the middle portion of 100 ms is replaced by a narrow-band noise of the same duration.

Because the intense noise should mask the offset and the onset of the pure tone portions delimiting the temporal gap, the following subevent cues are given to the auditory system:

```
abcdefghijkl
< = <=> = >/ .
```

This configuration is ungrammatical. The first offset (at "`g`") should be followed by a silence (`/`), but is not. Then the configuration can be reconstructed, and a silence can be inserted (at "`h`") as follows:

```
abcdefghijkl
< =     = >/
    <=>/    .
```

This is closer to a grammatical solution. Because the two fillings, or filling portions, in the upper stream (at "`c`" and "`i`") are from the same pure tone, they can be united as a single filling. Thus, we obtain:

```
abcdefghijkl
<    =    >/
    <=>/    .
```

This shows that the auditory continuity effect can be understood also in the framework of AG [17–19].

PERCEPTUAL EXTRACTION OF A MELODY

Finally, we would like to point out an auditory phenomenon, which we should be encountering often in our everyday life, especially in music. Suppose that a pure tone of 1047 Hz (C in music) and 1200 ms and another pure tone of 988 Hz (B in music) and 1000 ms are presented in such a way that they share an offset (Figure 5). We can hear two tones with different pitches, just as how the tones are presented, with asynchronous onsets:

```
abcdefghij
<    =  >
         /
  <  =  > .
```

However, it is also possible to hear out a sequence of two successive notes C and B:

```
abcdefghij
<=<  =  >/
C-B------
   (roughness).
```

The auditory system probably tries to find a coherent stream to make the percept as simple as possible. B is perceived with some roughness, which means that C (to begin at “a”) and B (to begin at “c”) have a perceptual interaction, but the presence of C is suppressed perceptually when B starts. A very simple melody “CB—” thus appears. If we played this melody “CB—” in an extremely reverberant room with a recorder, for example, a very similar acoustic pattern would appear, and to perceive a melody in this way should be *correct* in this case.

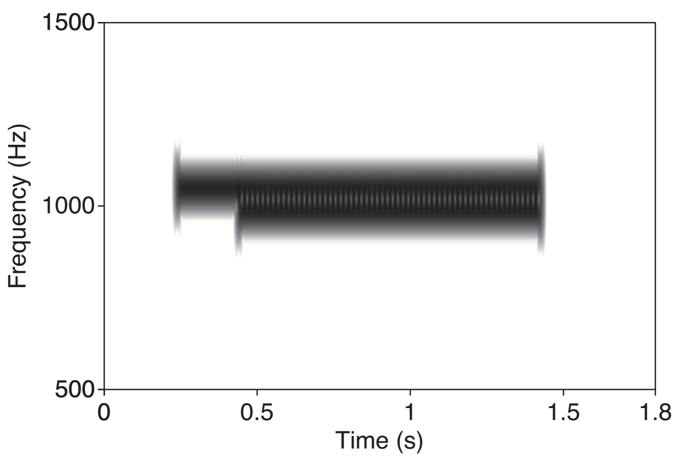


Figure 5. A subjectively constructed melody. Two pure tones of 1047 Hz (C in music) and 988 Hz (B in music) are presented with a 200-ms onset asynchrony. It is possible to hear a melody of C followed by B.

FINAL REMARKS

Some gestalt principles and AG can work together well to discover new auditory phenomena and to reinterpret our auditory experience. We invite people in many different fields to play with these toys to find something new themselves.

ACKNOWLEDGMENTS

Gert ten Hoopen worked with us to develop this theoretical framework. We are grateful to Kate Stevens for this opportunity to present part of our ideas in English. This study is supported by a Grant-in-Aid (25242002 to YN in FYs 2013-2017) from the Japan Society for the Promotion of Science.

REFERENCES

[1] A.S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambridge, MA, 1990

[2] Y. Nakajima, T. Sasaki, K. Ueda, and G.B. Remijn, *Auditory Grammar*, Corona Publishing, Tokyo, 2014 (in Japanese, with a CD-ROM)

[3] C.v. Ehrenfels, “Über ‘Gestaltqualitäten’”, *Vierteljahrsschrift für wissenshaftliche Philosophie*, **14**, 242-92 (1890)

[4] K. Koffka, *Principles of Gestalt Psychology*, Routledge and Kegan Paul, London, 1935

[5] S.S. Stevens and H. Davis, *Hearing: Its Psychology and Physiology*, Wiley and Sons, New York, 1938

[6] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd edition, Springer, Berlin, 2007

[7] E.C. Cherry, “Some experiments on the recognition of speech, with one and with two ears”, *Journal of the Acoustical Society of America* **25**, 975-979 (1953)

[8] G.A. Miller and J.C.R. Licklider, “The intelligibility of interrupted speech”, *Journal of the Acoustical Society of America*, **22**, 167-173, (1950)

[9] G.A. Miller and G.A. Heise, “The trill threshold”, *Journal of the Acoustical Society of America*, **22**, 637-638 (1950)

[10] D. Deutsch, “Grouping mechanisms in music”, in Deutsch, D. ed. *The Psychology of Music*, 3rd edition, Academic Press, Amsterdam, pp. 183-248 (2013)

[11] L.P.A.S. van Noorden, “Temporal coherence in the perception of tone sequences”, Unpublished doctoral thesis, Technical University, Eindhoven, 1975

[12] R.M. Warren, *Auditory Perception: An Analysis and Synthesis*, 3rd edition, Cambridge University Press, Cambridge, 2008

[13] C.I. Petkov, K.N. O’Connor, and M.L. Sutter, “Encoding of illusory continuity in primary auditory cortex”, *Neuron*, **54**, 153-165 (2007)

[14] L. Riecke, F. Esposito, B. Bonte, and E. Formisano, “Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex”, *Neuron*, **64**, 550-561 (2009)

[15] Y. Nakajima, T. Sasaki, K. Kanafuka, A. Miyamoto, G. Remijn, and G. ten Hoopen, “Illusory recouplings of onsets and terminations of glide tone components”, *Perception & Psychophysics*, **62**, 1413-1425 (2000)

[16] T. Sasaki, Y. Nakajima, G. ten Hoopen, E. van Buuringen, B. Massier, T. Kojo, T. Kuroda, and K. Ueda, “Time-stretching: Illusory lengthening of filled auditory durations”, *Attention, Perception, & Psychophysics*, **72**, 1404-1421 (2010)

[17] G.B. Remijn and Y. Nakajima, “The perceptual integration of auditory stimulus edges: An illusory short tone in stimulus patterns consisting of two partly overlapping glides”, *Journal of Experimental Psychology: Human Perception and Performance*, **31**, 183-192 (2005)

[18] T. Kuroda, Y. Nakajima, S. Tsunashima, and T. Yasutake, “Effects of spectra and sound pressure levels on the occurrence of the gap transfer illusion”, *Perception*, **38**, 411-428 (2009)

[19] T. Kuroda, Y. Nakajima, and S. Eguchi, “Illusory continuity without sufficient sound energy to fill a temporal gap: Examples of crossing glide tones”, *Journal of Experimental Psychology: Human Perception and Performance*, **38**, 1254-1267 (2012)

Note: The audio files for the auditory demonstrations mentioned in this paper are available via the Acoustics Australia website for downloading this paper.

LOW FREQUENCY SPATIALIZATION IN ELECTRO-ACOUSTIC MUSIC AND PERFORMANCE: COMPOSITION MEETS PERCEPTION

Roger T. Dean
austraLYSIS, Sydney, Australia
MARCS Institute, University of Western Sydney, Sydney, Australia
roger.dean@uws.edu.au

The article takes the perspectives of an electro-acoustic musician and an auditory psychologist to consider detection of localization and movement of low frequency sounds in reverberant performance environments. The considerable literature on low frequency localization perception in free field, non-reverberant environments is contrasted with the sparser work on reverberant spaces. A difference of opinion about reverberant environments has developed between on the one hand, audio engineers and many musicians (broadly believing that low frequency localization capacities are essentially negligible), and on the other, psychoacousticians (broadly believing those capacities are limited but significant). An exploratory auditory psychology experiment is presented which supports the view that detection of both localization and movement in low frequency sounds in ecological performance studio conditions is good. This supports the growing enthusiasm of electro-acoustic musicians for sound performance using several sub-woofers.

INTRODUCTION

Psychoacoustics now generally considers that there are three main mechanisms which allow humans to localize sounds, and detect movements of sonic sources (reviewed: [1-3]). The first two constitute the 'duplex' of inter-aural time differences (ITD) and inter-aural intensity differences (IID) distinguishing a single sound as heard at the two ears (see below for elaboration). The duplex theory was developed by Strutt (Lord Rayleigh) in the late 19th and early 20th century [4], following early observations on localizing the sound of a flute by Venturi in the 18th century. These two mechanisms are binaural: that is they require interaction between the auditory pathways flowing from both ears, and may exploit neural coincidence detection [5]. The third mechanism is monaural: the so-called spectral notch effect, whereby head-related transformations of the spectrum of an incoming sound, due mainly to the pinnae of the ears, provide location cues. All the mechanisms are influenced by head size and related anatomical features. The IID mechanism is largely restricted to frequencies above about 1000 Hz; the monaural mechanism operates in a fairly similar range. Only the ITD mechanism is usefully functional at lower frequency. Most psychoacoustic studies of sound localization have used free field conditions, that is conditions in which reflections and reverberation are virtually lacking, such as in an anechoic chamber or virtual environments in headphones. Those with primary involvement in auditory perception might note that this usage of the term free field is that of acoustics, and may not coincide with their normal usage.

None of these mechanisms of human sound localization are excellent when faced with low frequency sounds, and relatively few data specifically address frequencies below 250 Hz. Nevertheless, as I will describe, several fields of music exploit such very low frequencies substantially: for example music with or comprising of drones, noise music, and electroacoustic music at large. For this oeuvre, 250 Hz is hardly conceived as 'low frequency', rather more like mid-frequencies. Furthermore musical uses of frequencies below 250 Hz require presentation in reverberant environments, such as the performance studio, dance floor, or concert hall, and in moderately reverberant environments such as recording studios; conditions which contrast with those of the vast majority of psychoacoustic studies.

It is worth pointing out the distinctions between perceiving localization and lateralization. Biologically speaking, the important feature of a sound is the location of its source. However, in audio engineering and in electro-acoustic music in particular, more important is the degree of movement of the sounds and their components, expressed as their changing lateralization, that is their apparent positional spread in the listening space, whether virtual (headphones) or physical. One experimental approach to making this distinction is to have listeners represent the centre of a sound source on a one-dimensional scale from L to R [6], another to have them draw a graphical range of spread. Note the implication here that listeners may be perfectly aware of the disposition of the loudspeakers, yet perceive sounds as disposed almost anywhere in the listening space.

Thus the purpose of this article is to discuss low frequency perception in reverberant environments of ecological relevance to music, as an aspect of applied acoustics of particular importance to music composition, production and performance. I point to some of the gaps in the literature in relation to these applications, suggest and illustrate some useful analytical and experimental approaches, and also seek to provide pointers that may eventually be useful to electroacoustic composers and improvisers in constructing the spectral design of their music.

A SYNOPTIC REVIEW OF LOW FREQUENCY LATERALIZATION AND ITS IMPACTS

Sound spatialization in music performance and recording environments.

The sound of an orchestra is spatialized as a consequence of the disposition of the instruments in the performing space. For example, the low pitched instruments such as tuba (brass) and double bass (strings) are usually to the back, and at one or other side of the layout, while the high pitched flutes and oboes (woodwind) and violins (strings) are to the centre and further forwards, in relation to the position occupied by the conductor. Prior to the era of orchestral music, there were also notable compositional experiments with spatial dialogues between groups of instruments, as in the works of Gabrieli for brass chorale. Subsequent to the 19th/20th century dominance of the orchestra in western music, electroacoustic music since 1950 has enlarged this emphasis on spatialization to extremes, where hundreds of loudspeakers may be arrayed around a performing/listening space in a 3D organization, so that sounds can be projected from any point, or throughout, and can be 'moved' around [7, 8]. Composers have also sometimes had the opportunity to create grand architectural spaces (some temporary, some mobile) for such performances, for example Stockhausen and Xenakis [9], and in Australia, Worrall [10]. To electroacoustic music, even if presented in relatively humble stereo or quad, timbral flux around the listening space is a key feature.

Thus it is a matter of concern that we lack comprehensive data and mechanistic understanding of sound localization in some parts of the frequency spectrum: notably low frequencies. This is so much the case that particularly amongst studio and sound projection audio engineers there is a repeatedly argued view that low frequencies are poorly localized. The argument consequently suggests there is little point in having multiple subwoofers (the specialized speakers which most fully represent frequencies below about 200 Hz) in a performance space, unless the purpose is solely enhancing the loudness of those low frequencies. Furthermore, it is relevant that electroacoustic projection does not provide such detailed visual cues to sound lateralization as an orchestra provides, so the difficulty with low frequencies is not reduced by such cues. In neatly presenting the somewhat opposed views of psychoacousticians and audio engineers, a spatialization team from the University of Derby [11] present a concise review discussion. They conclude that on balance it is to be expected that in most environments low frequency localization is much better than the audio engineer community admits, yet concede there are major gaps in the data and our understanding in relation to ecologically relevant musical environments, which are mostly reverberant.

Electroacoustic composers have continued apace to exploit sound spatialization, relying mainly on their compositional intuitions and perceptual impressions as to what is or is not audible. Empirical evidence suggests that perception of low frequency lateralization may be masked to some degree by the presence of broad band energy in higher frequencies [12]. But other work emphasizes the positive impact of low frequency sound on auditory image size, and the additional benefit of stereo (or multiple) subwoofer conditions for perception of overall image size [13, 14]. Stereo subwoofers were thought to be distinct from mono pairs, with a limited number of popular music tunes, but the subjective preference experiments which followed were not statistically conclusive [15, 16]. Nevertheless, multispeaker systems, vector based amplitude panning and ambisonics have been used in attempts to enlarge the 3D impact of sound projections, and to foster impressions of envelopment [17] and engulfment, the latter seemingly related to the degree to which 3D impressions do superimpose on 2D [18, 19]. The multi-speaker system at ZKM Karlsruhe, neatly called Klangdom, has been developed alongside corresponding sound diffusion software, called Zirkonium [8]. This allows a composer or sound designer to specify the location of a sound, as a combination of outputs from three adjacent speakers, and allows a compositional dissociation between audio track and speaker. In contrast virtually all other software assumes that an individual audio track is sent to 1 or 2 speakers (according to whether the track is mono or stereo); some software, such as Ardour, MAXMSP and ProTools does permit a single audio stream to be sent to any combination of speakers, but they do not allow specification in terms of localization.

The Zirkonium system, and a few others can also well represent one of the most notable developments in spatializing electroacoustic music: what Canadian composer Robert Normandeau calls timbral spatialization [8]. This is effectively the systematic presentation of different frequency bands from different parts of the performance space, and their movement around the space. Convenient and accessible software, Kenaxis, based on MAXMSP, allows easy approaches to this technique even in live performance.

Such developments in electroacoustic music further emphasize the importance of understanding our localization, lateralization and movement detection abilities in relation to every band of the audible frequency spectrum.

The importance of low frequencies in electroacoustic and other music: relations to environmental and speech sounds

I mention here a few aspects of the increasing but long-standing musical importance of low frequency sound. Certain ritualistic musics, such as some of Tibet, and of the Australian didgeridoo, use repeated low frequency timbres, in a manner akin to the drone in Indian music (which is actually a wide band frequency pattern with strong bass), and to the specialized forms of shamanic [20] or trance-music which use drones. A drone in this music is a long-sustained low band sound, often changing very slowly or not at all.

The evolution of popular music through jazz and rock, via amplification has resulted since the 1960s in much higher sound levels in the bass instruments: as for example, the acoustic contrabass has often been heavily amplified or replaced by the electric bass or by bass motives played on a synthesizer. The bass 'riff', or accompanying repetitive rhythmic-melodic form of much music in rock and jazz has consequently been able to take a much more foregrounded position in the overall sound mass, a position which has been enhanced by mixing technologies. A similar but less obvious phenomenon took place in certain forms of contemporary classical instrumental and electroacoustic music and in dance electronica, where more visceral sensations [21] were created than generally sought before (with notable exceptions such as the medieval composer Marais). Thus it is interesting to compare Xenakis' influential orchestral piece *Metastaseis* [22] with, for example, an Elgar symphony in respect of the contribution of low frequencies to the overall acoustic intensity: there are sections in *Metastaseis* in which even the score shows clearly the dominant sweep of instrumental glissandi (pitch slides) in the lower string instruments, something rare in previous music. Similarly, Xenakis's electroacoustic piece *Bohor*, a work of considerable power, acoustically and affectively, has sections in which transitions in the low frequency intensities are the dominant feature, requiring high quality reproduction for full appreciation.

Electroacoustic music points to the possible relation between low frequency timbres in music, speech and environmental sound. Speech (and voice) is relevant because it is a common component of electroacoustic music, often heard both raw and digitally transformed [23-25]. One interesting aspect of this is that across speakers and conditions, the median frequency of the speech FO (the fundamental pitch) is around 130 and 220 Hz for men and women respectively [26, 27], well below the range of most psychoacoustic experiments on other sounds, and well into the 'low range' in any conception. Environmental sound is relevant partly because of the importance of overt environmental sounds in soundscape and other aspects of electroacoustic composition, but also because some of the physical associations of low frequency sounds in the environment (with large mass, low position, slow movement) can provide important metaphorical cues in the music.

Musical tension, often created by controlling the degree to which expectation is fulfilled [28], can then exploit countermanding cues, such as rapid movement of some sonic objects at the same time as a lack of movement of low frequency objects. Conversely, electroacoustic music, and noise music in particular, often focus on movement of low frequency sounds, again raising fundamental questions for the psychoacoustics of reverberant environments. Noise music is a large genre, springing from the work of Xenakis but extending from classical acousmatic composition to underground rock (and the edges of techno and drum 'n' bass). In the core of noise music, high intensity sounds of at most slowly varying complex timbres (high spectral flatness, poorly defined spectral centroid) are used. The timbres often start close to white (or sometimes pink) noise, and sculpt them slowly. Whereas in most previous music, melodic or more generally motivic structure has most often been delivered largely in the high frequency bands, this is no longer true in noise music, nor in many other aspects of electroacoustic music [29, 30].

All these observations on the current usage of low frequencies in music emphasise that understanding the perception of spatialization and movement of timbres comprising frequencies below 250 Hz is needed for composers to most fully and powerfully exploit it. Hence this is a worthwhile topic in applied acoustics, but as yet has not attracted the attention it deserves. I turn next to a brief summary of what is known about this, with particular reference to environments which are ecologically apt for music: in other words, reverberant rather than free field (non-reflective) environments.

Some relevant psychology and psychoacoustics: perception of low frequencies in reverberant environments

It is worthwhile to ask what psychoacousticians treat as 'low frequencies' and why; and to contrast that with even conventional compositional perspectives, let alone electroacoustic ones. There is a fairly clear lower limit to the frequencies in which the IID aspect of the duplex theory provides significant information: around 1000 Hz. It seems this may have driven the psychoacoustician (and perhaps acousticians and audio engineers) to treat frequencies below 1000 Hz as 'low', and hence rarely to venture below 250 Hz (see for example [31, 32]). In contrast, the frequency at the centre of a piano keyboard is the note called 'middle C', and it has a fundamental frequency of only about 260 Hz. This note also appears right in the middle of the two staves which notate two part tonal music. So for a classical composer, using acoustic instruments and notation to make 'pitch-based' music, low pitches are those at least an octave below middle C, in other words, below about 130 Hz; around the pitch referenced as the common male speech fundamental. Electroacoustic composers are often influenced by perceptions of, and maybe experience in playing acoustic instruments, so they share the conception of low pitch being around 100 Hz, even when they make 'sound-based' music, in which pitch may not be apparent and is certainly of limited importance [33]. One of the clearest and fullest studies in support of the idea that localization is possible at such low pitches [6], though dealing with headphones rather than reverberant spaces, indicates sensitivity with narrow band sounds down to 31.5 Hz, but with quite lengthy static stimuli (800 ms).

I provide some pointers to frame our further comments on perception of sonic movement, and perception of both location and movement in reverberant environments. The above-mentioned early and influential experiments of Lord Rayleigh 'on the lawn' (and sometimes with the participation of 'Lady Rayleigh') involved a tuning fork of low frequency (128 Hz), and did also involve speech. They lead to the duplex theory, and it was interesting that in the earliest papers (e.g. 1876) Rayleigh considered the tuning fork a 'pure' sound, but by 1907, when he reviewed [4] his overall work in this area of sound localization, he emphasized that it is a more complex sound, and brings to bear mechanisms at many frequencies. This timbre has been studied in detail subsequently, usually consisting of at least three harmonic components and several side-components, over a wide frequency range [34]. The duplex theory has largely survived empirical testing, as summarized in the two recent general review articles I reference [2, 3]. In depth discussion is provided in some of the empirical studies such as those of Middlebrooks and collaborators. They summarise the situation as follows : 'the duplex theory does serve as a useful description (if not a principled explanation for) the relative potency of ITD and ILD [which is what I term IID] cues in low- and high-frequency regimes (p. 2233) [35].

The main factors which create ITD and IID are clear: the geometry of the head in relation to the sound source. What is perhaps less obvious is why frequency should impact on the potency of the ITD and IID cues: but it seems that the diffusion of energy around the head and ears is such as to annul most low frequency IID cues, whereas the ITD remains. A given time delay represents a smaller proportion of the long cycle times of low frequencies, and this may provide more discrimination than with high frequencies. Monaural localization depends primarily on the influence of the pinnae of the outer ear on the sound transmission, creating transformed power spectra often with notches in mid to high frequencies, which can provide location cues that can be learnt even with only one ear (see [3, 5]). Neural pathways and possible mechanisms for localization have been investigated [5, 36] and computationally modelled [37].

All three localization mechanisms can be influenced by head movement, which consequently is an advantageous feature of listener behaviour [38], perhaps particularly for low frequencies [39]. Many experiments have restricted head movement, so as to control this influence. Of course, audio engineers, and music creators and listeners are used to taking advantage of it and hence experiments of ecological relevance to them do not restrict it (as in the exploratory experiment below). In favourable free-field conditions, the minimum audible angle (MAA) for localization is about 1° for broad band noise and the minimum audible movement angle (MAMA) is much higher [2]. Data suggest that for a 500 Hz tone MAMA is about 8° at velocity 90 degrees per second, and >20° at 360 degrees per second [40]. There is also evidence that a minimum stimulus duration between 150-300 ms may be required for motion detection [41].

There are also possible non-auditory components of low frequency localization, involving the vestibular system, or vibrotactile information, perhaps registered on the face, nose or other body parts [21, 42]. Of ecological relevance here is that many noise artists, used to very high sound intensities, have learnt to conserve their hearing by the use of ear plugs: for them, such non-auditory components of input sounds may be of even greater importance. The author, like many in electroacoustic music or electronic dance music, has experienced disconcertingly extreme intensity sound in some underground sound clubs, and used earplugs both there and when performing noise music. At some venues, earplugs are always given out at the entry. On the lower scale of acoustic intensity, orchestral wood wind players, usually seated just in front of the brass section, commonly have protective screens behind them in rehearsal and performance: unlike the screens used in recording studios, their purpose is simply reducing the sound impact on the woodwind players. The issue of earplugs and sonic localization is worthy of in depth experimental study.

Finally, I briefly summarize some of the known impacts of reverberation (i.e. rooms, or some partially sequestered spaces such as valleys). Many of the relevant studies use simulated reverberation, and have not been fully corroborated in experiments in real environments. The importance of reverberation diminishes when listeners are less than about 1 metre from the sound source, or when they are substantially eccentric to the speakers and/or the space. While speech perception is more difficult in reverberant environments, in some respects musical listening may be enhanced, and there is an important industry concerned with the design of environments and architectures for events, concerts, domestic and studio listening and living (e.g. [43]). Reverberation facilitates distance judgements concerning sound sources [44], but it generally diminishes other aspects of localization discrimination (reviewed [45, 46]). An important ongoing series of studies on localization of sound in rooms is being undertaken by Hartmann (e.g. [47, 48]). From this series, one salient observation is that the utility of the ITD declines below 800 Hz, though few data exist concerning frequencies below 200 Hz: sensitivity (in terms of ITD threshold time) is about twice as good at 800 as 200 Hz [49]. Some ITD-related theories of localization fail in this low frequency range [50]. Low frequency IIDs are even further reduced in reverberant environments, nevertheless, low frequencies may still be highly weighted in resultant localization judgements [46, 51].

Two illusions are particularly relevant to low frequencies and to reverberant environments. Precedence effects occur when two sound sources are separated in time, and determine whether the sounds undergo fusion: the opening sound may dominate the localization, preventing the realistic fission between the two. The duration of the sound is important in this, and in free field (particularly, anechoic) environments, the crucial transitions generally occur between 1 and 20 ms. However, this is much less clear for low frequency sound and for reverberant environments (see review [52]). The Franssen illusion is related: in this even a gradual transition of acoustic energy from one loudspeaker to another can be missed, and localization be determined by the opening onset location [2, 3, 53].

EXPERIMENT

An exploratory experiment on low frequency localization and movement detection in reverberant environments

Bearing the preceding discussion in mind, I conducted an exploratory experiment on this topic. It investigated whether and how fast a low frequency sound can be perceived to move in a performance studio environment (with quite short reverberation time). It complemented this with measures of location of sounds, using a four-alternative forced choice approach. I hypothesized that at frequencies below 200 Hz, localization to L or R of a listener would remain feasible in a reverberant environment, even for filtered noise. It was expected that accuracy would be higher as duration of the sound increased 0.2 s<2 s<6 s, the latter two durations chosen to correspond to feasible compositional durations for spatial movement. Similarly, I hypothesized successful detection of sonic motion at the latter two times, but to be poor or non-existent at 0.2 s. Note that the forced choice design does not aim to distinguish between localization and lateralization, as described above.

Our experimental procedure and studio environment is described more fully in the Appendix, and the legend to Figure 1, but the essence was as follows. I used two Genelec

7060B subwoofers as sound sources, each at 45° to, equidistant from, and at the same horizontal level as our seated listeners, who were musically trained people. There were five participants. Sounds were of three kinds, each low-pass filtered to reduce the presence of higher frequencies, and synthesized within MAXMSP software: sine tones of 30 and 60 Hz, and white noise. These were presented either for 0.2, 2 or 6 s, with 10ms on/off ramps, with the SPL (dBA) at the listeners' head set unequally on the basis of readily acceptable loudness for the three sounds. During presentation the sound originated

Table 1. Mixed effects analysis of accuracy, across all five participants.
Optimised Model: accuracy ~ duration * location + duration * sound + trial + (1 | participant). This means duration, location, sound and the interactions duration*location and duration*sound are the fixed effect predictors, and random effects for the intercepts by participant are required.

Predictor			
	Sequential ANOVA of the model		
	Degrees of Freedom (DF)	Sum of Squares	Mean Sum of Squares (by DF), which is also the F-value
Duration (sonds)	1	218.51	218.51
Location/Movement	3	221.55	73.85
Sound	2	25.16	12.58
Trial number (centred, rescaled)	1	3.11	3.11
Duration x Location	3	54.61	18.20
Duration x sound	2	36.37	18.18
	Random Effects parameter and Confidence Intervals for the Fixed effect coefficients		
Random effects	S.D.		
Intercept by participant	0.57		
Fixed Effects	2.5% Confidence Interval	97.5% Confidence Interval	
Intercept	2.41	4.36	
COEFFICIENTS:			
Duration	-0.10	+0.34	
Location: moving LR	-6.16	-4.41	
Location: moving RL	-5.59	-3.85	
Location: R	-1.70	+0.39	
Sound: 30Hz sine	0.87	1.73	
Sound: 60Hz sine	1.40	2.27	
Trial number	0.0004	0.15	
Duration x Location (moving LR)	0.51	0.95	
Duration x Location (moving RL)	0.42	0.85	
Duration x Location R	-0.10	+0.44	
Duration x Sound (30Hz)	-0.47	-0.15	
Duration x Sound (60Hz)	-0.64	-0.32	

either solely in the Left (L) or Right (R) speaker, or 'moved' with a linear constant power transit L->R or R->L over the whole sound duration. Listeners were required after the end of the sound (and not before) to indicate as quickly as possible which of the four categories of location/movement event they had just heard. There were in total 3(duration) x 3(sound source) x 4 (location/movement) = 36 different stimuli. These were presented in randomized order, each stimulus 8 times in each of two blocks, with the listener requested to take a break between the two. Thus each participant responded to 576 stimuli. This achieved the commonly used 'roving' of sound sources in terms of intensity and frequency, intended to minimize listeners recognition of specific colouring attached to individual speakers or positions in the space.

A mixed effects analysis of the data (using glmer from the lme4 package in R since the data are binomial) to model the accuracy of responses is summarized in Table 1. There were random effects for the intercept by participant; accounting for this reduces the likelihood of Type 1 errors in the analysis (e.g. [54]). The sequential ANOVA of the glmer model suggests that the main explanatory power is provided by duration, sound location/movement and their interaction; though it is important to note that the exact values in such a sequential ANOVA depend somewhat on the order in which the parameters are entered. The confidence intervals (which also reveal the mean coefficient, as the centre of the range) show that the effect of Duration is largely carried in the interaction with location (Duration itself is not effective). The moving sounds are both much worse identified than the static ones, which are not different from each other (as shown by the fact that the CI on the coefficient for R, which is referenced against L as base, breaches zero). The two sine tones are both better located than the noise sound (which is the base level, not shown in the table), though this modest effect is reduced as the sound duration increases: but note that they were not matched in dB at the listening position as I did not seek to understand the influence of timbre. There is a small improvement as the experiment proceeds (reflected in the positive coefficient on Trial number, which was centred and rescaled before analysis).

The modelling was guided by Bayesian Information Criteria values, coefficient significance and parsimony, and optimized by likelihood ratio tests. The optimized model had a BIC of 1814.74. Confidence intervals for the fixed effects parameters were determined by lme4's 'confint' function, using the 'profile' technique. The more approximate 'Wald' technique gave very similar values. The random effects are modelled as a distribution whose standard deviation is measured, but are not of primary interest here. The fixed effects are expressed as coefficients and confidence intervals (which are here symmetrical), and where a predictor has several categories (sound, location) or distinct continuous values (duration), the coefficients are the difference from the 'base' level, which is the level which does not appear in the table. The model was worsened by treating the location as comprising fixed vs moving, hence this approach is not shown.

Figure 1 summarizes the salient comparisons, as judged by the mixed effects analysis, and shows that for the static sounds, accuracy was extremely high, but it was much worse for the moving sounds, increasing with their duration. It also shows that participants had difficulty in locating the very short sounds, and this was almost entirely due to the short moving sounds (the interaction shown in the figure as Moves/0.2 seconds). The difference in performance for the different sounds (which were in any case intended as roving stimuli rather than maximally controlled, and so not shown) was very small.

The 0.2 s movement sounds had movement rates of 450 degrees/sond, beyond the rate at which MAMAs are optimal in free field conditions (discussed above). The inability to judge movement in these stimuli was thus expected. Correspondingly, Table 2 shows an aggregated contingency table for the responses in relation to the stimulus location/movement. It shows that the moving sounds created confusion, where generally the starting point of the movement was taken to be its static location when the participant failed to recognize the movement. This effect is similar to the Franssen illusion already described.

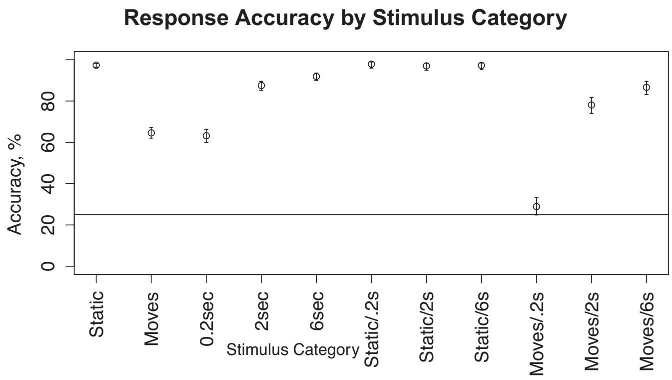


Figure 1. Summarized accuracies in detection of localization and movement in the various conditions. The graph shows percent hit rate with 95% confidence intervals, based as conventionally on pooling all five participants' data. Thus the categories overlap in what is shown: the combinations Static/Moves; the three durations; and the 6 duration/location interactions; each include all the data (2804 responses). All individual results (except Moves/.2s) are highly significant at $p < 0.00001$ in comparison with the chance rate (25%, shown as a horizontal line). A full analysis by mixed effects modelling is provided above. The text and Appendix describe the conditions in more detail. Confidence intervals were also determined by a more correct statistical meta-analysis of the separate confidence intervals determined on each (independent) participant for the stimulus categories shown, and these were a little larger, but confirmed all the conclusions.

Table 2: Contingency table of percentages of responses to each location/movement category of stimulus. Numbers on the diagonals are the correct response percentages.

Location/	Response:			
Movement	L	R	L->R	R->L
L	97.5	0.7	1.6	0.1
R	0.4	97.2	1.3	1.1
L->R	26.9	9.6	61.3	2.1
R->L	4.3	25.0	2.9	67.9

It is important to bear in mind (as detailed in the Appendix), that while the digital signal leaving the MAXMSP synthesis in each case contained very little energy at frequencies above 200 Hz, that generated by the subwoofers contained some energy detectable above the acoustic background of the studio, in some cases at frequencies up to about 600 Hz. Correspondingly, sending high dB sine tones to the speakers produced pitch-discernible audible sound at least up to 1000 Hz. Thus even 'clean' low frequency sounds as presented by these excellent speakers will always contain higher frequencies. The same observations held for larger (much more expensive!) Meyer subwoofers.

CONCLUSIONS

Can a composer use low frequency lateralization and movement, and hope for perceptibility?

I conclude that musically-experienced listeners have good capability in relation to location/lateralization and movement perception of low frequency sounds in our reverberant studio environment, though movement accuracy is much lower than location accuracy, as normally observed. There was worse performance with very short sounds. Our listeners may have learnt much about low frequency listening from their musical experience; but nevertheless, their performance improved slightly with trial as the experiment proceeded. On the other hand, I argue that most people learn from environmental sound around them, and that there is a biological advantage in gaining ability to localize even low frequency sounds, especially if they are moving. This remains to be more fully tested.

The results support the view that sound projection systems with multiple sub-woofers can add timbral flux and spatial control to composers' armories. It will be interesting to assess the influence of sub-woofers at different elevations [55], in 3D space, in addition to different azimuths in 2D space, as I have done here. This is especially so given the readily available Max patches for sound diffusion and movement, the panning software VBAP, and specialized facilities like those at ZKM introduced above, and the 22:4 system in our studio. In 3D spatialization, issues of front-back discrimination also come into play (not discussed here). These are generally far more problematic for listeners than lateral location or movement detection [3], and hence will require considerable attention.

Interesting in the longer run, both psychoacoustically and musically, are questions concerning the possible competition between musical low frequency and high frequency timbre spatialization. While already in practical use, the impact of these perceptually is little understood. Experimental acoustics and psychoacoustics will clearly have more to contribute to composers and performance space design in this area.

REFERENCES

[1] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, MA (1997)

[2] C. Stecker and F. Gallun, "Binaural hearing, sound localization, and spatial hearing", *Translational perspectives in auditory neuroscience: normal aspects of hearing*. San Diego (CA): Plural Publishing, Inc, (2012)

[3] T.R. Letowski and S.T. Letowski, *Auditory spatial perception: auditory localization*. 2012, DTIC Document.

[4] L. Rayleigh, "XII. On our perception of sound direction", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **13**, 214-232 (1907)

[5] A.R. Palmer, "Reassessing mechanisms of low-frequency sound localisation", *Current opinion in neurobiology* **14**, 457-460 (2004)

[6] M.I.J. Mohamed and D. Cabrera, "Human sensitivity to interaural phase difference for very low frequency sound", in *Acoustics 2008 - Australian Acoustical Society Conference*. p. p47.pdf.

[7] P. Lennox, Spatialization and Computer Music, in *The Oxford Handbook of Computer Music*, ed., R.T. Dean, Oxford University Press, New York, USA, 2009) p. 258.

[8] R. Normandeau, "Timbre Spatialisation: The medium is the space", *Organised Sound* **14**, 277 (2009)

[9] S. Kanach, *Music and Architecture by Iannis Xenakis*. 2008, Hillsdale, Pendragon Press.

[10] D. Worrall, "Space in sound: sound of space", *Organised Sound* **3**, 93-99 (1998)

[11] A.J. Hill, S.P. Lewis, and M.O. Hawksford, "Towards a generalized theory of low-frequency sound source localization", *Proceedings of the Institute of Acoustics* **34**, 138-149 (2012)

[12] P. Laine, H.-M. Lehtonen, V. Pulkki, and T. Raitio. "Detection and lateralization of sinusoidal signals in presence of dichotic pink noise", in *Audio Engineering Society Convention 122*. 2007. Audio Engineering Society.

[13] D. Cabrera, S. Ferguson, and A. Subkey. "Localization and image size effects for low frequency sound", in *Audio Engineering Society Convention 118*. 2005. Audio Engineering Society.

[14] W.L. Martens. "The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one?" in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*. 1999. Audio Engineering Society.

[15] A.J. Hill and M.O. Hawksford. "On the perceptual advantage of stereo subwoofer systems in live sound reinforcement", in *Audio Engineering Society Convention 135*. 2013. Audio Engineering Society.

[16] T.S. Welti. "Subjective comparison of single channel versus two channel subwoofer reproduction", in *Audio Engineering Society Convention 117*. 2004. Audio Engineering Society.

[17] J. Braasch, W.L. Martens, and W. Woszczyk. "Identification and discrimination of listener envelopment percepts associated with multiple low-frequency signals in multichannel sound reproduction", in *Audio Engineering Society Convention 117*. 2004. Audio Engineering Society.

[18] T. Nakayama, T. Miura, O. Kosaka, M. Okamoto, and T. Shiga, "Subjective assessment of multichannel reproduction", *Journal of the Audio Engineering Society* **19**, 744-751 (1971)

[19] G. Paine, R. Sazdov, and K. Stevens. "Perceptual Investigation into Envelopment, Spatial Clarity, and Engulfment in Reproduced Multi-Channel Audio", in *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*. 2007. Audio Engineering Society.

[20] M. Tucker, Dreaming with open eyes. The shamanic spirit in twentieth century arts and culture. Aquarian, London (1992)

[21] N.P.M. Todd and F.W. Cody, "Vestibular responses to loud dance music: A physiological basis of the "rock and roll threshold"?", *The Journal of the Acoustical Society of America* **107**, 496-500 (2000)

[22] R.T. Dean and F. Bailes, Event and process in the fabric and perception of electroacoustic music in *Proceedings of the international Symposium: Xenakis. The electroacoustic music* (http://www.cdmc.asso.fr/sites/default/files/texte/pdf/rencontres/intervention11_xenakis_electroacoustique.pdf), ed., (Centre de Documentation Musique Contemporaine, 2013) pp. Intervention11, pp.12.

[23] H. Smith, The voice in computer music and its relationship to place, identity and community, in *The Oxford Handbook of Computer Music*, ed., R.T. Dean, (Oxford University Press, New York, 2009) pp. 274-293.

[24] R.T. Dean and F. Bailes, "NoiseSpeech", *Performance Research* **11**, 134-135 (2007)

[25] H.A. Smith and R.T. Dean, "Voicescapes and Sonic Structures in the Creation of Sound Technodrama", *Performance Research* **8**, 112-123 (2003)

[26] M. Södersten, S. Ternström, and M. Bohman, "Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers", *Journal of Voice* **19**, 29-46 (2005)

[27] R.C. Smith and S.R. Price, "Modelling of human low frequency sound localization acuity demonstrates dominance of spatial variation of interaural time difference and suggests uniform just-noticeable differences in interaural time difference", *PLoS One* **9**, e89033 (2014)

[28] D. Huron, *Sweet anticipation.* MIT Press, Cambridge, MA (2006)

[29] R.T. Dean, *Hyperimprovisation: Computer interactive sound improvisation; with CD-Rom.* A-R Editions, Madison, WI (2003)

[30] P. Hegarty, "Just what is it that makes today's noise music so different, so appealing?", *Organised Sound* **13**, 13 (2008)

[31] D.R. Perrott and J. Tucker, "Minimum audible movement angle as a function of signal frequency and the velocity of the source", *The Journal of the Acoustical Society of America* **83**, 1522-1527 (1988)

[32] K. Saberi and D.R. Perrott, "Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate", *The Journal of the Acoustical Society of America* **87**, 1732-1737 (1990)

[33] L. Landy, Sound-based music 4 all, in *The Oxford Handbook of Computer Music*, ed., R.T. Dean, (Oxford University Press, New York, USA, 2009) pp. 518–535.

[34] T.D. Rossing, D.A. Russell, and D.E. Brown, "On the acoustics of tuning forks", *Am. J. Phys* **60**, 620-626 (1992)

[35] E.A. Macpherson and J.C. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited", *The Journal of the Acoustical Society of America* **111**, 2219-2236 (2002)

[36] J. Ahveninen, N. Kopčo, and I.P. Jääskeläinen, "Psychophysics and neuronal bases of sound localization in humans", *Hearing Research* **307**, 86-97 (2014)

[37] J. Liu, D. Perez-Gonzalez, A. Rees, H. Erwin, and S. Wermter, "A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation", *Neurocomputing* **74**, 129-139 (2010)

[38] W.O. Brimijoin, A.W. Boyd, and M.A. Akeroyd, "The contribution of head movement to the externalization and internalization of sounds", *PLoS One* **8**, e83068 (2013)

[39] W. Martens, S. Sakamoto, L. Miranda, and D. Cabrera. "Dominance of head-motion-coupled directional cues over other cues during walking depends upon source spectrum" in *Proceedings of Meetings on Acoustics*. 2013. Acoustical Society of America.

[40] D.R. Perrott and A. Musicant, "Minimum auditory movement angle: Binaural localization of moving sound sources", *The Journal of the Acoustical Society of America* **62**, 1463-1466 (1977)

[41] D.W. Grantham, "Detection and discrimination of simulated motion of auditory targets in the horizontal plane", *The Journal of the Acoustical Society of America* **79**, 1939-1949 (1986)

[42] N.P. Todd, A.C. Paillard, K. Kluk, E. Whittle, and J.G. Colebatch, "Vestibular receptors contribute to cortical auditory evoked potentials", *Hearing Research* **309**, 63-74 (2014)

[43] B. Blesser and L.-R. Salter, *Spaces speak, are you listening?* : MIT press, (2007)

[44] E. Larsen, N. Iyer, C.R. Lansing, and A.S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio", *The Journal of the Acoustical Society of America* **124**, 450-461 (2008)

[45] H. Al Saleh, *Effects of reverberation and amplification on sound localisation.* 2011, University of Southampton.

[46] A. Ihlefeld and B.G. Shinn-Cunningham, "Effect of source spectrum on sound localization in an everyday reverberant room", *The Journal of the Acoustical Society of America* **130**, 324-333 (2011)

[47] W.M. Hartmann, "Localization of sound in rooms", *The Journal of the Acoustical Society of America* **74**, 1380-1391 (1983)

[48] W.M. Hartmann and E.J. Macaulay, "Anatomical limits on interaural time differences: an ecological perspective", *Frontiers in neuroscience* **8**, (2014)

[49] A. Brughera, L. Dunai, and W.M. Hartmann, "Human interaural time difference thresholds for sine tones: the high-frequency limit", *The Journal of the Acoustical Society of America* **133**, 2839-2855 (2013)

[50] W.M. Hartmann and A. Brughera, "Threshold interaural time differences and the centroid model of sound localization", *The Journal of the Acoustical Society of America* **133**, 3512-3512 (2013)

[51] F.L. Wightman and D.J. Kistler, "The dominant role of low□ frequency interaural time differences in sound localization", *The Journal of the Acoustical Society of America* **91**, 1648-1661 (1992)

[52] R.Y. Litovsky, H.S. Colburn, W.A. Yost, and S.J. Guzman, "The precedence effect", *The Journal of the Acoustical Society of America* **106**, 1633-1654 (1999)

[53] W.M. Hartmann and B. Rakerd, "Localization of sound in rooms IV: The Franssen effect", *The Journal of the Acoustical Society of America* **86**, 1366-1373 (1989)

[54] H. Quené and H. Van den Bergh, "Examples of mixed-effects modeling with crossed random effects and with binomial data", *Journal of Memory and Language* **59**, 413-425 (2008)

[55] V.R. Algazi, C. Avendano, and R.O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies", *The Journal of the Acoustical Society of America* **109**, 1110-1122 (2001)

[56] C. Hak, R. Wenmaekers, and L. Van Luxemburg, "Measuring Room Impulse Responses: Impact of the Decay Range on Derived Room Acoustic Parameters", *Acta Acustica United with Acustica* **98**, 907-915 (2012)

Appendix A - The studio and experimental set up

The performance research space at MARCS Institute is roughly rectangular, 8 x 6.1 metres. It has 22 speakers in the roof area, up to 4 subwoofers on or near the floor, and two large retractable projection screens. These screens are on adjacent corrugated acoustic walls. The speakers for the experiment were 0.7 metres from these walls, and 3.3metres from the listening position, facing inwards. The other two walls of the studio contain glass windows allowing adjacent rooms to function as control rooms. The windows have retractable curtains, and these make a large difference to the coloration of sound in the studio, so the experiments were done with them completely open. The speaker and listener positions were chosen such that there was no consistent audible coloration distinction between the speakers with a wide range of input tones.

The Harlequin brand floor has a vinyl cover on top of wooden squares, providing suitable spring for dance use, and there were some normal studio items inside the room. The reverberation time (T30) measured from impulse responses in accordance with ISO 3382 [56] was 400 ms, and at 30 Hz this was extended to about 700 ms. The sounds were intended to be roving stimuli, and so they were not exactly matched for intensity: at the position of the listeners' heads they were measured to be 40-47 dB(A), using a Bruel and Kjaer 2250 sound level monitor. The noise sound was set at the lower intensity for listener comfort. Background noise levels in the studio were c. 30 dB.

The sounds were generated in MAXMSP as white noise, and as 30 Hz and 60 Hz sine tones. Each sound was digitally filtered through a MAXMSP low pass resonator (to reduce frequencies in the tones above 100 Hz, with 24 dB per octave roll off). The differences between the three spectra were obvious, and as expected. A MOTU 896 mk. III digital interface was used (sampling rate 44.1 kHz; output level -3dB), and the Genelec speakers were at default settings, with their nominal cut-off (equivalent to a cross-over) frequency being 120 Hz. They were visible, and the room was illuminated. It was noted that the loudness of all the sounds ramped rapidly to a maximum, but if sustained, then after 7-8 seconds it dropped to a new steady state which continued unchanged for a prolonged time. This was also observed with large Meyer subwoofers, and hence durations longer than 6 seconds were not suitable, and not used. The power spectrum of the sounds at the listeners' position was measured, and it showed the vast majority of energy to be below 200 Hz, but at higher frequencies there was slight energy above background levels, declining strongly and progressively with frequency. The 30 Hz stimulus was above background up to 400 Hz, the 60 Hz to 600 Hz (particularly during the transient ramp on), and the noise tone to 500 Hz.

The 5 participants (mean age 39.0 years, s.d. 17.2) all had musical experience; 4 also had experience of recording technology and practice; and there was one female. During an experiment, the stimuli were presented in randomized order, and in two blocks, each of 288 stimuli; participants could take a break between the blocks. They fixated on the computer screen while listening, and were asked to make their judgement of location/movement as quickly as possible after each sound had ended. Number keys were pressed to indicate the location (1 for L; 2 for R; 3 for LR; 4 for RL). Data was also recorded in the MAXMSP patch.

NEUROSCIENTIFIC INVESTIGATIONS OF MUSICAL RHYTHM

Daniel J. Cameron[1] and Jessica A. Grahn[1,2]
[1]Brain and Mind Institute, Western University, London, Canada
[2]Dept. of Psychology, Western University, London, Canada
dcamer25@uwo.ca

Music occurs in every human society, unfolds over time, and enables synchronized movements. The neural mechanisms underlying the perception, cognition, and production of musical rhythm have been investigated using a variety of methods. FMRI studies in particular have shown that the motor system is crucially involved in rhythm and beat perception. Studies using other methods demonstrate that oscillatory neural activity entrains to regularities in musical rhythm, and that motor system excitability is modulated by listening to musical rhythm. This review paper describes some of the recent neuroscientific findings regarding musical rhythm, and especially the perception of a regular beat.

INTRODUCTION

The temporal structure of music enables synchronized movement, such as tapping one's foot, clapping, or dancing to the 'beat' of musical rhythms. Such movement is precisely timed to align with the periodic, salient beats in the music, and with the movements of other individuals. Given this relationship between musical rhythm and movement, it is perhaps unsurprising that the brain's motor system is heavily involved in the neural processing of auditory rhythms. However, it is a relatively recent discovery that the motor system is involved even in the absence of movement – subtle differences in the temporal structure or context of an auditory rhythm can elicit robust differences in motor system activity. These discoveries are the topic of this review paper, with a focus on findings from functional magnetic resonance imaging (fMRI). FMRI measures the change in oxygenated blood levels following neural activity [see 1, 2]. This 'blood-oxygen-level dependent' (or BOLD) signal is considered to be an indirect measure of brain activity, and therefore increases in BOLD are termed 'activations' in this review. Findings from patient studies, as well as electroencephalography (EEG), magnetoencephalography (MEG), and transcranial magnetic stimulation (TMS) studies will also be discussed.

Although much theoretic and empirical work has sought to explain why certain temporal patterns elicit movement (e.g., dancing) while others do not [3-5], and the evolutionary basis for human sensitivity to musical rhythm [6-7], this review will focus on the neural substrates of rhythm perception and the role of individual differences, expertise, and sensory modality.

RHYTHM AND BEAT IN THE BRAIN

When human participants listen to rhythms (i.e., auditory sequences) with or without a beat, widespread activity is observed in the cortical motor system, especially in the supplementary motor area (SMA) and premotor cortex (PMC), as well as subcortical regions such as the basal ganglia and cerebellum [8-14]. Rhythms that are composed of intervals that are integer ratios of one another, and have accents occurring at regular intervals, tend to elicit the perception of a regular, emphasized beat, and beats are usually organized in a metre (a temporal structure determined by the cyclical pattern of strong and weak beats; see Figure 1). Compared to rhythms without a beat, listening to beat-based rhythms elicits more activity in the SMA and the basal ganglia [10]. The importance of the basal ganglia in beat perception was highlighted in a study demonstrating that patients with Parkinson's disease have impaired perceptual discrimination of changes in beat-based rhythms compared to healthy controls, but not in nonbeat rhythms [15]. This deficit in sensitivity to the beat structure in rhythms is presumably due to the degeneration of dopaminergic cells in a part of the basal ganglia called the substantia nigra; the death of these cells deprives the basal ganglia of dopamine, causing dysfunction. Overall, these findings suggest that the basal ganglia not only respond during beat perception, but are crucial for normal beat perception to occur.

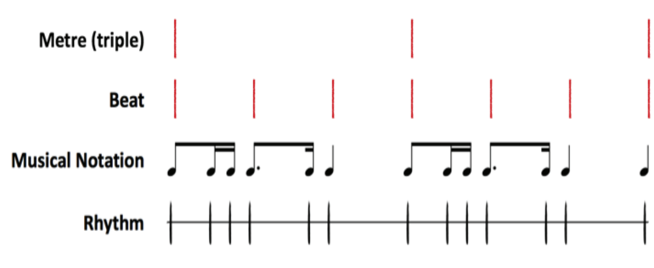


Figure 1. A depiction of rhythm, beat and metre. A rhythm is a sequence of auditory events, the onsets of which are separated by time intervals. The beat is the sequence of regular, salient time positions that are perceived in the rhythm. Metre is the hierarchical organization of beats into strong and weak (strong beats in the metrical structure are indicated in the top line).

In contrast to the basal ganglia, the cerebellum appears to play a different role in timing. Whereas the basal ganglia is important for beat perception and beat-based timing (i.e., timing of events relative to a regular and predictable beat), the cerebellum has been implicated in the perception of absolute time intervals (i.e., timing of events not relative to a beat). In one study, patients with cerebellar degeneration showed a deficit in absolute timing but not in beat-based timing [16]. A related study used TMS over the cerebellum to transiently disrupt function in that structure. After stimulation, participants performed worse in a single-interval timing task (i.e., a task that requires absolute timing), but not in a regularity (beat) detection task [17]. A subsequent fMRI study showed a dissociation between the cerebellum and basal ganglia (and the respective networks in which they operate) for absolute and beat-based timing: cerebellar regions and the inferior olive were more active for absolute timing, and regions of the basal ganglia, SMA, PMC, and other frontal cortical regions were more active for beat-based timing [18]. Both of these dissociable networks, however, are often active when hearing musical rhythms, suggesting that absolute and beat-based timing mechanisms are simultaneously engaged by rhythm processing.

THE TIME COURSE OF RHYTHM AND BEAT PERCEPTION

Beat perception is thought to have multiple stages: initially, when a rhythm is first heard, the beat must be detected, or found. 'Beat-finding' is followed by the creation of an internal representation of the beat, allowing the anticipation of future beats as the rhythm continues ('beat-continuation'). One fMRI study attempted to determine whether the role of the basal ganglia in beat perception was selective for finding or continuing the beat. Participants heard short, consecutive rhythms that either had a beat or not. Basal ganglia activity was low during the initial presentation of a beat-based rhythm, during which participants were engaged in beat-finding. Activity was high when beat-based rhythms followed one after the other, during which participants had a strong and continuing sense of the beat, suggesting that the basal ganglia are more involved in beat-continuation than beat-finding [19].

Another fMRI study compared activation during initial perception of a rhythm, when participants were engaged in beat-finding, to subsequent tapping of the beat as they heard the rhythm again (synchronized beat-tapping). In contrast to the previous study, basal ganglia activity was similar during finding and tapping (along with PMC and other temporal and frontal regions) [20]. The difference with respect to the previous study may be the result of differences in experimental paradigms, stimuli, or analyses, therefore further work remains to be done to clarify the role of the basal ganglia in beat finding versus beat continuation.

The consideration of the time course of rhythm and beat perception is an important topic of future research, as music and beat perception necessarily unfold over time and different stages may rely on distinct neural mechanisms (e.g., finding the beat, continuing the beat, and even adapting the beat rate in response when a rhythm changes). Most fMRI methods have a temporal resolution of about 1-2 seconds, but through appropriate experimental designs can be used to investigate the distinct stages of rhythm and beat perception, although not responses to each individual note.

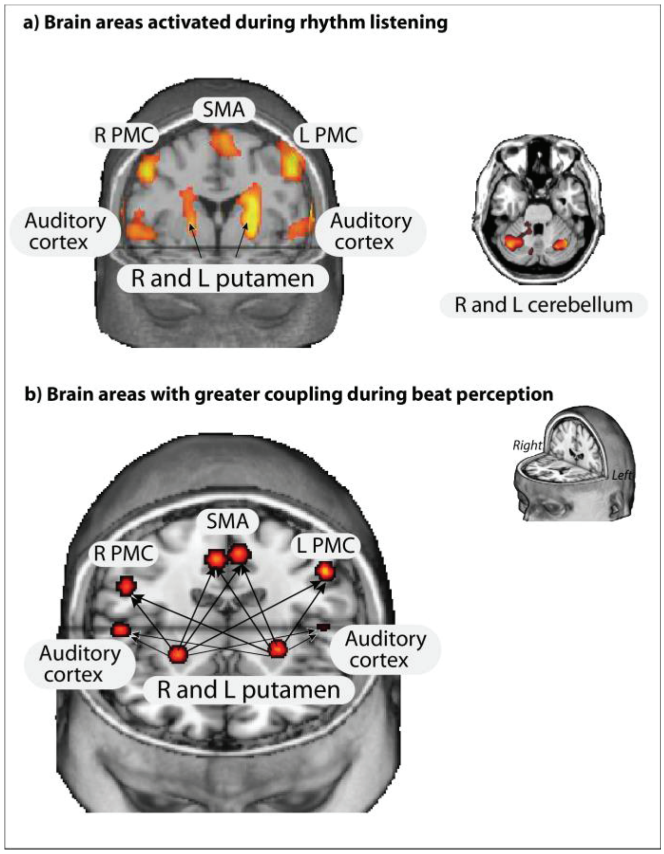


Figure 2. Neural regions that are a) active while listening to rhythms, and b) coupled (show greater correlation in activity) during beat perception [adapted from 10, 21]. PMC = premotor cortex, SMA = supplementary motor area, R = right, L = Left.

NEURAL CONNECTIVITY IN RHYTHM PERCEPTION

Individual brain regions do not act alone in the processing of musical rhythm, but rather function as networks. Using fMRI, the effective connectivity (the direct influence of one region's activity on that of another region) between the basal ganglia and several cortical areas, including the SMA, PMC and auditory cortex was found to be greater while participants listened to beat-based rhythms compared to nonbeat rhythms [21]. In another study, the functional connectivity (the non-directional correlation in activation) between PMC and auditory cortex was found to increase as the intensity of tones in beat positions of an isochronous sequence (or salience of the beat) increased [22]. Findings from these studies demonstrate that the connected activity, or coupling, between auditory and motor systems increases during rhythm and beat perception.

NEURAL OSCILLATIONS IN RHYTHM PERCEPTION

The studies discussed so far have used fMRI, which has poor temporal resolution – it is only sensitive to differences in activations occurring about 2 seconds apart. Although fMRI provides important insights about localization of neural activity due to its spatial resolution, other methods, such as EEG and

MEG, provide insight regarding brain responses on a much finer timescale. The fine temporal resolution of these methods can capture the oscillatory nature of neural activity. Neural populations typically have synchronized, oscillatory activity due to feedback connections. Although their functional role is not fully understood, neural oscillations at particular frequency bands have been linked to attention [23], stages of sleep [24], and movement [25]. EEG and MEG studies have demonstrated particular patterns of oscillatory activity in response to auditory rhythms. For example, one study found that when listening to a tone sequence of isochronous, alternating strong and weak beats, neural activity in the gamma band of oscillatory frequencies (in this case defined as 20-60 Hz) was greater for strong beats than weak beats [26]. In addition, when a tone in the sequence was occasionally omitted, anticipatory gamma responses occurred in the gaps where tones were expected. These findings suggest that gamma responses may index the perception and expectation of beats

In a similar study, participants heard a simple, repeating pattern: two identical tones followed by a silent gap. Participants imagined one of the two tones to be accented (emphasized), and oscillatory activity in the beta band (20-30 Hz; note that this overlaps with the gamma range as defined in the study discussed immediately above) was greater for the tone with imaginary accents than for the other tone [27]. Beta activity therefore shows a similar pattern to gamma activity in the study described above, but for imagined accents, rather than physical accents (greater intensity). Beta activity also appears to be sensitive to the expectation of when a tone will occur. When listening to isochronous sequences with varying rates, beta activity decreases following the onset of tones, but rebounds with a time course that is specific to the rate of the isochronous stimulus tones. The relationship between the timing of the beta rebound and the rate of the stimulus tones suggests that beta activity indexes anticipation of the timing of the next tone [28]. Thus, overall, both beta and gamma have been implicated in anticipation of the beat, but future research may elucidate whether they have distinct roles.

Recent work has shown entrainment of neural oscillations to the rate of the perceived beat in auditory rhythms. When participants heard a continuous isochronous tone sequence and imagined it as having either a duple metre or a triple metre (imagining an accent, or emphasis, on every two or three tones, respectively) there was increased power in entrained oscillations at the rate of the imagined emphasis, in addition to the rate of the tones themselves [29]. This specific enhancement of neural entrainment to the metrical beat rates in auditory rhythms was subsequently shown to occur for rhythms that are not isochronous (as in music) [30], and to occur in distinct brain regions for perception of the rhythm and tapping of the beat while listening [31]. These studies demonstrate entrainment to the beat in the auditory and motor systems of the brain.

MOTOR SYSTEM EXCITABILITY

Although fMRI and EEG/MEG provide complementary information about the location and timing, respectively, of neural activity, they are correlational methods. That is, they can only demonstrate associations between a condition or task, and neural activity. They cannot, however, indicate whether a particular brain response is causally related to the observed behaviour. Other methods, such as TMS, allow one to causally (and transiently) influence neural processing, and observe whether behaviour is affected. Recent studies have begun to use TMS to investigate the dynamics of excitability in the motor system during rhythm perception. Applying single pulses of TMS to the primary motor cortex can elicit a muscle twitch, or motor evoked potential (MEP). MEPs vary in magnitude depending on the excitability in the motor system at the time of stimulation. Three studies have used single pulse TMS to investigate how the perception of rhythm can modulate excitability in the motor system. One study found that MEPs elicited when tone sequences gave a strong sense of beat were larger than when sequences did not give a sense of beat [32]. A different study found that listening to music that is rated as having a high degree of 'groove' (or inducing a desire to move) modulates excitability at the time of the musical beat [33]. Another study found that excitability was modulated in correlation with how closely the rate of an isochronous sequence matched the participant's preferred tempo, determined by their spontaneous motor tempo (the rate at which they freely tapped) [34]. Together, these studies help show how the processing of rhythm by the brain's motor system can extend directly to the muscles, providing a mechanism by which rhythm might influence movement.

The relationship between rhythm and the motor system has also been studied with behavioural methods, including measuring the accuracy of synchronizing one's tapping with an auditory sequence [see 35], and theoretical and modelling approaches [e.g., 36). Rhythm's influence on the motor system is also exploited by movement rehabilitation in patients with motor disorders, such as Parkinson's Disease and stroke, using rhythmic auditory stimulation [see 37].

EXPERTISE AND INDIVIDUAL DIFFERENCES

The studies discussed thus far have usually manipulated rhythm structure in order to change perception (e.g., of the beat) in a general sample of participants. Individual differences among listeners, however, also influence rhythm and beat perception. Musical training is one difference that is commonly investigated. Musically trained and untrained individuals show similar coupling of activity between subcortical and cortical regions [21], and similar patterns of activity in the dorsal PMC, SMA, inferior parietal lobule and cerebellum, while listening to rhythms [38]. However, musicians show a greater increase in coupling between auditory cortex and the supplementary motor area when a beat is induced by the temporal structure of the rhythm, compared to when a beat is induced by regularly occurring volume accents [21]. Musicians also have greater activity in frontal regions and the cerebellum that covaries with the complexity of rhythms (defined by the metrical structure, varying from metrically simple to nonmetrical), compared to nonmusicians [38]. One study measured the contribution of several factors to individual differences in rhythm reproduction ability, including musical training, auditory short-term memory capacity, and sensitivity to the beat. All of these were found to be related to the ability to reproduce rhythms. Moreover, individual differences in these factors were associated with activity differences in response to hearing rhythms. Poor auditory short-term memory correlated with activity in auditory cortex, greater beat sensitivity correlated with activity in SMA and PMC, and musical training correlated with activity in both auditory and motor areas [39]. These findings underscore the importance of both motor and auditory systems in factors that lead to rhythmic ability. Another study found that while performing a temporal judgment task, individuals with strong beat perception had greater activity in the SMA, left PMC, insula, and inferior frontal gyrus compared to individuals with weak beat perception, who had greater activity in auditory cortex and right PMC [40]. In the normal human population there is a wide range of abilities and traits related to rhythm perception, and these studies only scratch the surface. Characterization of individual differences and their underlying causes needs to be addressed more comprehensively in the future.

AUDITORY SPECIFICITY OF BEAT PERCEPTION

The entrained behaviours associated with musical rhythms, such as dancing, do not generally occur in the real world for visual rhythms. Participants are much worse at tapping along with the beat of visual rhythms compared to auditory rhythms [41]. A few studies have tested the degree to which perception of a beat can be elicited by visual rhythms. In one study, participants were exposed to auditory and visual rhythms, and as expected, showed less sensitivity to the beat structure in visual rhythms. However, participants had a stronger sense of the beat in visual rhythms when they were preceded by identical rhythms in the auditory modality, suggesting that the timing mechanisms implicated in auditory beat perception can prime beat perception in other modalities [42]. Another study used a rhythm discrimination paradigm to show that sensitivity to the beat in rhythm can occur for visually presented rhythms, although overall performance is still worse for visual compared to auditory rhythms [43]. In that task, a rotating line was used to present the rhythms, and the added spatial information associated with each interval in the rhythm may have improved performance (i.e., compared to a simple flashing visual cue that always appears in the same spatial location). Other studies have used moving visual targets (such as a video of a finger tapping or a bouncing ball) and found that tapping in synchrony with these spatiotemporal visual stimuli was improved compared to purely temporal stimuli, but was still worse than for auditory stimuli [44, 45]. However, adding other visual information or using biological motion, may provide a means to more reliably induce beat perception from visual rhythms.

NONHUMAN PRIMATE STUDIES

The emerging research literature on rhythm and timing in nonhuman primates is bringing fresh insight to our understanding of the neuroscience of musical rhythm. Many studies have used a synchronization-continuation tapping task (in which monkeys were trained to synchronize their tapping with an isochronous auditory cue, then continue tapping at the same rate after the cue stopped) to compare rhythm and timing behaviour in humans and nonhuman primates. Rhesus monkeys and humans have similar performance when reproducing single intervals, but humans are far superior when synchronizing with sequences of multiple intervals [46]. Nonhuman primates are also worse at continuing tapping at the same rate after synchronizing with sequences of sounds, although there is preliminary evidence of at least one chimpanzee that shows some ability to do so [47]. One study found two populations of cells in the medial PMC of Rhesus monkeys that may provide distinct timing information during performance of the synchronization-continuation tapping task. The activity of some cells was sensitive to the duration of the interval being tapped, and the activity of other cells was sensitive to the time elapsed from the previous tap [48]. These two types of sensitivity could be used in conjunction with temporally precise movements, such as those required for musical rhythm production. A subsequent study showed that these mechanisms in the medial PMC are also used in the production of more complex, multiple-interval rhythms [49]. Distinct frequencies of oscillatory activity in the basal ganglia of Rhesus monkeys were found to relate to different aspects of rhythmic behaviour: activity in the gamma band (30-70 Hz) was more involved during synchronization of tapping, whereas activity in the beta band (13-30 Hz) was more involved in the continuation of tapping [50]. These results further implicate possible oscillatory mechanisms in cortical motor regions and the basal ganglia of humans that could underlie the particular roles these neural regions have for timing and movement synchronized to sound (such as required for the perception and production of musical rhythm). Behavioural work with nonhuman primates and other species will also help determine the degree to which beat perception is required for human-like synchronization-continuation [see 6, 7].

FUTURE DIRECTIONS

The future is promising for the neuroscience of musical rhythm. New and improving techniques (e.g., transcranial direct current stimulation, functional near-infrared spectroscopy), integration of existing techniques (e.g., computational modelling and neuroimaging), and the ongoing expansion of the research field (e.g., into animal research), will provide new insights into this topic.

Future research must face the persistent issue of stable and defined terminology. The use of 'beat', 'pulse', and 'complexity', for example, are not fully agreed upon in the literature, and the research community will benefit from standardizing the relevant terminology. Accounting for individual differences is becoming an increasingly apparent issue, with a wide range of rhythm abilities present in the normal population [39, 51].

The common comparison of musically trained and untrained groups (often based on a median split of years of musical training) is only an imprecise first step, insufficient to account for the variety of ways individuals can differ in their perception and processing of rhythm.

Future research will need to directly address the 'bottom-up' and 'top-down' influences on rhythm perception, and their inevitable interactions. For example, stimulus parameters like the temporal structure of intervals influence beat perception directly [3, 5], as do intensity dynamics in rhythms (i.e., compared to temporal structure alone) [21], but top-down influences of experience (e.g., culture, expertise) and of attention/intention (for paradigms requiring participants to intentionally impose a metrical beat structure onto rhythms) also influence beat perception. The interactions between these factors are largely uncharacterized, yet are critical for a complete understanding of the phenomena of interest.

Currently, the research literature is largely separated by methods, and future approaches could, for example, attempt to integrate across methods by accounting for entrained neural oscillations [e.g, 5], distinct timing mechanisms [19], and temporal expectation and probabilities in rhythms [e.g., 52] through convergent technical and theoretical methods.

CONCLUSIONS

Overall, recent neuroscientific investigations of the perception, behaviour, and neural processing related to musical rhythm have demonstrated the essential involvement of the brain's cortical and subcortical motor system. FMRI, EEG, TMS, and other methods have contributed to understanding the neural substrates, oscillatory frequencies, and excitability dynamics in relation to musical rhythm. Future work will likely focus on characterizing the exact neural pathways by which auditory and motor systems mutually influence each other during rhythm perception, and using this information to create neurobiologically plausible models of rhythm and beat perception.

ACKNOWLEDGEMENTS

The authors wish to thank Aaron Gibbings and Fleur Bouwer for their help creating Figure 1.

REFERENCES

[1] N. Logothetis and B.A. Wandell, "Interpreting the BOLD signal", *Annual Review of Physiology* **66**, 735-769 (2004)

[2] N. Logothetis, "What we can do and what we cannot do with fMRI", *Nature* **435**, 869-878 (2008)

[3] D.J. Povel and P.J. Essens, "Perception of temporal patterns", *Music Perception* **2**(4), 411–440 (1985)

[4] D.J. Povel and H. Okkerman, "Accents in equitone sequences", *Perception & Psychophysics* **30**(6), 565–572 (1981)

[5] E.W. Large and J.S. Snyder, "Pulse and meter as neural resonance", *Annals of the New York Academy of Sciences* **1169**, 46–57 (2009)

[6] H. Merchant and H. Honing, "Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis", *Frontiers in Neuroscience* **7**, (2013)

[7] A.D. Patel and J.R. Iversen, "The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis", *Frontiers in Systems Neuroscience* **8** (2014)

[8] S.L. Bengtsson, F. Ullén, H. Henrik Ehrsson, T. Hashimoto, T. Kito, E. Naito, H. Forssberg and N. Sadato, "Listening to rhythms activates motor and premotor cortices", *Cortex* **45**(1), 62–71 (2009)

[9] J.L. Chen, V.B. Penhune and R.J. Zatorre, "Listening to musical rhythms recruits motor regions of the brain", *Cerebral Cortex* **18**(12), 2844–2854 (2008)

[10] J.A. Grahn and M. Brett, "Rhythm perception in motor areas of the brain", *Journal of Cognitive Neuroscience* **19**(5), 893–9 (2007)

[11] P.A. Lewis, A.M Wing, P.A. Pope, P. Praamstra and R.C. Miall, "Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping", *Neuropsychologia* **42**, 1301–1312 (2004)

[12] J.M. Mayville, K.J. Jantzen, A. Fuchs, F.L. Steinberg and J.A.S. Kelso, "Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI", *Human Brain Mapping* **17**, 214–229 (2002)

[13] R.I. Schubotz and D.Y. von Cramon, "Interval and ordinal properties of sequences are associated with distinct premotor areas", *Cerebral Cortex* **11**, 210–222 (2001)

[14] F. Ullén, H. Forssberg and H.H. Ehrsson, "Neural networks for the coordination of the hands in time", *Journal of Neurophysiology* **89**(2), 1126-1135 (2003)

[15] J.A. Grahn and M. Brett, "Impairment of beat-based rhythm discrimination in Parkinson's disease", *Cortex* **45**(1), 54–61, (2009)

[16] M. Grube, F.E. Cooper, P.F. Chinnery and T.D. Griffiths, "Dissociation of duration-based and beat- based auditory timing in cerebellar degeneration", *Proceedings of the National Academy of Sciences of the United States of America* **107**(25), 11597–11601 (2010)

[17] M. Grube, K.H. Lee, T.D. Griffiths, A.T. Barker and P.W. Woodruff, "Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals", *Frontiers in Psychology* **1**, 171 (2010)

[18] S. Teki, M. Grube, S. Kumar and T.D. Griffiths, "Distinct neural substrates of duration-based and beat-based auditory timing", *Journal of Neuroscience* **31**, 3805–3812 (2011)

[19] J.A. Grahn and J.B. Rowe, "Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity", *Cerebral Cortex* **23**, 913-921 (2013)

[20] S.J. Kung, J.L. Chen, R.J. Zatorre and V.B. Penhune, "Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat", *Journal of Cognitive Neuroscience* **25**(3), 401-420 (2013)

[21] J.A. Grahn and J.B. Rowe, "Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat processing", *Journal of Neuroscience* **29**(23), 7540–7548 (2009)

[22] J.L. Chen, R.J. Zatorre and V.B. Penhune, "Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms", *NeuroImage* **32**, 1771–1781 (2006)

[23] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis", *Brain Research Reviews* **29**(2), 169-195 (1999)

[24] L. De Gennaro and M. Ferrara, "Sleep spindles: an overview", *Sleep Medicine Reviews* **7**(5), 423-440 (2003)

[25] D.J. McFarland, L.A. Miner, T.M. Vaughan and J.R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements", *Brain Topography* **12**(3), 177-186 (2000)

[26] J.S. Snyder and E.W. Large, "Gamma-band activity reflects the metric structure of rhythmic tone sequences", *Cognitive Brain Research* **24**(1), 117–126 (2005)

[27] J.R. Iversen, B.H. Repp and A.D. Patel, "Top-down control of rhythm perception modulates early auditory responses", *Annals of the New York Academy of Sciences* **1169**, 58-73 (2009)

[28] T. Fujioka, L.J. Trainor, E.W. Large and B. Ross, "Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations", *Journal of Neuroscience* **32**(5), 1791-1802 (2012)

[29] S. Nozaradan, I. Peretz, M. Missal and A. Mouraux, "Tagging the neuronal entrainment to beat and meter", *Journal of Neuroscience*, **31**(28), 10234-10240 (2011)

[30] S. Nozaradan, I. Peretz and A. Mouraux, "Selective neuronal entrainment to the beat and meter embedded in a musical rhythm", *Journal of Neuroscience*, **32**(49), 17572-17581 (2012)

[31] S. Nozaradan, Y. Zerouali, I. Peretz and A. Mouraux, "Capturing with EEG the neuronal entrainment and coupling underlying sensorimotor synchronization to the beat", *Cerebral Cortex* bht261v1-bht261 (2013)

[32] D.J Cameron, L. Stewart, M.T. Pearce, M. Grube and N.G. Muggleton, "Modulation of motor excitability by metricality of tone sequences" *Psychomusicology: Music, Mind, and Brain* **22**(2), 122-128

[33] J. Stupacher, M.J. Hove, G. Novembre, S. Schütz-Bosbach and P.E Keller, "Musical groove modulates motor cortex excitability: a TMS investigation", *Brain and Cognition* **82**(2), 127-136 (2013)

[34] K. Michaelis, M. Wiener and J.C. Thompson, "Passive listening to preferred motor tempo modulates corticospinal excitability", *Frontiers in Human Neuroscience* **8** (2014)

[35] B.H. Repp, "Sensorimotor synchronization: A review of the tapping literature" *Psychonomic Bulletin & Review* **12**(6), 969-992 (2005)

[36] N.P.M. Todd, C.S. Lee and D.J. O'Boyle, "A sensorimotor theory of temporal tracking and beat induction" *Psychological Research* **66**, 26-39 (2002)

[37] M.H. Thaut and M. Abiru, "Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research", *Music Perception* **27**(4), 263-269 (2010)

[38] J.L. Chen, V.B. Penhune and R.J. Zatorre, "Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training", *Journal of Cognitive Neuroscience* **20**(2), 226-239 (2008)

[39] J.A. Grahn and D. Schuit, "Individual differences in rhythmic ability: Behavioral and neuroimaging investigations", *Psychomusicology: Music, Mind, and Brain* **22**(2), 105-121 (2012)

[40] J.A. Grahn and J.D. McAuley, "Neural bases of individual differences in beat perception", *NeuroImage* **47**, 1894-1903 (2009)

[41] A.D. Patel, J.R. Iversen, Y. Chen and B.H. Repp, "The influence of metricality and modality on synchronization with a beat", *Experimental Brain Research* **163**(2), 226-238 (2005)

[42] J.A. Grahn, M.J. Henry and J.D. McAuley, "FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa", *Neuroimage* **54**(2), 1231-1243 (2011)

[43] J.A. Grahn, "See what I hear? Beat perception in auditory and visual rhythms", *Experimental Brain Research* **220**(1), 51-61 (2012)

[44] M.J. Hove, J.R. Iversen, A. Zhang and B.H. Repp, "Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome", *Psychological Research* **77**(4), 388-398 (2013)

[45] M.J. Hove, M.J. Spivey, and C.L. Krumhansl, "Compatibility of motion facilitates visuomotor synchronization", *Journal of Experimental Psychology: Human Perception and Performance* **36**, 1525-1534 (2010)

[46] W. Zarco, H. Merchant, L. Prado and J.C. Mendez, "Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys", *Journal of Neurophysiology* **102**(6), 3191-3202, (2009)

[47] Y. Hattori, M. Tomonaga and T. Matsuzawa, "Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee", *Scientific Reports* **3**, (2013)

[48] H. Merchant, W. Zarco, O. Pérez, L. Prado and R. Bartolo, "Measuring time with different neural chronometers during a synchronization-continuation task", *Proceedings of the National Academy of Sciences of the United States of America* **108**(49), 19784-19789 (2011)

[49] H. Merchant, O. Pérez, W. Zarco and J. Gámez, "Interval tuning in the primate medial premotor cortex as a general timing mechanism", *Journal of Neuroscience* **33**(21), 9082-9096 (2013)

[50] R. Bartolo, L. Prado and H. Merchant, "Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping", *Journal of Neuroscience* **34**(11), 3910-3923 (2014)

[51] J. Launay, M. Grube, and L. Stewart "Dysrhythmia: a specific congenital rhythm perception deficit", *Frontiers in Psychology* **5**, 18 (2014)

[52] G.M. Cicchini, R. Arrighi, L. Cecchetti, M. Giusti and D.C. Burr, "Optimal encoding of interval timing in expert percussionists", *The Journal of Neuroscience* **32**(3), 1056-1060 (2012)

MUSIC TRAINING: LIFELONG INVESTMENT TO PROTECT THE BRAIN FROM AGING AND HEARING LOSS

Nina Kraus[1,2] & Travis White-Schwoch[1]

[1] Auditory Neuroscience Laboratory, www.brainvolts.northwestern.edu, and Department of Communication Sciences, Northwestern University, Evanston, IL, USA

[2] Department of Neurobiology & Physiology, Northwestern University, Evanston, IL, USA and Department of Otolaryngology, Northwestern University, Chicago, IL, USA

nkraus@northwestern.edu

Age-related declines in the auditory system contribute strongly to older adults' communication difficulties, especially understanding speech in noisy environments. With the aging population growing rapidly there is an expanding need to discover means to offset or remediate these declines. Music training has emerged as a potential tool to set up the brain for healthy aging. Due to the overlap between neural circuits dedicated to speech and music, and the strong engagement of cognitive, sensorimotor, and reward circuits during music making, music training is thought to be a strong driver of neural plasticity. Comparisons of musicians and non-musicians across the lifespan have revealed that musicians have stronger neural processing of speech across timescales, ranging from the sentence and word level to consonant features on a millisecond level. These advantages are also present in older adult musicians, and they generalise to advantages in memory, attention, speed of processing, and understanding speech in noise. Excitingly, even older adult musicians with hearing loss maintain these neurophysiological and behavioural advantages, outperforming non-musicians with normal hearing on many auditory tasks. Delineating the neurophysiological and behavioural advantages associated with music experience in older adults, both with normal hearing and hearing loss, can inform the development of auditory training strategies to mitigate age-related declines in neural processing. These prospective enhancements can provide viable strategies to mitigate older adults' challenges with everyday communication.

INTRODUCTION

Older adults often have difficulty communicating in noisy or reverberant listening environments. This temporal processing deficit is compounded by presbycusis, meaning that older adults with hearing loss find speech comprehension in noise especially challenging. These communication challenges are not trivial: from talking on the telephone to ordering in a noisy restaurant, poor speech understanding in degraded listening environments can contribute to stress, social isolation, and depression [1]. In 2009 there were approximately 37 million adults over age 65 in the United States alone, a figure expected to double by 2030 [2]. And it is estimated that nearly two-thirds of adults over age 70 have hearing loss [3]. Therefore, there is an expanding need to develop strategies to mitigate older adults' communication challenges. More generally, there is a burgeoning interest in tools to promote "healthy aging" – investments made at any age to bolster cognition, health, and quality of life in senescence – including through diet, exercise, and vocational activities [4].

Music training has emerged as an exciting candidate for this prevention and remediation. Due to the overlap between neural circuits dedicated to speech and music, and the distributed network of cognitive, sensorimotor, and reward circuits engaged during music making, it would appear that music training is a potent driver of experience-dependent plasticity [5–7]. For the majority of studies discussed herein, music training is operationalized as engaging in active music training regularly, for a minimum of 20 minutes twice a week since childhood. These individuals are described as "musicians" as a shorthand, however the benefits for auditory processing associated with music training are likely observed in many individuals who have pursued training less rigorously [8,9], and those who may not self-identify as "musicians" in a formal sense [10,11]. This training is associated with neurophysiological benefits for encoding speech that cascade to heightened auditory-cognitive skills across the lifespan [12,13]. Therefore, music training may hold special promise to set up the communicating brain for healthy aging.

THE AGING AUDITORY SYSTEM: IMPACT ON COMMUNICATION

Pervasive age-related changes occur in the auditory system, which degrade the precision and stability of signal processing. These changes compound age-related declines in cognitive functions (speed of processing, memory, attention, etc.) that, taken together, create challenges for everyday communication [14,15]. In cases of age-related hearing loss, these communication challenges are even greater [16]. A large series of behavioural and neurophysiological studies has characterized the maladaptive plasticity incurred by aging and presbycusis (see [17] for an authoritative review).

Age-related declines in the auditory system

Most age-related changes that occur irrespective of hearing loss affect the fine temporal resolution required for coding fast-changing elements in speech, such as consonants. Age-related declines in temporal processing have been observed in psychophysical studies that have pinpointed a loss of temporal resolution as a hallmark of auditory aging [18–20]. Neurophysiologically, this temporal processing deficit is likely due to a pervasive reduction in inhibitory neurotransmitter function throughout the auditory neuraxis [for review, see 21]. This inhibitory loss is compounded by an increased postsynaptic recovery time [20] and likely a loss of ribbon synapses [22]. These declines cause a reduction in the *neural synchrony* that is required for speech perception in noise [23].

Anderson and her colleagues used the auditory brainstem response to complex sounds (cABR) to investigate age-related changes in the neural precision of speech encoding in older adults with normal hearing [24]. The cABR is a variant of the auditory brainstem response that is elicited in response to complex sounds such as speech or music. By using these sounds, cABR can measure the neural processing of both transient and sustained acoustic elements, providing unique insight into submillisecond temporal processing (neural phaselocking occurring predominantly between 100-1000 Hz). This rapid neural processing is important to encode details in speech such as formants and temporal fine structure that provide perceptual clarity and convey information about phonemic categories and sounds' locations. These features also support listening in the "dips" of certain kinds of maskers [16,25,26]. Five age-related declines in neural processing were described by Anderson *et al*. (see Table 1; Figures 1 & 2). In older adults, responses were *smaller*, including for representation of the fundamental frequency and harmonics in speech. Responses were more *variable* on a trial-by-trial basis, and there was an increase in timing jitter across frequencies. Older adults exhibited a *selective timing delay* on the order of a few milliseconds. This timing delay was only present for time regions of the response corresponding to the onset and consonant-vowel transition in speech; importantly, there was no neural timing delay for the vowel. cABR timing bears strongly on speech-in-noise perception and communicative skills broadly, with sub-millisecond timing differences distinguishing performance on speech perception tasks between groups [27,28]. Finally, there was an age-related increase in *spontaneous neural activity*, putatively representing more "neural noise." These findings have been replicated by a variety of studies [29–31], and conform well to complementary investigations in humans and animals [18,19,32,33].

Hearing loss

Presbycusis exacerbates the communication challenges posed by central auditory aging [16,34]. Although it is difficult to disentangle the effects of aging from those of age-related hearing loss, some of the classic age-related declines in auditory processing are exaggerated, such as the loss of inhibitory neurotransmitter function and misbalance of excitatory and inhibitory neurotransmitters [21,35]. Distinct cochlear pathologies, such as loss of outer hair cells and spiral ganglion cells, likely contribute strongly to age-related hearing loss as well [36]. These losses can modulate cochlear filter properties, eventually leading to downstream changes that may cause maladaptive gain mechanisms in peripheral and central structures [37–39].

A hot topic is how hearing loss exacerbates cognitive decline. Operating under the hypothesis that effective and active engagement with sound supports the maintenance of cognitive skills in older adults, Lin and his colleagues have demonstrated that age-related hearing loss can speed up the rate of cognitive decline in older adults [40]. The same group has found that older adults with better hearing thresholds have larger brain volumes, suggesting retention of cytoarchitectonic integrity through the active and meaningful engagement with sound that is facilitated by good hearing [41]. These studies illustrate two important points. For one, hearing loss affects more than the ear and, in fact, affects more than auditory system function. Changes in the *quality and consistency* of auditory experiences—theoretically, for the better or the worse—can propagate to cognitive functions. But for two, these studies are cases in point that the declines in central processing that are associated with aging and presbycusis may not be *fait accompli*: active engagement with sound can reinforce auditory-cognitive skills, potentially bolstering communicative abilities despite aging and hearing loss.

MUSIC TRAINING ACROSS THE LIFESPAN

Myriad correlational and cross-sectional studies have evaluated the impact of music training on the nervous system and associated behavioural functions. A recurring theme is that music training has a profound impact on *auditory* perception and cognition, in addition to its underlying neurophysiology. Although debates persist as to innate vs. trained differences in studies of music training [42], it would appear that irrespective of intelligence and other personality factors music training can effect changes in nervous system function (although in most cases music training likely interacts with several other factors to dictate final behavioural outcomes). These benefits are grounded in enhanced neural processing of speech [6], occurring across timescales, from slower features such as sentence-level processing, to syllable-level features such as pitch contours and phonemic cues such as voice onset time, to very rapid processing of millisecond features such as formant changes [43–48].

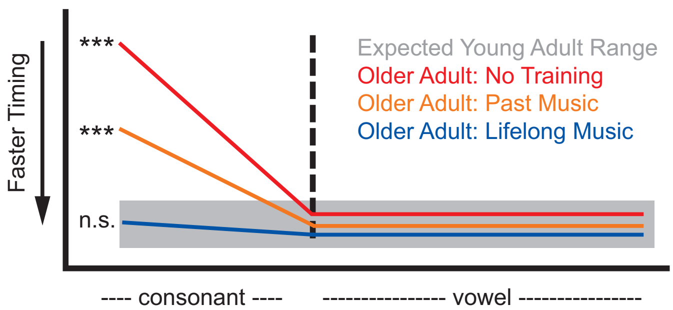


Figure 1. Schematic illustrating neural timing in older adults with no music training (red), past music training (orange), and lifelong music training (blue). Aging slows neural responses to consonants, however music training mitigates this effect. Lifelong musicians have neural timing within a typical young adult range.

The Kraus Laboratory has conducted a series of studies investigating the biological impact of music training on the nervous system across the lifespan, once again using the cABR. By measuring the precision of millisecond-level neural encoding of speech features, these techniques have pinpointed which acoustic aspects of speech processing are enhanced through music training, delineating a *neural signature* [5,12,49]. Briefly, musicians have faster neural timing in response to consonants, enhanced neural encoding of speech harmonics, more synchronous responses to speech, and more resistance to noise degradation (see Table 1).

But these benefits are not solely reflected in neural processing. In fact, these neural enhancements likely underlie a series of behavioural advantages in auditory perception and cognition. Compared to their non-musician peers, musicians have better speech understanding in noise, refined auditory temporal resolution, and heightened auditory memory and attention skills [5,12]. These enhancements combine to make the musician's brain a powerful canvas for auditory processing, tuned into behaviourally relevant sounds and primed to encode them precisely. Importantly, all of the domains where child and young-adult musicians outperform their non-musician peers are areas of decline in aging. This raises the question: can a life of music training mitigate age-related loss?

OLDER ADULTS: MUSIC, AGING, AND HEARING LOSS

A smaller number of studies have considered the biological impact of music training on the older adult's brain. Most of these have considered lifelong musicians, and have asked whether a life of playing music abates age-related declines in auditory processing. The answer is a resounding yes. Both behaviourally and neurophysiologically, older adult musicians do not exhibit many of the age-related declines in auditory function commensurate with typical aging.

Musicians and aging

Parbery-Clark and her colleagues conducted a series of studies of older adult musicians (ages 45-65) to describe the age-related changes—or lack thereof—that occur in neural speech processing. The first set of studies considered older adult musicians with normal hearing. Unlike their non-musician peers, older adult musicians do not exhibit the age-related neural timing delay in response to consonants in speech [50]. In fact, older adult musician's neural timing matches young adult *non-musicians* (see Figure 1). These musicians also had more robust representation of speech harmonics, more consistent responses to speech, and more resilient responses to noise degradation [51]. All told, four of the five signature aging effects on the neural encoding of speech appear to be absent in lifelong musicians (see Table 1; Figure 2). Impressively, these older adult musicians also outperform their non-musician peers on behavioural tests of speech understanding in noise, auditory temporal processing, and auditory working memory [52, see also 53,54]. Therefore, these biological enhancements appear to be linked to advantages in auditory perception and cognition as well.

Table 1. Summary of aging effects and whether they are offset by lifelong music training in older adults with normal hearing or hearing loss.

	Is the aging effect offset by music training in older adults with …	
Aging effect	**… normal hearing?**	**… hearing loss?**
Lower cognitive and perceptual performance	**YES**	**YES**
Neural timing delays in response to consonants	**YES**	**YES**
Decreased response magnitude for fundamental frequency	**NO**	**YES**
Decreased response magnitude for spectral harmonics	**YES**	**NO**
Decreased neural synchrony	**YES**	**YES**
Less precise encoding of the temporal envelope	**YES**	**NO**
Increased in spontaneous neural activity ("neural noise")	**NO**	**NO**

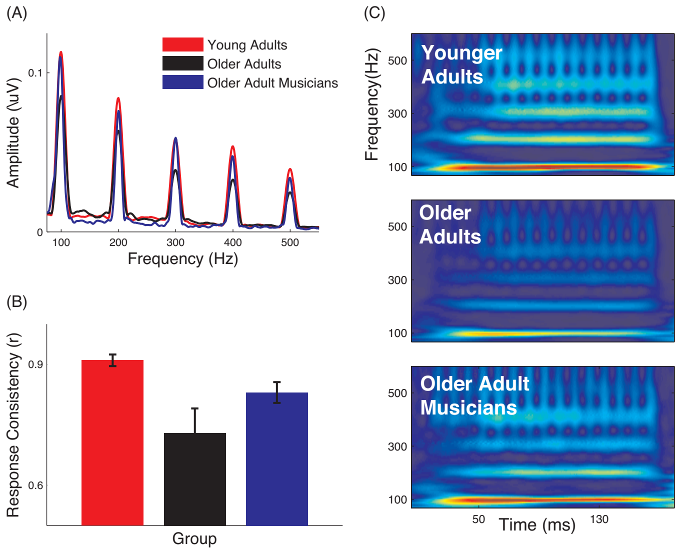


Figure 2. Aging effects that are offset by lifelong music training. (A) Older adult musicians have stronger encoding of the fundamental frequency and harmonics of speech, in line with young adults. (B/C) Age-related declines in neural synchrony are absent in lifelong musicians. This is reflected in the trial-by-trial stability of the neural response to speech overall (B) and on a frequency-specific basis (C).

A complementary series of studies has come from Zendel and Alain, who have compared auditory processing and attentional allocation in musicians and non-musicians throughout the lifespan. In a cross-sectional comparison of musicians and non-musicians (ages 18-91) Zendel and Alain found consistent lifelong advantages in central auditory processing in musicians irrespective of hearing thresholds [13]. Subsequent neurophysiological studies, considering auditory processing occurring predominantly below 100 Hz, demonstrated that older adult musicians had enhanced temporal processing at slower time scales, including for auditory stream segregation [55] and compensatory attention-dependent activity [56].

Taken together, these studies are consistent with the idea that lifelong music training can set up the brain for healthy aging. This is true for auditory tasks requiring processing across seconds and minutes all the way to sub-millisecond neural processing of very fine speech features. Given the gamut of slow-to-fast processing that appears enhanced in older adults, one may extrapolate to additional advantages that may be present in older musicians but have yet to be characterized biologically. For example, aging degrades the neural processing of voice onset time [57], a temporal cue in speech that informs phonemic categorization; music training, however, enhances this processing [44]. By making timing sensitivity behaviourally relevant to the listener, music training may engender biological enhancements particularly tailored to counteract the older adult's temporal resolution deficit.

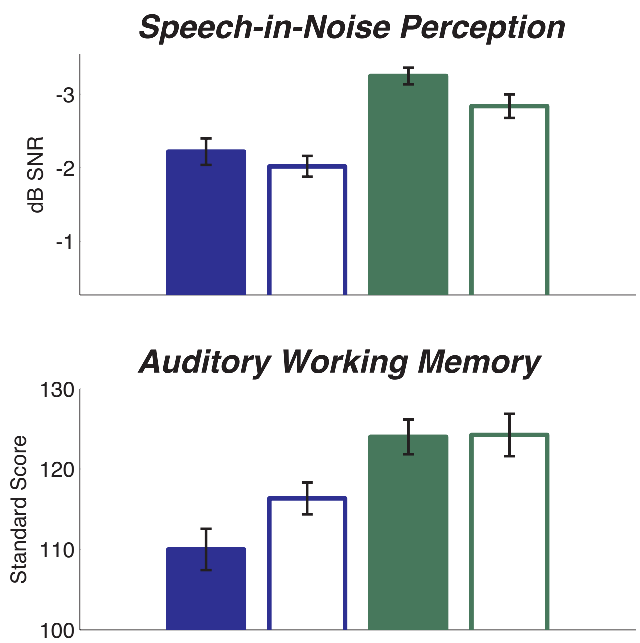


Figure 3. Behavioral advantages found in older adult non-musicians (blue; grouped on left) and musicians (green; grouped on right) musicians with normal hearing (solid bars) and hearing loss (open bars). Musicians outperform their non-musician peers on tests of speech perception in noise and auditory working memory. In fact, musicians with hearing loss even outperform non-musicians with normal hearing on the same tasks.

Musicians with hearing loss

Building upon their work in older musicians with normal hearing, Parbery-Clark and her colleagues considered the impact of music training on older adults with mild age-related hearing loss [58]. Older adult musicians with hearing loss exhibited several of the same biological enhancements as their normal-hearing peers, namely: faster neural timing in response to rapidly-changing sounds, more synchronous responses to speech, and greater resistance to noise degradation. However, in their neural responses older musicians with hearing loss had greater amplitudes at the fundamental frequency of speech as opposed to the enhanced harmonics seen in normal-hearing musicians. This unique neural signature is thought to be a compensatory mechanism developed to maintain robust encoding of sound despite a loss of peripheral function. Remarkably, these musicians with hearing loss outperformed normal hearing non-musicians on behavioural tests of speech-in-noise perception and auditory working memory (see Figure 3). These musicians' enhanced encoding of the fundamental frequency of speech may underlie their maintained behavioural advantage; indeed, in non-musician older adults robust neural encoding of the fundamental frequency supports speech perception in noise [15,59].

DOES MUSIC TRAINING HAVE TO BE LIFELONG?

The studies discussed so far have focused on individuals who played an instrument for their entire lives. This is a rare breed, especially in the context of senescence; it is much more common to encounter individuals who played music for a number of years as children and adolescents but then stopped as young adults. This raises an intriguing question: does the brain continue to benefit from these early experiences?

Skoe and Kraus [10] compared young adults (ages 18-31 yr) with varying levels of music training during childhood. More years of music lessons were associated with a higher signal-to-noise ratio in the neural response to sound, reflecting "cleaner" and more robust neural processing. Inspired by this work, White-Schwoch and colleagues [11] evaluated older adults (ages 55-76 yr) who played instruments from 1-14 years as children but had not touched an instrument for decades. Despite an intervening 40-50 years without training, older adults with past music experience had faster neural responses to consonants in speech than their peers—counteracting older adults' hallmark age-related temporal processing deficit [24].

Anderson and her colleagues [15] further investigated the impact of past music training on auditory perception, namely, the ability to understand speech in noise. They used structural equation modelling to elucidate the cognitive, central, peripheral, and lifestyle factors that contributed to older adults' abilities to understand speech in noisy listening environments. They dichotomized their subjects into two groups: one with no music training and a second with any amount (1-71 yr). The older adults with past music training relied more on cognitive functions such as working memory and attention to achieve the same performance on the speech-in-noise perception tasks, *irrespective of the amount of music training* and *irrespective of hearing status*. In older adults with no past music experience, life experience still informed mechanisms of hearing in noise, with socioeconomic status affecting hearing in noise and central auditory functions playing a stronger role.

These experiments are in line with animal studies that have demonstrated a lifelong impact of early sensory experience on auditory processing see, for example, [60]. A theme of these studies is that past auditory training—especially music training—may teach listeners to listen more meaningfully to sound. By directing attention to the most salient and acoustically complex elements in a soundscape, music training may subtly change the substrate mechanisms a listener uses to process novel sounds, even after said training has stopped. An intriguing possibility is that these listeners, even if they do not outperform peers on a cognitive or perceptual task [61], have achieved a different mode of *automatic* auditory processing that may "set the stage" for future auditory experiences [11]. If so, these individuals may be good candidates for auditory training to remediate challenges in auditory perception or cognition.

CONCLUSIONS

Taken together, the work reviewed here revolves around three general themes:

1. Age-related declines in auditory processing are not inevitable. They may be offset by the quality and consistency of everyday auditory experience.
2. Music training appears to be a powerful strategy to support meaningful interactions with sound, mitigating age-related decline in nervous system function.
3. Early auditory experiences, such as through music, are investments in healthy aging that pay lifelong dividends for auditory processing.

But why music? Music training directs special attention to meaningful acoustic features in the environment while engaging motor, cognitive, and emotional circuits. This rich series of networks combine into a powerful driver of experience-dependent neural plasticity. By allowing a listener to make sound-meaning connections, these listening activities can refine the *automatic* state of auditory processing, even during future listening tasks. This neural remodelling primes the musician's brain for effective and efficient auditory processing.

Despite the many accomplishments made with respect to aging, hearing loss, and music training, there remain open questions. Answering these questions can further inform the use of music training as a tool to remediate age- and hearing loss-related declines:

1. What is the impact of music training resumed or initiated early in life? In light of evidence that the nervous system retains substantial potential for plasticity into older age [see, for example, 62-64], and the large series of aforementioned benefits conferred by lifelong and early music, music training later in life may have special potential to engender improvements in auditory processing. Although there are some promising early studies [65,66], to date there have been no systematic investigations, and there have been no studies of biological changes following music training later in life. *Community-based* interventions are particularly appropriate for large-scale interventions in senior centres and retirement communities. Music would seem an especially suitable training regimen for these settings because music lends itself to group performance in choirs and ensembles.
2. How does music training compare to other training strategies for auditory rehabilitation, and what predictions can be made about who is a good candidate for which training strategy? Music training may not be a panacea, especially because there are some people who simply are not drawn to music. Understanding the pros and cons of different training strategies can inform clinicians interested in using auditory training as a part of their practice. Dosage studies can lead to best-practice protocols to identify candidates not just for auditory training in general, but for *particular kinds* of training.
3. What is the role of social and emotional engagement in music-related neuroplasticity? Listening to and performing music engages a large series of emotional networks, and it is thought that the emotional salience and motivation of a training regimen can bolster its neuroplastic potential.

Nevertheless, there is resounding evidence that music training can set up the brain for healthy aging. By bolstering the very aspects of auditory processing that decline with aging and hearing loss, music may prevent or mitigate the challenges that aging and hearing loss pose to spoken communication. Hopefully, music training at any point across the lifespan may lead to improvements in the quality of older adults' lives by bolstering everyday communication, grounded in improved auditory perception and cognition.

ACKNOWLEDGEMENTS

Thanks to Emily Spitzer for her assistance preparing figures. Supported by NSF BCS-1057556 and the Knowles Hearing Center.

REFERENCES

[1] Heine C and Browning C 2002 Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions *Disabil. Rehabil.* **24** 763–73

[2] US Census Bureau 2011 *2008 National Population Projections* (Washington, D.C.: US Census Bureau)

[3] Lin F R, Thorpe R, Gordon-Salant S and Ferrucci L 2011 Hearing loss prevalence and risk factors among older adults in the United States *J. Gerontol. A. Biol. Sci. Med. Sci.* **66** 582–90

[4] Paganini-Hill A, Kawas C H and Corrada M M 2011 Activities and mortality in the elderly: The Leisure World Cohort Study *J. Gerontol. A. Biol. Sci. Med. Sci.* **66A** 559–67

[5] Kraus N and Chandrasekaran B 2010 Music training for the development of auditory skills *Nat. Rev. Neurosci.* **11** 599–605

[6] Patel A D 2011 Why would musical training benefit the neural encoding of speech? The OPERA hypothesis *Front. Psychol.* **2**

[7] Herholz S C and Zatorre R J 2012 Musical training as a framework for brain plasticity: Behavior, function, and structure *Neuron* **76** 486–502

[8] Putkinen V, Tervaniemi M and Huotilainen M 2013 Informal musical activities are linked to auditory discrimination and attention in 2–3-year-old children: an event-related potential study *Eur. J. Neurosci.* **37** 654–61

[9] Tierney A, Krizman J, Skoe E, Johnston K and Kraus N 2013 High school music classes enhance the neural processing of speech *Front. Psychol.* **4** 855

[10] Skoe E and Kraus N 2012 A little goes a long way: How the adult brain is shaped by musical training in childhood *J. Neurosci.* **32** 11507–10

[11] White-Schwoch T, Woodruff Carr K, Anderson S, Strait D L and Kraus N 2013 Older adults benefit from music training early in life: Biological evidence for long-term training-driven plasticity *J. Neurosci.* **33** 17667–74

[12] Strait D L and Kraus N 2014 Biological impact of auditory expertise across the life span: musicians as a model of auditory learning *Hear. Res*. **308**, 109-121.

[13] Zendel B R and Alain C 2012 Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* **27** 410

[14] Salthouse T A 1985 *A theory of cognitive aging* Amsterdam: North Holland

[15] Anderson S, White-Schwoch T, Parbery-Clark A and Kraus N 2013 A dynamic auditory-cognitive system supports speech-in-noise perception in older adults *Hear. Res.* **300** 18–32

[16] Lorenzi C, Gilbert G, Carn H, Garnier S and Moore B C 2006 Speech perception problems of the hearing impaired reflect inability to use temporal fine structure *Proc. Natl. Acad. Sci.* **103** 18866–9

[17] Gordon-Salant S 2014 Aging, hearing loss, and speech recognition: stop shouting, I can't understand you *Perspectives on Auditory Research* Popper, Arthur, Fay, Richard R. (Eds.), New York: Springer. pp. 221-228.

[18] Gordon-Salant S and Fitzgibbons P J 1993 Temporal factors and speech recognition performance in young and elderly listeners *J. Speech Lang. Hear. Res.* **36** 1276

[19] Pichora-Fuller M K, Schneider B A, MacDonald E, Pass H E and Brown S 2007 Temporal jitter disrupts speech intelligibility: A simulation of auditory aging *Hear. Res.* **223** 114–21

[20] Walton J P 2010 Timing is everything: Temporal processing deficits in the aged auditory brainstem *Hear. Res.* **264** 63–9

[21] Caspary D M, Ling L, Turner J G and Hughes L F 2008 Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system *J. Exp. Biol.* **211** 1781–91

[22] Sergeyenko Y, Lall K, Liberman M C and Kujawa S G 2013 Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline *J. Neurosci.* **33** 13686–94

[23] Kraus N, Bradlow A R, Cheatham M A, Cunningham J, King C D, Koch D B, Nicol T G, McGee T J, Stein L and Wright B A 2000 Consequences of neural asynchrony: a case of auditory neuropathy *JARO-J. Assoc. Res. Otolaryngol.* **1** 33–45

[24] Anderson S, Parbery-Clark A, White-Schwoch T and Kraus N 2012 Aging affects neural precision of speech encoding *J. Neurosci.* **32** 14156–64

[25] Smith Z M, Delgutte B and Oxenham A J 2002 Chimaeric sounds reveal dichotomies in auditory perception *Nature* **416** 87–90

[26] Moore B C 2008 The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people *JARO-J. Assoc. Res. Otolaryngol.* **9** 399–406

[27] Anderson S, Skoe E, Chandrasekaran B and Kraus N 2010 Neural timing is linked to speech perception in noise *J. Neurosci.* **30** 4922

[28] Banai K, Hornickel J, Skoe E, Nicol T, Zecker S and Kraus N 2009 Reading and subcortical auditory function *Cereb. Cortex* **19** 2699–707

[29] Vander Werff K R and Burns K S 2011 Brain stem responses to speech in younger and older adults *Ear Hear.* **32** 168

[30] Clinard C G and Tremblay K in press Aging degrades the neural encoding of simple and complex sounds *J. Am. Acad. Audiol.*

[31] Bidelman G M, Villafuerte J W, Moreno S and Alain C 2014 Age-related changes in the subcortical-cortical encoding and categorical perception of speech *Neurobiol. Aging* in press.

[32] Parthasarathy A and Bartlett E 2011 Age-related auditory deficits in temporal processing in F-344 rats *Neuroscience* **192** 619–30

[33] Ruggles D, Bharadwaj H and Shinn-Cunningham B G 2012 Why middle-aged listeners have trouble hearing in everyday settings *Curr. Biol.* Vol 22, pp. 1417–1422.

[34] Dubno J R, Dirks D D and Morgan D E 1984 Effects of age and mild hearing loss on speech recognition in noise *J. Acoust. Soc. Am.* **76** 87–96

[35] Kotak V C, Fujisawa S, Lee F A, Karthikeyan O, Aoki C and Sanes D H 2005 Hearing loss raises excitability in the auditory cortex *J. Neurosci.* **25** 3908–18

[36] Engle J R, Tinling S and Recanzone G H 2013 Age-related hearing loss in rhesus monkeys is correlated with cochlear histopathologies *PloS ONE* **8** e55092

[37] Schaette R and McAlpine D 2011 Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model *J. Neurosci.* **31** 13452–7

[38] Henry K S and Heinz M G 2012 Diminished temporal coding with sensorineural hearing loss emerges in background noise *Nat. Neurosci.* **15** 1362–4

[39] Anderson S, Parbery-Clark A, White-Schwoch T, Drehobl S and Kraus N 2013 Effects of hearing loss on the subcortical representation of speech cues. *J. Acoust. Soc. Am.* **133** 3030–8

[40] Lin F R, Yaffe K, Xia J, Xue Q-L, Harris T B, Purchase-Helzner E, Satterfield S, Ayonayon H N, Ferrucci L, Simonsick E M and others 2013 Hearing loss and cognitive decline in older adults *JAMA Intern. Med.* **173** 293–9

[41] Lin F R, Ferrucci L, An Y, Goh J, Metter E J, Davatzikos C, Kraut M and Resnick S (2013) Hearing loss and changes in brain volume in older adults. *Assoc. Res. Otolaryngol. Abs*. **312**

[42] Corrigall K A, Schellenberg E G and Misura N M 2013 Music training, cognition, and personality *Front. Psychol.* **4**

[43] Parbery-Clark A, Skoe E and Kraus N 2009 Musical experience limits the degradative effects of background noise on the neural processing of sound *J. Neurosci.* **29** 14100–7

[44] Chobert J, François C, Velay J-L and Besson M 2012 Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time *Cereb. Cortex*

[45] Strait D L, O'Connell S, Parbery-Clark A and Kraus N 2013 Musicians' enhanced neural differentiation of speech sounds arises early in life: Developmental evidence from ages three to thirty *Cereb. Cortex* in press

[46] Strait D L and Kraus N 2011 Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise *Front. Psychol.* **2**

[47] Bidelman G M, Gandour J T and Krishnan A 2011 Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem *J. Cogn. Neurosci.* **23** 425–34

[48] Zuk J, Ozernov-Palchik O, Kim H, Lakshminarayanan K, Gabrieli J D, Tallal P and Gaab N 2013 Enhanced syllable discrimination thresholds in musicians *PloS One* **8** e80546

[49] Kraus N and Nicol T 2014 The cognitive auditory system *Perspectives on Auditory Research* Springer Handbook of Auditory Research ed R Fay and A Popper (Heidelberg: Springer-Verlag) pp 299–319

[50] Parbery-Clark A, Anderson S, Hittner E and Kraus N 2012 Musical experience offsets age-related delays in neural timing *Neurobiol. Aging* **33** 1483.e1–1483.e4

[51] Parbery-Clark A, Anderson S, Hittner E and Kraus N 2012 Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults *Front. Aging Neurosci.* **4**

[52] Parbery-Clark A, Strait D L, Anderson S, Hittner E and Kraus N 2011 Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise *PLoS ONE* **6** e18082

[53] Hanna-Pladdy B and MacKay A 2011 The relation between instrumental musical activity and cognitive aging. *Neuropsychology* **25** 378

[54] Amer T, Kalender B, Hasher L, Trehub S E and Wong Y 2013 Do older professional musicians have cognitive advantages? *PloS ONE* **8** e71630

[55] Zendel B R and Alain C 2013 The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation *J. Cogn. Neurosci.* **25** 503–16

[56] Zendel B R and Alain C 2014 Enhanced attention-dependent activity in the auditory cortex of older musicians *Neurobiol. Aging* **35** 55–63

[57] Tremblay K L, Piskosz M and Souza P 2003 Effects of age and age-related hearing loss on the neural representation of speech cues *Clin. Neurophysiol.* **114** 1332–43

[58] Parbery-Clark A, Anderson S and Kraus N 2013 Musicians change their tune: How hearing loss alters the neural code *Hear. Res.* **302**, 121–131.

[59] Anderson S, Parbery-Clark A, Yi H-G and Kraus N 2011 A neural basis of speech-in-noise perception in older adults *Ear Hear.* **32** 750

[60] Zhou X and Merzenich M M 2008 Enduring effects of early structured noise exposure on temporal modulation in the primary auditory cortex *Proc. Natl. Acad. Sci.* **105** 4423–8

[61] Gaab N and Schlaug G 2003 The effect of musicianship on pitch memory in performance matched groups *NeuroReport* **14** 2291–5

[62] Anderson S, White-Schwoch T, Parbery-Clark A and Kraus N 2013 Reversal of age-related neural timing delays with training *Proc. Natl. Acad. Sci.* **110** 4357–62

[63] Anderson S, White-Schwoch T, Choi H J and Kraus N 2013 Training changes processing of speech cues in older adults with hearing loss *Front. Syst. Neurosci.* **7**

[64] Anguera J, Boccanfuso J, Rintoul J, Al-Hashimi O, Faraji F, Janowich J, Kong E, Larraburo Y, Rolle C, Johnston E and others 2013 Video game training enhances cognitive control in older adults *Nature* **501** 97–101

[65] Bugos J, Perlstein W, McCrae C, Brophy T and Bedenbaugh P 2007 Individualized piano instruction enhances executive functioning and working memory in older adults *Aging Ment. Health* **11** 464–71

[66] Johnson J K, Louhivuori J, Stewart A L, Tolvanen A, Ross L and Era P 2013 Quality of life (QOL) of older adult community choral singers in Finland *Int. Psychogeriatr.* 1–10

PITCH PROCESSING IN MUSIC AND SPEECH

Barbara Tillmann[1,2,3]

[1] CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics team, Lyon, F-69000, France

[2] University of Lyon 1, Lyon, F-69000, France

[3] MARCS Institute, University of Western Sydney

barbara.tillmann@cnrs.fr

The present paper proposes an overview of research that investigates pitch processing by considering cognitive processes (related to context, learning, memory and/or knowledge) for both music and language materials. Research investigating cross-domain influences of expertise (either in music or tone languages) and deficits (as in congenital amusia), referred to as positive and negative transfer effects, also contributes to our understanding of domain-specificity or –generality of mechanisms involved in pitch processing.

INTRODUCTION

A highly-debated question is to what extent music and language share processing components. Beyond syntax and temporal structure processing, one studied aspect is pitch-processing in a given domain and across domains (e.g., [1]).

Pitch processing is crucial in music. For example, in Western tonal music, it is a form-bearing dimension (next to temporal structures). Pitch processing is also crucial in speech, notably for discriminating questions and statements, as well as for communicating emotional expressions. This is valid for non-tone intonation languages (e.g., English, French) and tone languages (e.g., Mandarin, Thai, Vietnamese). However, for tone languages, pitch processing is even more crucial as pitch information is used for communicating word meaning. Tone languages comprise 70% of the world's languages and are spoken by more than 50% of the world's population. In these languages, tone variations (comprising predominantly fundamental frequency (F0) height and contour parameters) at the syllabic level have the same effect on word meaning as do vowel and consonant variations in non-tone languages. For example, the syllable /ma/ combined with different tones (e.g., tones describing contours of rather constant level, rising dipping or falling patterns in Mandarin) represents different lexical items.

Interestingly, music and language differ in the size of the pitch differences that are relevant for each of the systems (speech intonation versus musical structures). For speech intonation of non-tonal languages, F0 variations, in particular those indicating statements and questions, are typically coarse (up to more than 12 semitones[1] for the pitch rise of the final word in a question; e.g., [2]). For music (as in most research, we are referring here mostly to the Western tonal system), however, the pitch variations are typically more fine-grained (1 or 2 semitones; see [3]). In tone languages, the range of F0 variations can be as small as in music of the Western tonal system or larger depending on the tones and the tone languages.

The present paper proposes an overview of research that investigates pitch processing by considering cognitive processes (related to context, learning, memory and/or knowledge) for both music and language material. While extensive research has focussed on pitch processing in musical material as well as the influence of musical expertise (e.g., comparing musicians and nonmusicians), research investigating pitch processing in tone language material for either musicians, nonmusicians or tone language speakers (or both), provides complementary information about underlying mechanisms. Furthermore, investigating deficits (such as in amusia) provides further insights into the potential domain-specificity or domain-generality of pitch processing mechanisms.

PITCH PROCESSING: INFLUENCE OF CONTEXT, KNOWLEDGE AND MATERIAL

Pitch processing has been investigated in detail within the research domain of psychoacoustics. However, psychoacoustic studies investigating early processes of pitch discrimination have mostly used pure tones or complex tones (see [4] for a review) and the rare studies using verbal material focused on vowel formants or one of the formants (e.g., [5]).

Even though no study has investigated pitch discrimination thresholds within the same participants for both non-verbal and verbal materials, some studies have examined thresholds for each domain separately. The findings suggest that pitch discrimination is more precise for musical material than for speech material (in typical listeners). For example, pitch discrimination thresholds (above 100Hz) are around 0.2% for complex tones [4] and are 10 times larger (2%) for vowels [6]. A recent study has compared for the first time the same series of pitch changes for music and speech in the same listeners

FOOTNOTE

[1] A semitone (or half tone) is the smallest musical interval commonly used in Western tonal music (e.g., the interval between tone C and tone C#); one semitone corresponds to 100 cents.

(even though it was not measuring thresholds). The findings of typical nonmusician listeners are consistent with these previous data, notably reporting higher accuracy for pitch processing when instantiated in musical (non-verbal) material than in verbal material [7]. Interestingly, individuals afflicted by congenital amusia (i.e., a lifelong deficit of music processing) showed better performance for verbal material than for musical material, even though they were generally impaired (see below). This benefit of speech for pitch discrimination in congenital amusia (and the reverse data pattern for typical nonmusician listeners) might be due to differences at early perceptual processing steps (e.g., exploiting differently the energy distribution in the sounds' spectrum linked to the presence vs. absence of formants) and/or higher level processing steps (i.e., strategic influences, attention and memory) (see [7] for further discussion).

Another characteristic of numerous psychoacoustic studies is the investigation of pitch processing of a single sound, which is presented in an isolated way without other surrounding sounds, thus "without a sound context" (besides those of the experimental context of the paradigm). In contrast, research in cognitive psychology has reported powerful effects of sound contexts as well as knowledge-driven influences (also referred to as top-down influences) on perceptual processing. These cognitive influences interact with the stimulus-driven influences (also referred to as bottom-up influences). Some research strands have investigated these influences on pitch processing in particular. In a natural environment, a sound rarely occurs on its own, but it rather occurs in a context with other sounds (e.g., presented in a sequence). The context may contain local regularities (such as a repeating pattern, e.g., AABAABAAB) or may refer to a structural system, which is based on statistical regularities (i.e., how the events are used together and thus define regularities of frequencies of occurrence or co-occurrence, for example). If listeners have knowledge about the underlying sound system, this knowledge allows listeners to develop expectations for future incoming sound events, and these expectations then influence event processing. Musical systems, which include structural regularities based on events differing in pitch, and listeners' knowledge thereof are examples of these cognitive influences on pitch processing.

Cognitive influences for the processing of musical materials have been reported not only for musician listeners, but also for nonmusician listeners who have been reported to have knowledge about the musical system of their culture (e.g., [8]). Performance of nonmusicians is compared to performance of musicians, that is individuals who are trained in performing music, with theoretical background knowledge and several thousands of hours of music production. While for some tasks, nonmusicians perform as musicians, for other tasks they are outperformed by the musicians. The musicians' musical background has been shown to influence pitch processing not only for musical, non-verbal material, but also for speech material. Conversely, it is interesting to also investigate the influence of language background on pitch processing in speech and musical material – in particular, native speakers of tone languages. This research thus also integrates into the research domain investigating whether processes and/or cognitive and neural correlates are domain-specific or domain-general.

CONTEXT EFFECTS: TOP-DOWN, KNOWLEDGE-DRIVEN EFFECTS ON PITCH PROCESSING

The influence of context effects on pitch processing has been shown not only for musical materials, but also for non-musical materials, that is, for example, contexts made out of several tones defining Gestalt-like features (e.g., a series of tones describing a descending melodic contour). When a sound context is defined by such a descending melodic contour (i.e., sounds constantly decreasing in pitch height from high to low frequency), participants are more sensitive to the detection of a target whose pitch is placed in the continuity of the instilled contextual movement than when it violates this movement [9]. The perceptual expectations developed with contextual movement function as an indirect signal, which facilitates the detection of the target sound. Music cognition research has investigated whether expectations based on listeners' tonal knowledge might also serve as an indirect indication of the target's pitch height and thus facilitate its processing. More specifically, processing the pitch of a tonally strongly related tone at the end of the melody (and thus supposed to be expected tone) should be facilitated in comparison to processing the pitch of an unexpected or less-related tone.

For musician listeners, Francès [10] has shown that tonal expectations influence listeners' perceptual ability to detect changes in the shift of the F0 of a musical tone (the target), which was followed by another tone (thus presented as a tone pair, defining an interval). When presented without additional, surrounding tones, the performed mistunings of the target tone, which either reduced or increased the pitch interval defined by the two tones, were detected by the participants (all musicians). However, when placed in a musical context (i.e., additional tones, which defined tonal structures and a tonality, were presented before the same tone pair), the same mistunings of the target tone were only perceived when they increased the pitch intervals and conflicted with listeners' expectations linked to musical anchoring, but not when they were in agreement with the musical patterns of tension and relaxation induced by the musical structures (see also [11]).

For nonmusician listeners, Warrier and Zatorre [12] have shown that an increasingly long tonal context (i.e., melodies with increasing duration) improved participants' capacity to process pitch information despite timbre changes. The authors interpreted this benefit by suggesting that the contextual tones create a stronger reference point from which to judge the pitch, and that the tonal structure of the melodies provides cues that facilitate pitch processing. In the experimental material, the to-be-judged tone played the role of the tonic, the most referential tonal function, and this function might have been also beneficial for pitch processing.

Other research focuses more specifically on the influence of tonal functions and structures on pitch processing, and this in particular for nonmusician listeners (see [8, 13] for reviews). Most of this research investigated harmonic structures by using chord sequences, but some more recent research applied these research questions to melodic materials and to tonal structures implemented in melodies, thus allowing for the investigation of pitch processing more specifically. In Marmel et al. [14], the influence of tonal function (also referred to as tonal relatedness) was investigated for pitch processing in melodies. Tonal expectations for target tones were manipulated in melodic contexts by changing just one tone of the context (which could be repeated). This subtle manipulation changed the tonal function of the final tone. It allows investigation of the influence of tonal expectations on pitch perception while controlling expectations based on information stored in sensory memory buffer. Excluding this kind of sensory influence (which otherwise would be a parsimonious, stimulus-driven explanation) allows for the investigation of cognitive influences, thus to provide evidence for listeners' tonal knowledge, which influences perception. Results showed that even for nonmusician listeners, the tonal relatedness of a target tone influences not only listeners' subjective judgments of tuning/mistuning (using a subjective rating scale), but also the speed of processing: in-tune tones that are tonally related are processed faster than in-tune tones that are less related (shown by using a priming task). Most interestingly, the tonal expectations also influence pitch discrimination: participants' performance in a pitch comparison task requiring the processing of small mistunings was better when the to-be-compared tones were tonally related to the melodic context (i.e., functioning as a tonic tone rather than as a subdominant tone). The findings suggest facilitation of early perceptual processing steps via knowledge- and attention-related processes (and not only later cognitive processing steps related to, for example, decision making for the experimental task). This has been further confirmed with finer differences in tonal relatedness (excluding the central tonic; [15]) and with even more controlled experimental material, using melodies played by pure tones to strip off potential sensory influences of the complex harmonic spectrum of the musical timbre used to play the context [16]. It is worth underlying that these findings have been obtained for nonmusician listeners. These tonal context effects do not require explicit musical knowledge, but point to the power of implicit cognition (here listeners' knowledge about the musical system, just acquired by mere exposure; see [8, 17]). The early influence of tonal expectations has been further supported by results of an Evoked-Related Potential study: tonal expectations modulated tone processing within the first 100 msec after tone onset, resulting in an Nb/P1 complex that differed in amplitude between tonally related and less-related conditions. The results suggest that cognitive tonal expectations can influence pitch perception at several steps of processing, starting with early attentional selection of pitch [18].

Benefits of tonal knowledge on pitch processing have also been shown for pitch structure knowledge newly acquired in the laboratory. A recent behavioural study combined implicit learning and priming paradigms [19]. Participants were first exposed to structured tone sequences without being told about the underlying artificial grammar of the sequences. They then made speeded judgements on a perceptual feature of target tones in new sequences (i.e., in-tune/out-of-tune judgements). The target tones respected or violated the structure of the artificial grammar and were thus supposed to be expected or unexpected in that grammatical framework. Results of this priming task revealed that grammatical tones were processed more quickly and more accurately than ungrammatical ones. These findings show that top-down expectations based on newly acquired structure knowledge (i.e., acquired in the lab) influence processing speed (i.e., response times) of the target tones. It remains to be shown whether these top-down expectations can go beyond this influence and are powerful enough to facilitate early perceptual processing steps (e.g., pitch processing per se) and not only processes linked to decisions and other task-related processes.

Tonal structures and listeners' knowledge thereof does not influence only performance in perceptual tasks focusing on pitch, but also performance in memory tasks, which require processing of the pitch dimension (e.g., comparing two tones or two tone sequences separated by a delay and indicating whether these are the same or different). Participants show better memory performance for tonal compared to atonal chord sequences and melodies [20-21]. Tonal sequences are tone sequences that respect Western musical regularities; atonal sequences are those that do not. The benefit of tonal structures on memorizing tone sequences has been shown for both Western nonmusicians and musicians [22]. The benefit was observed when the task required maintenance of tone information, but not when manipulation was required (comparing two sequences and judging whether they were same or different, with "same" being defined as all tones played correctly in the backward order).

However, for the simpler task (requiring only maintenance, that is by comparing both sequences with tones in the same (forward) order), the benefit of the tonal structure on short-term memory was reduced for individuals with congenital amusia who have been reported to have deficits in music perception and production. Congenital amusia is a life-long deficit of music processing without brain damage or sensory deficits. Individuals with congenital amusia have difficulties recognizing familiar tunes without lyrics and detecting an out-of-key or out-of-tune note. This musical disorder occurs despite normal performance on tests of intelligence, auditory processing, cognitive functioning, language processing, and it is not due to a lack of environmental stimulation to music (see [23-25] for extensive testing).

This condition has been described as being based on impaired processing of the pitch dimension, notably with a deficit of pitch perception (e.g., [24]), but particularly of pitch memory (e.g., [26]). When tested for short-term memory with tone sequences containing tonal structure, this population did not show the benefit of tonal sequences over atonal sequences in terms of accuracy. However a benefit was shown for response times, notably with faster response times for tonal sequences than for atonal sequences (as observed for the control participants). These findings suggest that some implicit processing of tonal structures is potentially preserved in congenital amusia, which can also influence pitch memory [27]. This observation conforms with data of other studies suggesting implicit processing of pitch despite congenital amusia (e.g., [28-29]).

This section reviewed findings for top-down influences due to listeners' tonal knowledge or newly learned tone structure knowledge based on an artificial grammar. Future research now needs to further investigate the kind of top-down influences that are driven by listeners' knowledge of linguistic structure from the language of their culture. Some research has investigated this for tone languages where the pitch dimension is crucial, with pitch carrying meaning (see next section), but no research has investigated the influence of knowledge based on context and/or whether question or statement, which can be indicated by pitch markers, will be presented.

DOMAIN SPECIFICITY OF PITCH PROCESSING IN MUSICAL AND VERBAL MATERIAL?

Regarding the debate of domain-specific or domain-general processing, some findings have suggested common pitch processing mechanisms in music and speech, notably by showing some beneficial influences (positive transfer) across domains (i.e., music and speech) due to expertise in music or in tone languages.

Musical training or expertise has been shown to improve pitch perception not only in musical contexts, but also in speech contexts. For example, musicians show improved pitch processing for prosody of non-tonal language material [30-31] and for tone-language material, such as Thai tones [32-33] and Mandarin tones [34-38]. Comparing musicians and nonmusicians is informative, but also raises the criticism that differences between the two populations might not be due to musical training, but have rather existed before starting to learn music. Longitudinal studies of musical training in the short term, e.g., within the experimental framework, have started to investigate this issue to reject the raised criticism: Nonmusician children are allocated to two groups, for example musical training versus some other kind of training (painting, drama) for several months. Comparing the performance of the children before and after training as well as between the groups after training allows investigation of the potential effects of musical training on neural correlates (anatomical, functional) and sensory and cognitive processes involved in music processing as well as in language processing. For example, Moreno et al. [39] reported that after musical training, the children of the music group processed better small pitch changes (for music and speech materials) than did the children of the painting group. And this benefit for pitch processing in speech was observed not only for their mother tongue, but also for a different, foreign language – a phenomenon that could facilitate the learning of new languages [40].

These studies have all focused on the Western tonal system and compared Western tonal musicians to nonmusicians. This thus reveals another research area where cultural investigation bias needs to be overcome (see [41] for a discussion). Notably, it would be interesting to test whether musicians who are experts for musical systems that are based on microtonal structures (that is, the octave is divided in more than 12 semitones, thus containing smaller intervals than a semitone, as for example in traditional Indian music or some African musical systems) would be even better in pitch processing for both musical and verbal materials.

Similarly to musical expertise, expertise or training in a tonal language can facilitate pitch perception and production with musical materials: Mandarin, Vietnamese and Cantonese speakers have been found to be more accurate at imitating musical pitch and discriminating intervals than English speakers ([42], see also [43]), as can be also reflected in subcortical pitch tracking (e.g., [44]). The influence of tone language background has been mostly observed for relative pitch processing (e.g., intervals, contours). Stevens et al. [45] more specifically investigated pitch contour processing in spoken Thai and English items (speech task) as well as in matched musical items for participants with tonal (Thai) and non-tonal (Australian English) language backgrounds. The overall findings suggest that expertise in tonal language leads to perceptual attunement for contour processing in spoken items as well as in musical items (even though here restricted to the speed of processing rather than extending to accuracy of processing). However, the influence of tone language background might also lead to difficulties in pitch contour processing when non-speech target sounds resemble features of linguistic tones [46]. In contrast to these results for relative pitch processing, it has been shown that listeners with tone language background did not differ from listeners with non-tone language background for absolute pitch discrimination of non-speech sounds (e.g., [42, 46]). Interestingly, in musically-trained participants, there is a link between tone language background and single pitch processing: absolute pitch (i.e., the ability to label a tone without a reference pitch) appears to be more prevalent among tone language speakers than among non-tone language speakers [47-48].

Regarding potential neural correlates of these expertise effects and their cross-domain effects, it has been proposed that musical training might shape basic sensory circuitry as well as corticofugal tuning of the afferent system, which is context-general and thus also has positive side-effects on linguistic pitch processing (e.g., [38]). Similar findings suggesting experience-dependent corticofugal tuning have been recently reported for the effects of tone language expertise on musical pitch processing [35].

In contrast to these expertise/training-related improvements of pitch processing from one domain to the other, recent research has investigated the influence of a pitch perception deficit, which has been first described for music (as in the condition of congenital amusia), on pitch perception in speech. This could be also labelled as "negative transfer" – in parallel to the positive transfer and benefit of expertise, discussed above.

As introduced above, congenital amusia has been thought to result from a musical pitch-processing disorder. Individuals with congenital amusia have impaired perception of pitch directions for pure tones [25] and for detecting pitch deviations that are smaller than two semitones in sequences of piano notes [49] as well as in note pairs [24]. Initial reports have suggested that the deficit was restricted to pitch processing in music, and did not extend to pitch processing in speech material. Individuals with congenital amusia have been reported to be unimpaired in language and prosody tasks, such as learning and recognizing lyrics, classifying a spoken sentence as statement or question based on final falling or rising pitch information (e.g., [23-24]).

However, more recent studies revealed deficits also for pitch processing in amusia – for prosody and for tone languages. Amusics showed mild deficits in processing speech intonation (questions vs. statement) or emotional prosody in their native language (English or French; [2, 50-51]), or in processing pitch contrasts in tone language words (Mandarin or Thai) - whether for native Mandarin speakers [52-54] or native French speakers [55-56]. Note that when tested with natural speech, which involves multiple acoustic cues, Mandarin people with amusia were not impaired [57-58]. Interestingly, people with amusia who are native speakers of Cantonese show better pitch processing abilities than those with amusia who are native speakers of non-tonal languages (English, French; [38]). For non-tonal language speakers, it has been shown that people with amusia performed better with speech than with musical analogues, especially for those individuals with amusia and high pitch discrimination thresholds (over one semitone), even though they were also impaired for speech material in comparison to the control participants. This data pattern was observed both for tone language material (Thai) and a single repeated syllable (/ka/), both in comparison to their non-verbal/musical analogues [7, 56]. Nevertheless, for both verbal and musical materials, French-speaking people with amusia were impaired in comparison to their control participants. In conclusion, speech may enhance pitch processing in amusia, even though it does not necessarily restore normal processing.

In addition to these positive and negative transfer effects (due to expertise (in music or tone language) or deficit (in congenital amusia)) on pitch processing in music and speech materials, pitch processing capacities have been recently linked to phonological and phonemic awareness abilities [60-61]. Some findings have led to the hypothesis that there might be shared or common neural bases for pitch-related impairments in amusia/tone-deafness and phonemic awareness (i.e., the ability to manipulate phonemes and syllables in spoken words) in dyslexia [61].

TRAINING OF PITCH PROCESSING: CROSS-DOMAIN AND CROSS-MODAL EFFECTS

The previous section has reviewed some research investigating pitch processing in both verbal and musical materials. Expertise and training have beneficial effects for pitch processing in both music and language. And even though individuals with congenital amusia have pitch-processing deficits in both domains, the deficits are less pronounced for the verbal material (at least for the individuals with amusia with higher pitch discrimination thresholds). Based on this finding, one might thus wonder in how far it might be possible to exploit this observation to train pitch processing – that is to train pitch processing with language material, aiming for an improvement of pitch processing in musical material. However, some previous findings also suggest that pitch processing/ learning is not independent in music and speech: For example, Wong and Perrachione [62] reported an association between participants' ability to learn pitch patterns in syllables, their ability to perceive pitch patterns in non-lexical contexts and their previous musical experience.

Regarding training and rehabilitation for pitch processing, another approach exploits beneficial effects of audio-visual integration. It has been previously shown that the combination of sensory information across senses can modify perception. The simultaneous presentation of an auditory signal has been shown to improve visual performance for various tasks, even when the auditory signal was not informative regarding the visual task (e.g., [63-64]). Interestingly, these benefits based on cross-modal interactions are maximally effective when the perception of one of the signals (i.e., in only one modality) is weak. Caclin et al. [64], for example, reported the benefit of an (uninformative) auditory cue on visual processing in particular for myopic participants with poor visual performance.

Another population (beyond those with amusia) with strong deficits in pitch processing are hearing-impaired patients with cochlear implants. Cochlear implants are surgically implanted devices that directly stimulate the auditory nerve in individuals with profound deafness. However, current technology is limited in transmitting spectral information, which leads to impoverished pitch processing, thus affecting both music and speech processing (i.e., prosody). Galvin et al. [65] have started to propose training programs with short musical sequences, which are presented in addition to visual cues (informative for pitch and contour). Using this kind of display combining auditory and visual information in training, has led to improved melodic contour identification tasks (also for new contours, thus showing some kind of generalization) in patients with cochlear implants.

More recently, audio-visual interactions and benefits have been exploited for testing individuals with congenital amusia. Albouy et al. [66] investigated whether audio-visual facilitation can be observed in congenital amusia, notably by presenting uninformative visual cues in addition to the to-be-processed sound sequences (requiring pitch change detection). Results revealed that individuals with amusia and control participants benefited from simultaneous visual information: accuracy was higher and response times shorter in the audiovisual condition than in the auditory-only condition. These findings suggest that individuals with amusia can benefit from multisensory integration to improve pitch processing. The results thus provide the first step towards the possibility of exploiting multisensory paradigms to help reducing pitch-related deficits in congenital amusia.

CONCLUSION

The research reviewed here investigates pitch processing for music and speech materials, with a focus on the influence of cognitive processes (related to context, learning, memory, knowledge and/or expertise). The strength of cognitive influences (based on listeners' knowledge) on pitch perception (even down to early attentional selection of pitch) has been shown in particular for musical materials. Complementary information for our understanding of pitch processing and the domain-specificity or –generality of potentially involved mechanisms has been provided by research investigating cross-domain influences of expertise (either in music or tone languages) and deficits (as in congenital amusia). Results rather point to mostly domain-general mechanisms or shared mechanisms, with some specificities in pitch processing depending on the material, which need to be further investigated. As pointed out above, most research suffers from a cultural investigation bias; this is particularly the case for music cognition research, which focuses on Western tonal music, while some research on pitch processing in speech also includes tonal languages (and not only non-tonal ones). Future research should thus open up to the investigation of pitch processing in other cultural materials and situations, as previously investigated by research on emotional connotation in music and speech (including the role of pitch) [67].

ACKNOWLEDGEMENTS

This work was supported partly by a grant from the "Agence Nationale de la Recherche" (ANR) of the French Ministry of Research ANR-11-BSH2-001-01, and conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060).

REFERENCES

[1] Patel, A. D. (2008). *Music, language, and the brain*. NY: Oxford University Press.

[2] Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, **25**, 357-368.

[3] Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music Perception*, **6**, 383-396.

[4] Moore, B. C. J. (2008). Basic auditory processes involved in the analysis of speech sounds. *Philosoph Transact Royal Soc B Biol Sci*, **363**, 947-963.

[5] Lyzenga, J., and Horst, J. W. (1995). Frequency discrimination of band-limited harmonic complexes related to vowel formants. *J. Acoust. Soc. of Am.* **98**, 1943-1955.

[6] Smith, D. R. R., Patterson, R. D., and Turner, R. (2005). The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.*, **117**, 305-318.

[7] Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umilta, C. & Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *Journal of the Acoustical Society of America*, **130**, 4089-4096.

[8] Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, **100**, 100-130.

[9] Howard, J. H., O'Toole, A. J., Parasuraman, R., & Bennett, K. B. (1984). Pattern-directed attention in uncertain-frequency detection. *Perception & Psychophysics*, **35**, 256-264

[10] Francès, R. (1958). *La perception de la musique* (2° ed.). Paris: Vrin.

[11] Bharucha, J. J. (1996). Melodic anchoring. *Music Perception*, **13**, 383-400.

[12] Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, **64**, 198-207.

[13] Tillmann B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Ann N Y Acad Sci.*, **1060**, 100-110.

[14] Marmel, F., Tillmann, B. & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, **70**, 841-852.

[15] Marmel, F. & Tillmann, B. (2009). Tonal priming beyond tonics. *Music Perception*, **26**, 211-221.

[16] Marmel, F., Tillmann, B., Delbé, C. (2010) Priming in melody perception: tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception & Performance*, **36**, 1016-1028

[17] Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of music: A Self-Organizing Approach. P*sychological Review*, **107**, 885–913.

[18] Marmel, F., Perrin, F. & Tillmann, B. (2011). Tonal expectations influence early pitch processing: Evidence from ERPs. *Journal of Cognitive Neuroscience*, **10**, 3095-104.

[19] Tillmann, B. & Poulin-Charronnat, B. (2010) Auditory expectations for newly acquired structures. *Quarterly Journal of Experimental Psychology*, **63**, 1646-1664.

[20] Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: hierarchies of stability as a function of context. *Cognition*, **13**, 63-102.

[21] Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, **50**, 305-313.

[22] Schulze, K., Dowling, W. J., Tillmann, B. (2012). Maintenance and Manipulation of tonal and atonal sequences in non-musician and musician listeners. *Music Perception*, **29**, 255-267.

[23] Ayotte, J., Peretz, I., and Hyde, K. L. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* **125**, 238-251.

[24] Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., and Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* **33**, 185-191.

[25] Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain* **127**, 801-810.

[26] Tillmann, B., Schulze, K. & Foxton, J. (2009). Congenital amusia: A short-term memory deficit for nonverbal, but not verbal sounds. *Brain & Cognition*, **71**, 259-264.

[27] Albouy, P., Schulze, K., Caclin, A. & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, **1537**, 224-32

[28] Peretz I., Brattico E., Järvenpää M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*, **132**, 1277-1286.

[29] Tillmann, B., Gosselin, N., Bigand, E. & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex*, **48**, 1073-1078.

[30] Magne, C., Schön, D. & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, **18**, 199-211.

[31] Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, **41**, 341-349.

[32] Burnham, D., & Brooker, R. (2002) Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. In *7th International Conference on Spoken Language Processing*, Denver, USA, pp. 257-260.

[33] Schwanhäußer, B. & Burnham, D. (2005) Lexical tone and pitch perception in tone and non-tone language speakers. In *Proceedings of the 9th European Conference on Speech Communication and Technology*. Bonn, Germany, pp 1701-1704.

[34] Alexander, J. A., Wong, P. C. M. & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *Proceedings of Interspeech 2005*, Lisbon, Portugal.

[35] Bidelman, G.M., Gandour, J.T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, **23**, 425-434.

[36] Delogu, F., Lampis, G. & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception, *European Journal of Cognitive Psychology*, **22**, 46-61.

[37] Lee, C-Y. & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, **124**, 325-3248.

[38] Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci,* **10**, 420-422.

[39] Moreno S., Marques C., Santos A., Santos M., Castro S., L. & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, **19**, 712-723.

[40] Marques C., Moreno S., Castro S. L., Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J Cogn Neurosci*., **19**, 1453-63.

[41] Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, **4**, 653-667

[42] Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attent Percep Psychophys*., **71**, 1385-1398.

[43] Hove, M. J., Sutherland, M. E. & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin & Review*, **17**, 310-316.

[44] Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, **25**, 161-168.

[45] Stevens, C. J., Keller, P. E., & Tyler, M. D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, **41**, 59-74.

[46] Bent, T., Bradlow, A. R. & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 97-103.

[47] Deutsch, D., Dooley, K., Henthorn, T. and Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *Journal of the Acoustical Society of America, April*, **125**, 2398-2403.

[48] Deutsch, D., Henthorn, T., Marvin, E., & Xu H-S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, **119**, 719-722.

[49] Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science* **15**, 356-360.

[50] Liu, F., Patel, A. D., Fourcin, A., et al (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain*, **133**, 1682-93.

[51] Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc Natl Acad Sci USA*, **109**, 19027-32.

[52] Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain*, **133**, 2635-42.

[53] Jiang, C, Hamm, J. P., Lim, V. K., et al (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. Neuropsychologia, **48**, 2630-9.

[54] Jiang, C, Hamm, J. P., Lim, V. K., et al (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Mem Cognit*, **40**, 1109-21.

[55] Nguyen, S., Tillmann, B., Gosselin, N., et al (2009). Tonal language processing in congenital amusia. *Ann N Y Acad Sci*, **1169**, 490-3.

[56] Tillmann, B, Burnham, D, Nguyen, S, et al (2011). Congenital Amusia (or tone-deafness) Interferes with pitch processing in tone languages. *Front Psychol*, **2**, 120.

[57] Liu, F, Jiang, C, Thompson, W. F., et al (2012). The mechanism of speech processing in congenital amusia: evidence from Mandarin speakers. *PLoS One*, **7**, e30374.

[58] Liu, F, Xu, Y, Patel, A. D., et al (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: evidence from Mandarin speakers. *Brain Cogn*, **79**, 209-15.

[59] Wong, P. C., Ciocca, V., Chan, A. H., et al (2012). Effects of culture on musical pitch perception. *PLoS One*, **7**, e33424.

[60] Jones, J. L., Lucker, J., Zalewski, C., et al (2009b). Phonological processing in adults with deficits in musical pitch recognition. *J Commun Disord.*, **42**, 226-34.

[61] Loui, P., Kroog, K., Zuk, J., et al (2011). Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia. *Front Psychol.*, **2**, 111.

[62] Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, **28**, 565-585.

[63] Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform.*, **26**, 1583-1590.

[64] Caclin, A., Bouchet, P., Djoulah, F., Pirat, E., Pernier, J., and Giard, M. H. (2011). Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Res.*, **1396**, 35-44.

[65] Galvin, J. J., Fu, Q. J. & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear & Hearing*, **28**, 301-319.

[66] Albouy, P., Leveque, Y., Hyde, K. L., Bouchet, P., Tillmann, B. & Caclin, A. (2014). Boosting pitch encoding with audiovisual interaction in congenital amusia. Manuscript submitted for publication.

[67] Thompson, W. F. & Balkwill, L-L. (2010). Cross-cultural similarities and differences. In Patrik Juslin and John Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 755-788). Oxford: Oxford University Press.

COCHLEAR IMPLANTS CAN TALK BUT CANNOT SING IN TUNE

Jeremy Marozeau, Ninia Simon and Hamish Innes-Brown
The Bionics Institute, East Melbourne, Australia
jmarozeau@bionicsinstitute.org

The cochlear implant is rightfully considered as one of the greatest success stories in Australian biomedical research and development. It provides sound sensation to hundreds of thousands of people around the world, many of whom are able to understand and produce speech. The device was developed in order to optimize speech perception, and parameters such as the choice of frequency bands and signal processing used were chosen in order to maximise perceptual differences between speech vowels. However, these settings are far from being suited for the perception of music, which might partly explain why many cochlear implant recipients cannot enjoy music through their implant.

INTRODUCTION

The cochlear implant was developed simultaneously in the sixties in France, the US and Australia [1]. Initially the cochlear implant , CI, was designed to increase the ability of the profoundly deaf to navigate through the environment by providing them with sound sensation. Due to the complexity of the cochlea it was thought that improvements in speech perception or directional hearing for future CI recipients would be of minor impact [2]. However, speech understanding outcomes improved rapidly and nowadays many CI users are able to understand words in sentences in quiet listening environments without any other aid [2, 3]. Understanding speech is important for verbal communication; however, many other sounds are also important for non-verbal communication and understanding of the auditory environment. Music is an obvious example. Listening to music forms important social bonds, and is ubiquitous in our society, from the movies to the supermarket. Music is invariably present in situations where group and community reinforcement occurs (e.g. concerts, weddings, funerals, protests, State occasions). A survey of the musical habits of over 100 CI listeners found that music was generally less enjoyable post-implantation [4]. Although there is a wide range of musical abilities among CI users, many display major limitations in their ability to perceive and enjoy music [for a review see 3, 5, 6].We will argue in this paper that the strong focus on speech processing has been detrimental to the perception of music.

THE COCHLEA

The cochlea is a spiral structure with a shape similar to the Nautilus shells sometimes found washed up on beaches, and its job is to convert mechanical vibrations within the cochlea into electrical pulses in the auditory nerve. The cochlea is embedded in the temporal bone, and in a healthy ear contains rows of hair cells lined up along its length. These hair cells stimulate auditory nerves when they are moved by vibrations in the basilar membrane, in which they are mounted. The basilar membrane has mechanical properties causing it to resonate at different frequencies along its length. The hair cells are thus set in motion at different points along the membrane depending on the frequency of the sound. If hair cells close to the middle ear vibrate, a high-pitched sound is heard. The pitch gradually gets lower as the region of hair cells that vibrate get deeper into the cochlear.

Table 1. Left panel: Electrode position (the distance from the apex in mm) for a patient with 33 mm cochlear length and an insertion of 20 mm, the corresponding frequency according to the Greenwood function. Right panel: Frequency allocations of a standard Cochlear© cochlear implant.

Distance From Apex [mm]	Frequency Predicted by Greenwood function. [Hz]	Electrode Number	Lower Freq [Hz]	Upper Freq [Hz]
13	966	22	188	313
13.9	1124	21	313	438
14.8	1305	20	438	563
15.7	1512	19	563	688
16.6	1748	18	688	813
17.5	2017	17	813	938
18.5	2326	16	938	1063
19.4	2678	15	1063	1188
20.3	3080	14	1188	1313
21.2	3540	13	1313	1563
22.1	4065	12	1563	1813
23.0	4665	11	1813	2063
23.9	5350	10	2063	2313
24.8	6133	9	2313	2688
25.7	7028	8	2688	3063
26.6	8050	7	3063	3563
27.5	9218	6	3563	4063
28.5	10552	5	4063	4688
29.4	12076	4	4688	5313
30.3	13818	3	5313	6063
31.2	15807	2	6063	6938
32.1	18080	1	6938	7938

Based on the work of the Nobel laureate Georg von Békésy, Greenwood [7] developed a function that predicts the position of maximum excitation on the cochlea as a function of the frequency of the input sound. Table 1 shows some of these relationships for a patient with a 33 mm long cochlea. For example, when a patient with a CI is presented with a 700 Hz tone, electrode 18 will be activated. Stimulation at this electrode will produce a maximum gradient potential around 16.6 mm from the apex. In normally-hearing listeners, this region is activated by a tone with a frequency of 1748 Hz.

There are around 3500 rows of hair cells along the length of the basilar membrane. If they are damaged, the auditory neurons cannot be excited, leading to a loss of the sensation of hearing.

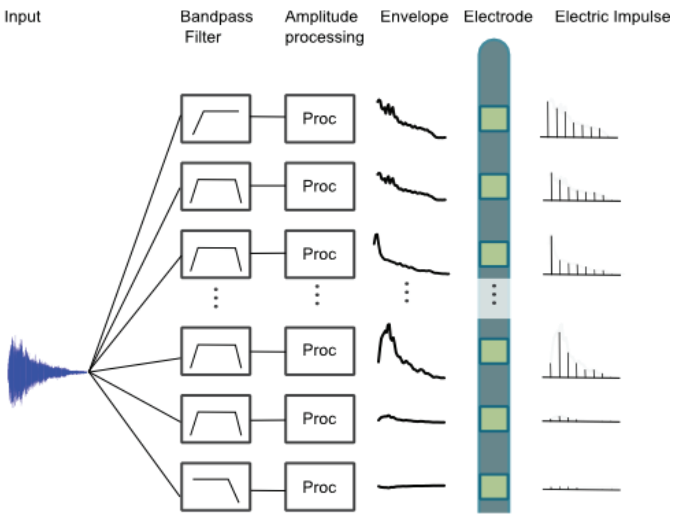


Figure 1. Simplified block diagram of a cochlear implant. First the sound is decomposed into frequency bands. Then the output of each band is processed through compression and an envelope detector. A series of electrical impulses at fixed rate modulated with the envelope will be produced on selected electrodes. The frequency allocation of each electrode is predetermined by the frequency of each channel (represented in Table 1 for Cochlear Device.)

THE COCHLEAR IMPLANT

A cochlear implant largely replaces the function of the outer, middle, and most of the inner ear – up to the level of the auditory nerve. It consists of two main parts. First, the sound processor is worn externally, and hooks behind the ear. It contains microphones, batteries, and a miniaturized computer system that converts the acoustic signal received at the microphones into a series of electric pulses according to a programmable software algorithm called a 'strategy.' Second, the implant itself is implanted in the mastoid bone behind the ear. It receives power, as well as the electrical signals from the sound processor via a wireless link through the skin. At the end of the implant is a very fine linear array of up to 22 electrodes, which is inserted about half-way into the spiral-shaped cochlea. These electrodes stimulate the auditory nerve, thus replacing the function of the hair cells that are lost or damaged in sensorineural deafness. When an electrode is activated, it delivers a series of biphasic pulses, normally with phase durations of 25 µs and an 8-µs inter-phase gap.

The strategy embedded in the sound processor determines which combinations of electrodes to stimulate according to the acoustic signal received by the microphone. Figure 1 shows a simplified block diagram of the cochlear implant. The most commonly used strategy divides the incoming sound signal into as many frequency bands as there are electrodes (22 in the Cochlear Ltd Nucleus devices), selects a small number of the bands with the highest amplitude (typically the 8 highest of the total 22 available), and then stimulates those electrodes at a current level related to the smoothed amplitude in each band.

If a high-frequency pure tone is played, about 1400Hz for example, electrode #13 is stimulated (see Table 1). The audiologist may change the frequency allocations of each electrode individually. However, in a typical clinical session, the allocation will only be changed in case of dysfunctional electrodes.

SPEECH SIGNALS

Speech signals convey semantic meaning though a rapid succession of vowel and consonant sounds. Vowel sounds (such as the /ɪ/ in 'heed', /ɛ/ in 'head', /æ/ in 'had', /ʌ/ in 'hud', /ɒ/ in 'hod', /ɔ/ in 'horde', /ʊ/ in 'hood', /u/ in 'who'd') are produced without significant constrictions in the vocal tract, and are generally "voiced" – that is the vocal folds vibrate and produce a harmonic sound. Vowels in English and other non-tonal languages generally have a fairly consistent voicing frequency (F0), with a unique pattern of harmonics called formants, labelled F1, F2, F3 etc. Depending on the vowel sound produced, the first formant in English varies between 300-770 Hz, and the second between 900-2300 [8, 9]. Most vowel sounds can be distinguished by the first and second formants alone [10].

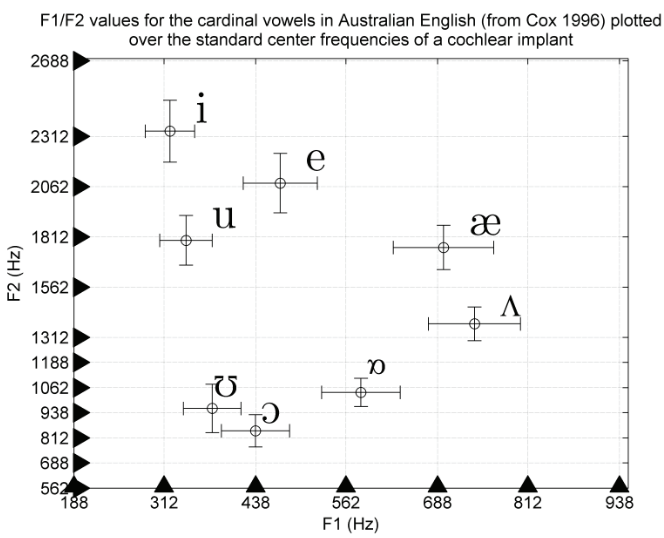


Figure 2. Eight vowels plotted according to the frequency of their first and second formants (F1, F2). The grid overlayed corresponds with the edge of the electrode frequency bands specified in a default CI map (see Table 1).

In a CI, the steady first formant activates one or more of the lowest electrodes, and a number of the higher electrodes are activated by the higher formants – with a different pattern activated for each vowel sound. Thus the CI user receives a fairly unique pattern of electrode activation for each vowel sound. Figure 2 shows the F1 and F2 frequencies for some of the cardinal vowel sounds in Australian English with a male speaker [11], overlaid on a grid representing the edges of the default CI frequency bands. For example the vowel sound "i" has a first formant around 320 Hz and a second around 2320 Hz. It will therefore activate mostly electrodes #22, #21, #10 and #9. Crucially, speech can be understood with a relatively small number of vowel sounds, so that despite the problems of overlapping filter bands and current spread, there is enough frequency resolution using 22 electrodes for many CI users to successfully distinguish between many of the vowels [12, 13]. It is therefore important to allocate frequency bands to electrodes in order to maximise the difference between vowels. As shown in Table 1, a typical frequency allocation table serves that purpose. The low-frequency bands (#22 to #14) increase linearly with a fixed width of 125 Hz. The bandwidth of the higher frequency bands increases logarithmically and reaches 1 kHz for the highest band (#1). The lowest 5 electrodes (inserted deeply) are associated with the first formants of most vowels, the middle 8 electrode to the second formants, and the highest to the third formants and other high pitched sounds.

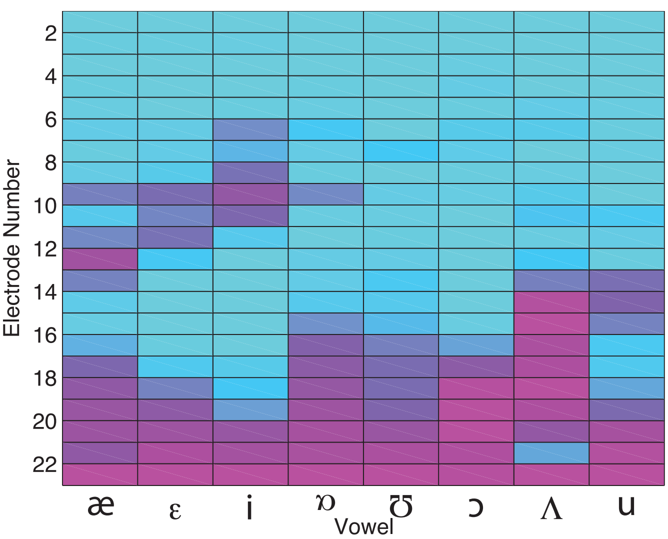


Figure 3. shows the activation pattern of 8 Australian vowels averaged across time. The ordinate represents the average activation of each of the 22 electrodes. The colour codes the amplitude of the activation.

Figure 3 shows that each vowel produces a specific pattern of activation across electrodes that can be learned by the CI recipient. The vowel /ɔ/ will mostly activate the lower frequency electrodes (22 to 17). The vowel /ɪ/ clearly shows the activation of two zones corresponding to the two formants. Compared to vowel sounds, which are distinguishable mostly on the basis of spectral information, consonants are mostly distinguishable on the basis of how the overall amplitude varies in time [14]. The rate of stimulation pulses in CIs can vary from around 200 Hz up to 1200 Hz. At these rates, gross temporal cues can be transmitted fairly well. Thus, despite the complex acoustic nature of consonant sounds, many of the time-based cues used to distinguish between consonant sounds are successfully transmitted to the listener [15].

MUSICAL PITCH

Most musical sounds share these same basic features; spectral parameters encode pitch, melody, and tonal aspects of timbre while time-varying parameters encode rhythm and impulsiveness aspects of timbre. However, musical signals are acoustically more complex than speech. Unfortunately, the signal processing employed in most standard CIs destroys many of the acoustic parameters in the signal, only passing the smoothed amplitude envelopes of a series of band pass filters.

The perception of pitch is based on the fundamental frequency (F0) of an acoustic signal. It is not completely clear how pitch is coded in the auditory system, but research so far points to the conjunction of three physiological cues. First, as described in the above section, different auditory nerves are stimulated depending on the frequency of the acoustic signals. Therefore, frequency information can be transmitted to the brain by detecting which auditory nerves have been activated. This cue is called **place coding**. Second, the basilar membrane within the cochlea resonates, and therefore triggers the auditory nerves at a rate related to the input frequency (at least up to about 1-4 kHz). This temporal pattern of neural firing can also convey pitch information. This is called **temporal coding**. Third, as the high frequencies excite a portion of the membrane located at the entrance of the membrane, and the low frequencies a portion at the end of the membrane, the delay of excitation will be different according to the frequency – the low frequencies will be delayed by the time needed to travel along the cochlea. Therefore, pitch information can also be conveyed through the timing of activation of the nerves (the high frequencies will arrive first). For example, the delay of a 200 Hz tone is about 6-ms longer than the delay of an 8 kHz tone [16]. This is called **phase coding**.

In current sound processing strategies, pitch information is, for the most part, conveyed by temporal cues (amplitude modulation) and place coding, as different electrodes are activated according to the frequency. It might be possible in the future to introduce more electrodes; however, due to the spread of current, it is unclear if this will improve the frequency resolution [17]. Furthermore, as the frequency allocations are not designed to convey musical pitch, the electrical output of the sound processor cannot accurately reproduce a musical scale. Figure 4 shows the activation pattern of all 22 electrodes for a piano note as a function of its fundamental frequency. As the lowest frequency band (electrode 22) ranges from between 188 and 313 Hz, the fundamental frequency of any note one octave below Middle C will not be transmitted. The activation will only be caused by higher harmonics. When the fundamental frequency reaches this octave, the fundamental frequency will start to activate the lowest electrode. Therefore, an increase in the lower musical note will not always be associated with an increase in position of the electrodes activated. This will negatively impact the perception of lower pitch notes.

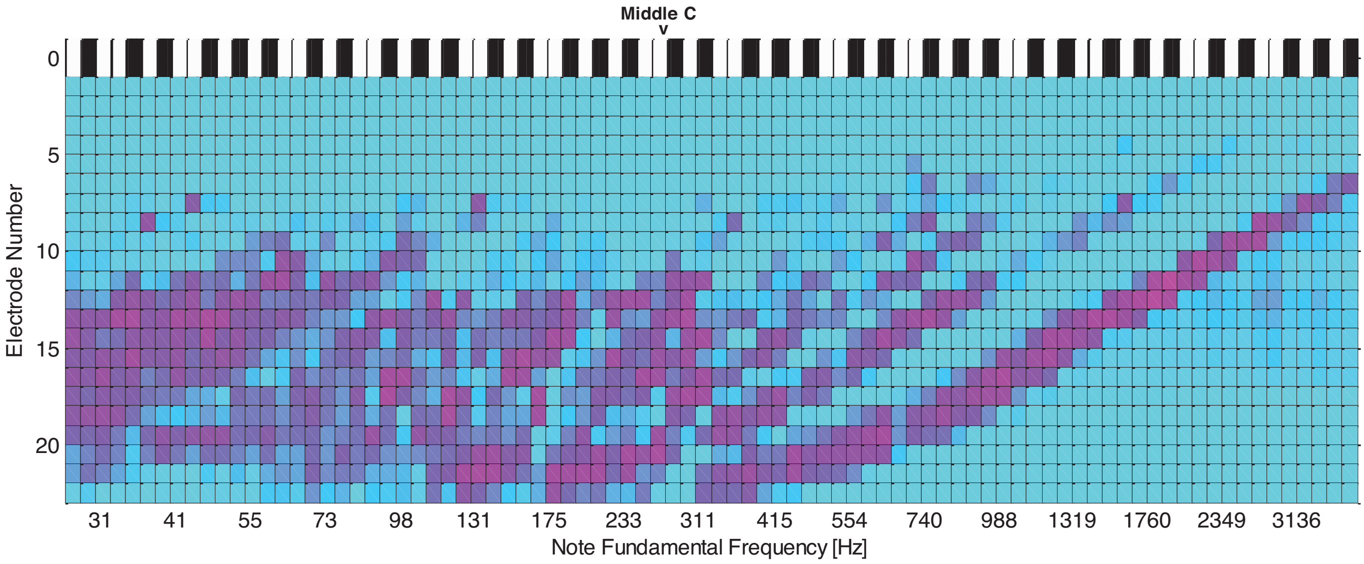


Figure 4. Average electrical excitation diagram of a piano note as function of the fundamental frequency (in Hz on bottom axis, or as piano key top axis). Red, hot colour represents a strong activation, and blue, cold colour, no activation.

In most current CI recipients, the pulse rate is fixed at 900 Hz and therefore amplitude modulation for frequencies lower than half the pulse rate can be transmitted. Some CI recipients use this temporal cue to perceive the pitch of the lower octaves of the piano, where no place cues are available (as showed in Figure 4) [3]. It is possible to increase the pulse rate, however this does not appear to improve pitch perception [18], but does decrease battery life.

Finally, in most current strategies the phase delay is not implemented, so recipients cannot benefit from this cue. Experimental strategies have been tested to determine whether the addition of a phase delay will improve speech perception. Results have found small but significant improvement for speech perception in noise [19].

In summary, CIs only partially convey two out of the three main pitch cues. This explains their poor results in pitch discrimination tasks. A study has shown that most CI recipients could not reliably identify a pitch direction change below three semi-tones or that only 20% could identify a well known melody without rhythm cues [20].

When comparing Figure 3 and Figure 4, it is clear that the frequency allocations of the bands are well suited for speech: every vowel produces a specific pattern of the activation. On the other hand there is no clear increase in activation pattern with successive musical notes below Middle C.

THE INTERACTION BETWEEN SPEECH AND MUSIC IN COCHLEAR IMPLANT RECIPIENTS

We have argued that the strong focus on speech in the development of the cochlear implant was detrimental to the perception of music. However, in some cases speech signals have been shown to enhance music perception.

Experiments using familiar musical items revealed that verbal cues increase CI users' ability to positively identify the musical material [21-24]. Indeed, it was reported that some CI listeners were able to extract linguistic information from sung lyrics [25] and correlations were found between melody recognition with lyrics and speech perception [21].

On the other hand, vowel identity and pitch can be conveyed by the same cues (electrode position), therefore, speech signals could have a detrimental effect on music. Vowels sung to the same pitch should be distinguishable through their differences in spectral shape. As the formants of each vowel will activate different electrodes, this might be perceived as a difference in timbre. For example, a transition from the word "head" to "hod" might be confused with an increase of pitch instead of a change of vowel.

CIs and hearing-aids were developed to assist their users with speech perception and understanding. Advanced sound processing algorithms now include a sound classification system for the automatic recognition of the acoustic environment [26]. Hearing devices are now programmed to automatically distinguish between sound classes such as 'clean speech', 'speech in noise', 'noise', and 'music'. When a speech in noise situation is detected, the algorithms supress the noise and amplify speech signals. Unfortunately, these automatic algorithms can often mis-classify music as speech in noise. This can lead to an inappropriate processing of music.

STRATEGIES DEDICATED TO IMPROVE MUSIC PERCEPTION IN COCHLEAR IMPLANT

As a cochlear implant that can propose reliable speech and music will be an important commercial success, many new sound processors strategies have been tested [6]. Some strategies tried to improve the amount of temporal information conveyed by enhancing the amplitude modulation of each electrode [for example 27, 28], some others by adapting the rate of stimulation to the F0 of the signal on each electrode [for example 29, 30] or by increasing the overall rate of stimulation (such as the HiRes strategy of Advanced Bionics or FSP strategy of Medel devices). Unfortunately none of those strategies succeed in bringing music perception to a satisfactory level [6].

CONCLUSIONS

Although the cochlear implant has restored the communication abilities for hundreds of thousands of people around the world, it has short-comings in music perception. Given the complexity and very restricted size of the cochlea, no hearing device will be able to restore hearing perfectly, at least not with the current technology. It is natural that speech will be the primary focus while developing the cochlear implant. However, a cochlear implant that will be designed specifically for music perception might be possible, but can be unfavourable to speech perception.

ACKNOWLEDGEMENTS

The Bionics Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program. Ninia Simon was financially supported by a scholarship from the German Academic Exchange Service. Portions of this material were published in the Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM); Innes-Brown H, Au A, Stevens C, Schubert E & Marozeau J (2012). New music for the Bionic Ear: An assessment of the enjoyment of six new works composed for cochlear implant recipients, where HIB's attendance was funded by the Harold Mitchell Travelling Fellowship 2012.

REFERENCES

[1] Ramsden, R.T., History of cochlear implantation. *Cochlear Implants International*, 2013. **14**(S4): p. 3-5.

[2] Galvin, J.J., Q.J. Fu, and R.V. Shannon, Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, 2009. **1169**(1): p. 518-533.

[3] McDermott, H.J., Music perception with cochlear implants: a review. *Trends in Amplification*, 2004. **8**(2): p. 49-82.

[4] Looi, V. and J. She, Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *International journal of audiology*, 2010. **49**(2): p. 116-128.

[5] Limb, C.J. and J.T. Rubinstein, Current research on music perception in cochlear implant users. *Otolaryngologic Clinics of North America*, 2012. **45**(1): p. 129-140.

[6] McDermott, H., Music Perception, in Auditory Prostheses: New Horizons, A.N.P. Fan-Gang Zeng, Richard R. Fay, Editor 2011, Springer: New York. p. 305-339.

[7] Greenwood, D.D., A cochlear frequency□position function for several species—29 years later. *The Journal of the Acoustical Society of America*, 1990. **87**(6): p. 2592-2605.

[8] Hillenbrand, J., et al., Acoustic characteristics of American English vowels. *J Acoust Soc Am*, 1995. **97**(5 Pt 1): p. 3099-111.

[9] Cox, F., The acoustic characteristics of/hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, 2006. **26**(2): p. 147-179.

[10] Pols, L.C., L.J. van der Kamp, and R. Plomp, Perceptual and physical space of vowel sounds. *J Acoust Soc Am*, 1969. **46**(2): p. 458-67.

[11] Cox, F., Vowel change in Australian English. *Phonetica*, 1996. **56**(1-2): p. 1-27.

[12] Eddington, D.K., Speech discrimination in deaf subjects with cochlear implants. *Journal of the Acoustical Society of America*, 1980. **68**(3): p. 885-91.

[13] Blamey, P.J., et al., Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *Journal of the Acoustical Society of America*, 1987. 82(1): p. 48-57.

[14] Diehl, R.L., A.J. Lotto, and L.L. Holt, Speech perception. *Annual review of psychology*, 2004. **55**: p. 149-79.

[15] Shannon, R.V., et al., Speech recognition with primarily temporal cues. *Science*, 1995. **270**(5234): p. 303-4.

[16] Donaldson, G.S. and R.A. Ruth, Derived band auditory brain□ stem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects. *The Journal of the Acoustical Society of America*, 1993. **93**(2): p. 940-951.

[17] Shannon, R.V., Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hearing Research*, 1983. **12**(1): p. 1-16.

[18] Vandali, A.E., et al., Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and hearing*, 2000. **21**(6): p. 608-24.

[19] Taft, D.A., D.B. Grayden, and A.N. Burkitt, Across-frequency delays based on the cochlear traveling wave: enhanced speech presentation for cochlear implants. *IEEE transactions on bio-medical engineering*, 2010. **57**(3): p. 596-606.

[20] Gfeller, K., et al., Accuracy of Cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing*, 2007. **28**(3): p. 412-423.

[21] Leal, M.C., et al., Music perception in adult cochlear implant recipients. *Acta Otolaryngol*, 2003. **123**(7): p. 826-35.

[22] Fujita, S. and J. Ito, Ability of nucleus cochlear implantees to recognize music. *Ann Otol Rhinol Laryngol*, 1999. **108**(7 Pt 1): p. 634-40.

[23] Looi, V. and J. She, Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *Int J Audiol*, 2010. **49**(2): p. 116-28.

[24] Gfeller, K., et al., The effects of musical and linguistic components in recognition of real-world musical excerpts by cochlear implant recipients and normal-hearing adults. *J Music Ther*, 2012. **49**(1): p. 68-101.

[25] Gfeller, K., et al., The impact of voice range and instrumental background accompaniment on recognition of song lyrics. *7th Asia Pacific Symposium on Cochlear Implants and Related Sciences*, 2009: p. 42-46.

[26] Allegro, S., M. Büchler, and S. Launer. Automatic sound classification inspired by auditory scene analysis. in *Proc. of EUROSPEECH 2001, 7th European Conference on Speech Communication and Technology*, 2001, Aalborg, Denmark

[27] Laneau, J., J. Wouters, and M. Moonen, Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurotology*, 2006. **11**(1): p. 38-52.

[28] Vandali, A.E., et al., Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies. *Journal of the Acoustical Society of America*, 2005. **117**(5): p. 3126-38.

[29] Grayden, D.B., et al. A cochlear implant speech processing strategy based on an auditory model. *Proc. of Intelligent Sensors, Sensor Networks and Information Processing Conference* 2004. Melbourne.

[30] Li, X., et al., Improved perception of music with a harmonic based algorithm for cochlear implants. Neural Systems and Rehabilitation Engineering, *IEEE Transactions* on, 2013. **21**(4): p. 684-694.

ACOUSTICS FORUM

Elizabeth Beach
National Acoustic Laboratories, Australian Hearing Hub, Macquarie University, NSW, Australia

INTRODUCING THE NOISE DATABASE

Over the past several years researchers from the National Acoustic Laboratories (NAL) and the HEARing CRC have been measuring noise levels at some of Australia's noisier recreational environments using personal noise dosimeters. Recently, these measurements have been collected together to form the NOISE (Non-occupational Incidents, Situations and Events) Database, which is available online via a new website: http://noisedb.nal.gov.au/.

The data are organised into seven main categories: Attendance at sports venues, Active recreation and sport, Arts and cultural activities, Attendance at entertainment venues, Travel, Domestic activities, and Other. These, in turn, are organised into various subcategories. Measurements come from a wide range of venues, including inner-city nightclubs, large concert venues and sporting stadia, international race tracks, the New York subway, and local school discos!

Researchers, acousticians, policy makers and other people interested in knowing more about the hearing health risks of recreational noise are invited to search the NOISE database and consider whether they have similar research data they could contribute to the database. Currently there are almost 900 measurements on the database and we hope to expand this over time.

There are two main features of the website. There is a search function, in which you can obtain a list of noise measurements in a particular category and subcategory. If one conducts a 'simple' search, the output will be shown onscreen. The database computes the average noise level (in L_{Aeq}) and exposure (in Pa^2h) and provides a list of measurements. If one conducts an 'advanced' search, more detailed output is provided in a spreadsheet and users can specify whether they want to restrict the output to just NAL data or all data; data from Australia or elsewhere; and data from calibrated equipment or all types. The second main feature of the NOISE database is an upload function, in which users can contribute their own data. Contributors are invited to format their data using the instructions provided online and then upload it via the website. Once the data are received, they will be verified, and if appropriate, the data will be added to the NOISE database.

It is hoped that the NOISE database will be a useful source of leisure noise measurements that are reliable, up-to-date and accurate. It is intended to support acousticians, researchers, policy makers and hearing health advocates to identify the leisure activities that pose a real risk to hearing health, so that we can make positive steps towards reducing noise exposure in these environments.

We look forward to welcoming you to the NOISE Database.

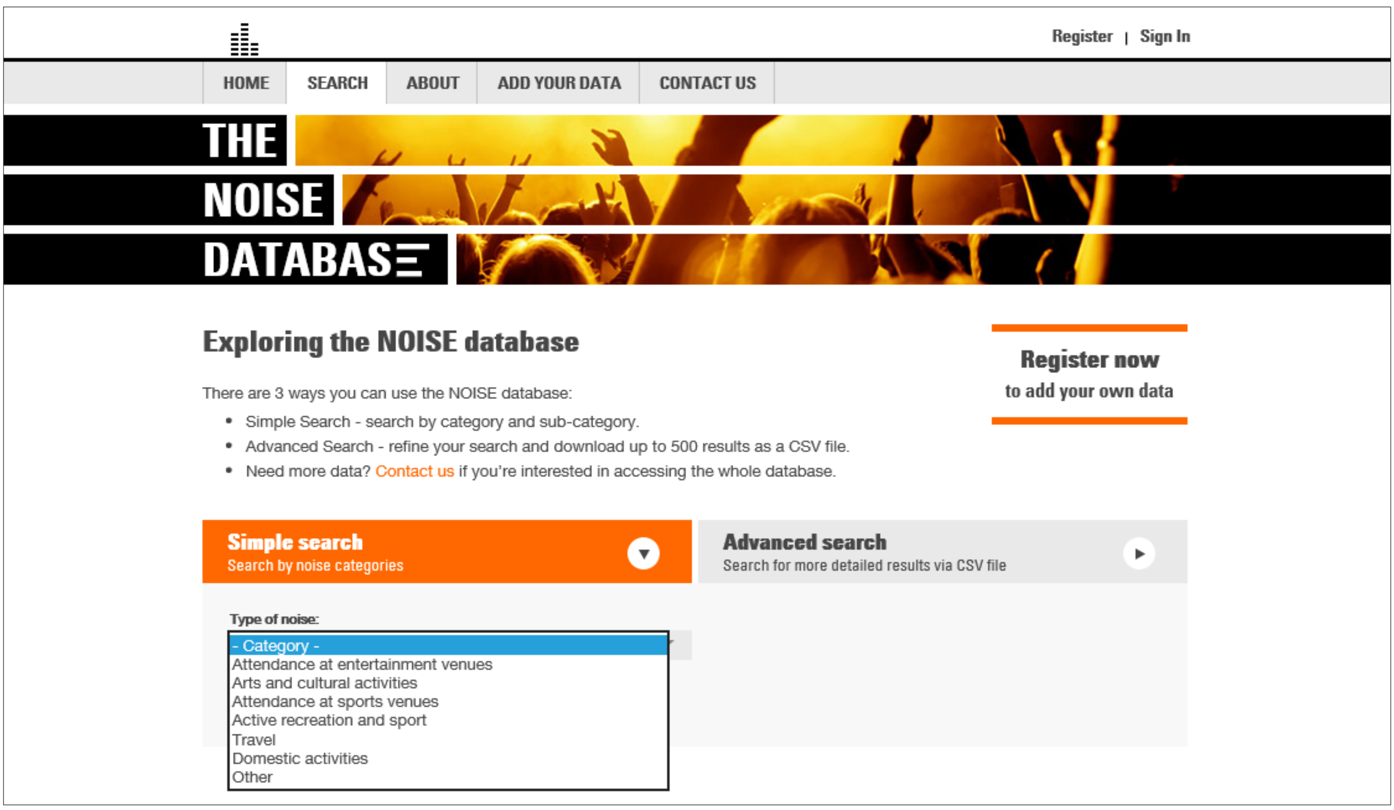


Figure 1 Search screen of the NOISE database.

LETTER TO THE EDITOR

Henri Boutin, John Smith and Joe Wolfe
School of Physics, University of New South Wales, Sydney

INFRASOUND SENSITIVITY AND HARMONIC DISTORTION

Human sensitivity to infrasound (IS) and the problems that it may or may not cause have become topical recently because of wind farms and both have been discussed here [1-5]. Determination of sensitivity to IS and assessment of its effects are complicated for several reasons.

One complication is that some common sources of infrasound with significant intensity, such as wind turbines and surf beaches, also emit power in the audible range, and that audible power is modulated at one of the IS frequencies. This makes the source audible and readily identified. If the source is regarded as an unattractive rather than a desirable feature, this modulated audible sound may become a nuisance and so could possibly trigger reactions that are not directly due to the IS component. (In contrast, one of the authors can report that three decades of living opposite a surf beach seem to have had no negative effects.)

Another complication is the difficulty of determining human sensitivity to IS. For low frequencies and high power, loudspeakers usually have harmonic distortion of one percent (–20 dB) or more. Because the sensitivity of the human hearing increases dramatically with increasing frequency below about 100 Hz, even small amounts of harmonic distortion of pure IS tones could produce audible distortion at higher harmonics.

In a recent research project studying nonlinear acoustical effects in a completely unrelated field, the present authors desired a powerful, low frequency source with very low harmonic distortion. This led us to develop a method of cancelling harmonic distortion and intermodulation products using an iterative method [6]. To demonstrate this, we used a loudspeaker that, at 107 dB, initially produced 10% harmonic distortion when driven with an input sine wave of 52 Hz (chosen to avoid confusion with harmonics of mains noise). This iterative method reduced the harmonic distortion over the audible range to less than –65 dB. One of the purposes of this letter is to draw this method to the attention of researchers seeking pure, low frequency sine waves for psychophysical measurements.

Henri Boutin, John Smith and Joe Wolfe
School of Physics, University of New South Wales, Sydney

REFERENCES

[1] V.V. Lenchine, "Amendment for wind farms environmental noise guidelines", *Acoustics Australia*, **37**, 24-25 (2009)
[2] S. Cooper, "Wind farm noise – an ethical dilemma for the Australian Acoustical Society?", *Acoustics Australia*, **40**, 139-142 (2012)
[3] P. Miskelly, Response to S. Cooper, *Acoustics Australia*, **41**, 139 (2013)
[4] P.Alway, "Wind farm noise", *Acoustics Australia*, **41**, 195 (2013)
[5] A. Bell, "Annoyance from wind turbines: role of the middle ear muscles", *Acoustics Australia*, **42**, 57 (2014)
[6] Boutin, H., Smith J. and Wolfe, J. "Producing undistorted acoustic sine waves" *J. Acoust. Soc. America.* **135**, 1665-1667 (2014)

NEWS

Draft Revision of AS 2021

The public review period for the draft revised version of AS 2021 "Acoustics - Aircraft noise intrusion - Building siting and construction" (DR AS 2021:2014) commenced on 8 August and closes on 10 October 2014. This Standard was prepared by the Australian members of Standards Australia Committee EV-011, Aircraft and Helicopter Noise, to supersede AS 2021-2000. After consultation with stakeholders, Standards Australia decided to develop this Standard as an Australian Standard rather than an Australian/New Zealand Standard. AS 2021 provides guidance on the siting and construction of buildings in the vicinity of airports to minimize aircraft noise intrusion. The assessment of potential aircraft noise exposure at a given site is based on the Australian Noise Exposure Forecast (ANEF) system (for processes and details of this system refer to Appendices A and B).

This revised edition provides expanded aircraft noise tables and incorporates various associated amendments to the text. A new Appendix has been added to describe the process that should be followed in producing an Australian Noise Exposure Forecast (ANEF) chart for use in applying this Standard. To access the free draft for public review go to the Standards Australia homepage and click the link to view "Draft Standards Open for Public Comment" and with some persistence you will eventually get to download the public review draft of AS 2021.

VIC EPA Policy Review

The EPA Victoria and the Department of Environment and Primary Industries have commenced a review of the two State Environment Protection Policies for noise: State Environment Protection Policy (Control of Noise from Industry, Commerce and Trade) No. N-1 (SEPP N-1) and State Environment Protection Policy (Control of Music Noise from Public Premises) No. N-2 (SEPP N-2).

SEPP N-1 applies to commercial and industrial premises, such as factories, shops, restaurants, farms, mines and quarries. The review of the noise SEPPs includes consideration of the approach to regulating commercial and industrial noise in regional Victoria. Currently, SEPP N-1 only applies in metropolitan Melbourne, while commercial and industrial noise outside metropolitan Melbourne is covered by the EPA guideline Noise from industry in regional Victoria (NIRV). SEPP N-2 applies to music noise from public premises such as pubs, nightclubs, outdoor entertainment venues, restaurants, public halls, gymnasiums and retail stores. SEPP N-2 covers all of Victoria.

To support the first stage of public consultation, a discussion paper has been released and is open for comment until 15 October 2014. The discussion paper, a short survey and further information about the noise SEPPs review can be found on EPA Victoria's website. (www.epa.vic.gov.au/our-work/setting-standards/environmental-standards-reform/noise). A number of workshops, seminars and engagement activities are planned for Melbourne and regional Victoria later in 2014. Details about these will be made available on EPA Victoria's website where you will also be able to register your interest to attend.

Assessing Vibration Update

NSW EPA published "Assessing Vibration: a technical guideline" in March 2006. Since that time, the British Standard, on which much of the guideline was based, has been updated. The principal changes made to this Standard include moving to a geocentric (rather than basicentric) coordinate system and changing the vertical axis weighting system from W_g to W_b. After reviewing the comments that are due by the end of July, NSW EPA plans to publish "Assessing Vibration: a technical guideline application note" to include the changes. www.epa.nsw.gov.au/noise/index.htm.

Validation of Inversion Strength Estimation Method

NSW EPA has published the results of a project looking into the measurement and estimation of atmospheric temperature inversions. The project found that significantly different inversion strength estimates can be calculated depending on the heights from which the temperature values are taken. This has implications for environmental regulators, community and industry. A number of policy, technical and regulatory recommendations that EPA is considering are contained in a discussion paper prepared by EPA. www.epa.nsw.gov.au/noise/140011invstrength.htm EPA is keen to receive feedback on the recommendations, in particular the suggestion to move towards F stability Class from 3degC/100m as the default inversion strength in the Industrial Noise Policy www.epa.nsw.gov.au/noise/industrial.htm. EPA is also interested in receiving comments on standardising the method of directly measuring temperature inversions. Submissions to larry.clark@epa.nsw.gov.au before 12 September 2014.

Developments and Music Venues

Residential developers building next to existing live music venues will have to foot the bill for sound-proofing under new planning changes to be introduced by the Victorian Government. The "agent of change" principle, which will also apply to new music venues, will put the responsibility of noise mitigation on the new development, not the existing businesses and residents nearby. Planning Minister Matthew Guy said once the new regulations are in place, established pubs and clubs will not be forced to close due to noise complaints from new neighbours. He said the reverse would also apply to new live music venues opening their doors near existing residential areas.

UK Wind Turbine Noise Notes

The UK Institute of Acoustics (IOA) has published four supplementary guidance notes to its Good Practice Guide to the Application of ETSU-R-97 for the Assessment and Rating of Wind Turbine Noise. The notes cover: sound power level data, wind shear, post completion measurements and noise propagation over water for on-shore wind turbines. The Institute plans to publish the final two notes (data collection and data processing and derivation of ETSU-R-97 background curves) shortly. The publication follows an extensive consultation exercise. The guidance notes are available from www.ioa.org.uk/publications/good-practice-guide

Good Practice Guide on Quiet Areas

On 30 April 2014, the European Environment Agency published its 'Good practice guide on quiet areas'. Noise pollution is a growing concern in Europe. Of particular importance is noise from transport and industrial sources, which are addressed by Directive 2002/49/EC relating to the assessment and management of environmental noise, otherwise known as the Environmental Noise Directive (END). The European legislation aims to reduce noise pollution and also highlights the need to preserve areas that are currently unaffected. These so called quiet areas may be found, not only in rural areas, but also inside our busiest cities. This good practice guide offers a digest of actions from all across Europe to identify and protect environments with good acoustic quality. It has been drawn up to help policymakers, competent authorities and any other interested parties understand and fulfil the requirements of the END. The guide makes recommendations based on examples of good practice in assessing and managing quiet areas in Europe. www.eea.europa.eu/publications/good-practice-guide-on-quiet-areas.

NZ Transport Agency News

The NZ Transport Agency has finalised its Guide to state highway road surface noise www.nzta.govt.nz/resources/road-surface-noise/index.html

The NZ Transport Agency is trialling a noise camera to manage engine braking noise in Tauranga. www.nzta.govt.nz/about/media/releases/3191/news.html

Good Practice Guide for Underwater Noise Measurement

The National Physical Laboratory has released a guide in March 2014 that presents best practice for in-situ measurement of underwater sound, data processing and the reporting of the measurements using appropriate metrics. It addresses the need for a common approach and is designed for those making in-situ measurements of underwater sound, for example consultants, offshore developers, oil and gas companies, and developers of marine renewable energy, as well as regulators wishing to base their requirements on a firm scientific foundation. www.npl.co.uk/publications/guides/good-practice-guide-for-underwater-noise-measurement

Traffic Noise in Cities and Health

A forum organised on behalf of the Council of the CAETS, International Council of Academies of Engineering and Technological Sciences, was held following the Internoise 2013 in Innsbruck. This CAETS forum was a follow-up to the 2008 June CAETS workshop on the design of low-noise transportation vehicles that was held at the Institute of Sound and Vibration Research, Southampton, U.K. The task for this forum was to find out what is possible to achieve with the best of today's known technology and planning instruments to improve the acoustic environment in our major cities. How much can be accomplished if all actors do their best? What is possible to substantially improve the acoustic environment? What are the lead times? The report on this forum "Lessening the Severe Health Effects of Traffic Noise in Cities by Emission Reductions" prepared by Tor Kihlman, Wolfgang Kropp, and William Lang is available from: www.ta.chalmers.se/downloads/open/intro/QuieterCities.pdf

SLR Canberra on the Move

SLR Consulting Australia has recently relocated its Canberra office to the heart of the Canberra CBD, in closer proximity to local clients. "Additionally, by moving to a more central location, all staff are committed to using public transport, walking or riding to work, therefore reducing our office's carbon footprint and improving the environment we live in." commented SLR Canberra Office Manager, Jason Watson.

The new SLR Canberra contact details are: Suite 3, Level 4, 11 London Circuit, Canberra ACT 2600. Ph: 02 6287 0800

Workplace Health and Safety News

Revision of AS/NZS 1269.4

Standards Australia published a revised edition of AS/NZS 1269.4 Occupational noise management Part 4: Auditory assessment on 24 April 2014. The main change from the 2005 edition is the replacement of the table of “maximum acceptable background noise levels” required for audiometric testing using particular makes and models of audiometer earphones/enclosures, with a method to calculate what is now called “maximum permissible ambient noise levels” for any earphone/enclosure combination for which valid attenuation data is available. This allows new makes and models of earphones to be considered for use, whilst an informative appendix still lists the ambient noise levels required for existing makes and models.

Sound in Healthcare Facilities

A report on a Swedish Symposium on Sound in Healthcare Facilities can be found at www.careforsound.com/Publications/Careforsound-publication.pdf

Excellence in Hearing Loss Prevention Awards

The winners of the US The 2014 Safe-in-Sound - Excellence in Hearing Loss Prevention Awards were announced in March and can be viewed at: www.safeinsound.us/winners.html

Evaluation of Smartphone Measurement Applications

Researchers at the US NIOSH have published a paper “Evaluation of smartphone measurement applications”. See http://blogs.cdc.gov/niosh-science-blog/2014/04/09/sound-apps/. Their results showed that, for A-weighted sound level measurements, three of the ten iOS apps tested had mean differences within ± 2dB(A) of the reference measurements and so could potentially be used for preliminary surveys in workplaces. Note though that performance in variable noise and impulsive noise has not been tested yet.

UK Fines for Hand Arm Vibration

A company responsible for maintaining the grounds of a naval base in Cornwall UK has been fined £10,000 and ordered to pay £10,000 in costs after three workers were diagnosed with Hand Arm Vibration Syndrome and/or Carpal Tunnel Syndrome, debilitating conditions that left them with permanent nerve damage. They had been exposed to high levels of hand-arm vibration caused by using tools such as hedge cutters and trimmers for long periods.

http://press.hse.gov.uk/2014/company-in-court-after-workers-suffer-nerve-damage/

Some Recent Findings on Noise Effects

A report found that for each dB of hearing loss, a statistically significant risk of increased hospital admission for injury was observed. A statistically significant association was also found between working in an occupational ambient noise ≥100 dB(A) and the risk of injury. See abstract at: www.ncbi.nlm.nih.gov/pubmed/24639292

Results indicated that prolonged duration of noise exposure (≥36.5 years) was associated with an increased risk of Cardio Vascular Disease death, as compared with shorter duration (<27 years), and that moderate and severe NIHL were also associated with an increase in risk of CVD death. See abstract: www.ncbi.nlm.nih.gov/pubmed/24792922

Researchers Li et al from the US National Institutes of Health have recently published a paper connecting hearing impairment and depression, ‘Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005-2010’. See abstract at: www.ncbi.nlm.nih.gov/pubmed/24604103.

AAS NEWS

From the General Secretary

There have been 29 applications for membership to AAS since January this year, with most being approved. Members who have not received a subscription invoice for membership for 2014-2015, please contact: General Secretary GeneralSecretary@acoustics.asn.au

AAS Research Grant

With funding for research declining in Australia, AAS has made available funds for this purpose to help achieve the objectives of the Society. The AAS Federal Council has authorised a Research Grant Plan for the AAS Research Grants. The Plan identifies key research needs (projects) and provides that information to AAS members as guidance while they develop research projects. The Plan is not meant to take the initiative from applicants, but rather to use input from AAS members to identify strategic research needs that are appropriate.

All successful applicants will be required to provide matching funding from other sources and report to the AAS at predetermined stages. The limit for individual grants is currently $50,000 with a pool of funding up to $ 100,000 over 3 years. The approved AAS priority research areas are:

- Windfarm noise assessment
- Underwater noise monitoring & detection
- Environmental noise modelling and assessment
- Sleep disturbance assessment
- Road traffic assessment

Applications can be submitted at any time; however the closing date for review each year is 5pm Australian Eastern Standard Time on October 1, and announcement of each application will be made at the annual conference in November each year.

Details and application from via "Awards and Grants" menu item on the AAS website.

QLD Division

The year to date has been very busy for the Queensland Division. On 25 February members of the Society visited Saint Peters Lutheran College at Indooroopilly to see and hear their new Performing Arts Centre in action. The Society was treated to a special musical performance by the school before James Heddle, the Principal/ Managing Director of James Heddle Pty Ltd Acoustical Consultants, talked about the acoustical design features that were incorporated into the design. The site visits enabled the members the rare opportunity to be shown the principles used to acoustically design a dedicated performance venue.

On Wednesday 9 April the Society was treated to two presentations titled the *Investigation of Digital Audio Manipulation Methods*, presented by Brendan Trevorrow, awardee of the AASQ 2013 Acoustic Bursary and *Detailed Analysis of Acoustical Performance of Smart Phone Devices* presented by Sam Denman, awardee of the AASQ 2013 RJ Hooker Bursary. These presentations were of very high quality, with both presenters being very knowledgeable about their topics.

A site visit was conducted by members of the Society on Tuesday 3 June to the Autex factory at Acacia Ridge to see how just-in-time manufacturing is used to create a range of polyester insulation products. The quality control measures implemented by Autex were of interest to the members, along with the no-wastage manufacturing process. Thanks must be given to Donovan Cochrane, Account Manager (Insulation) and Rob Jones, Market and Development Manager from Autex for facilitating the factory tour.

On Wednesday 16 July the Society hosted a presentation from Frank Henry, Program Delivery Manager Pollution Control and Alex Marchuk, Senior Program Officer Noise Policy of Brisbane City Council on the new *Brisbane City Plan 2014*. This event was opened up to non-members of the Society and as a result there were 55 attendees, consisting of 38 members and 17 non-members. The presentation provided an overview and rationale of the noise content of *Brisbane City Plan 2014*, including noise criteria, the *Noise Impact Assessment Planning Scheme Policy* and relevant codes. Key changes in relation to noise impact assessment were highlighted, which will be useful to our members who practice in acoustic consulting.

NSW Division

In March the NSW Division was delighted to be able to host its 2nd Technical Talk for 2014 at a brand new venue: The Australian Hearing Hub, Macquarie University, North Ryde. In January 2013 NAL and Australian Hearing vacated Greville Street and moved into The Australian Hearing Hub. The Hub brings together some of the country's best researchers, educators and service providers to improve the lives of people who experience hearing and language disorders. Dr Jörg Bucholz, Senior Research Scientist, Macquarie University (currently seconded to the NAL) and several key researchers gave an overview on some of the work currently being undertaken at The Hub. This was followed by a tour of the new facilities, including the new 'quiet', 3-D Sound room.

The 3rd NSW Division Technical Talk for 2014, held on 13 May, again took advantage of the new venue (thanks to Warwick Williams for coordinating). The landmark NSW Court of Appeal decision to uphold the Land and Environment Court decision to halt the expansion of the Mount Thorley Warkworth coal mine was the topic for the evening. The Court found that the expected economic gains from expanding the mine were insufficient to justify the environmental impacts, in particular noise and dust impacts. Paul Mitchell and Najah Ishac, managing directors of EMGA Mitchell McLennan Pty Ltd, gave us an overview of the environmental approvals process and the Judgement for the mine expansion, as well as a more intimate discussion of the noise assessment process and lessons learned. A video of the talk was made available in the members only part of the AAS website.

The NSW Division aims to video all technical talks to host on the AAS website, with the permission of the presenter. The purpose of this is to enable members outside the Sydney Metropolitan Area to have access to these informative talks. Note that the videos are usually only accessible on the website for a limited time, so take advantage of the opportunity whilst you can.

This year NSW Division is pleased to be able to provide funding for 9 recipients of the AAS NSW Division Travel Awards, to assist with their costs attending InterNoise 2014. NSW Division offers 3 categories of Travel Award:

1. MAAS Travel Award of up to $1,500, offered to current AAS NSW Division members who have held membership for 12 months or more;
2. ECR Travel Award of up to $1,500, offered to Early Career Researchers in acoustics who are AAS members of the NSW Division; and
3. RS Travel Award of up to $1,000, offered to Research Students who are AAS student members of the NSW Division as well as research students endorsed by AAS members of the NSW Division.

This year we had a total of 13 applications for the various awards. All applications were of an extremely high standard. The review process was very difficult and much debated. To allow the Awards to assist the most applicants possible, some of the funding was increased and in some cases, where it was too difficult to select an outright winner, the award was split between two recipients. The NSW Division would like thank all those who applied for the 2014 Travel Awards. We congratulate the following award winners and look forward to seeing your paper at InterNoise 2014:

- Jeffrey Peng (MAAS Award - shared)
- Gareth Forbes (MAAS AWARD - shared)
- Herwig Peters (ECR AWARD - shared)
- Sebastian Oberst (ECR AWARD - shared)
- Zhi Zhang (RS Award)
- Manuj Yadav (RS Award)
- Daipei Liu (RS Award)
- Hongjian Wu (RS Award - shared)
- Samaneh Fard (RS Award - shared)

SA Division

Early in 2014 the South Australian division advertised supplementary funding of up to $1000 per project for up to two final-year university acoustics-related projects. We are pleased to announce that two applications from students at the University of Adelaide have been successful. One of the successful applications, submitted by Kieran Doherty, Matthew Tripodi and Francesco Larizza, is the '*Active Exhaust Silencer*' in which the team of students are developing an adaptive quarter wavelength tube tuned by varying air temperature to control exhaust noise from an internal combustion engine. The second successful application, submitted by Anna Koefer and Pablo Blanco, is the '*Investigation of Noise from Flow Control Valves*', in which the students are building a noise test-rig, the data from which they will synthesise with analytical noise models and CFD flow simulations to better understand the flow-induced noise mechanisms of the valves under consideration.

On 9 July, the SA division held a technical meeting. The guest speaker was Peter Exton, a Senior Consultant at Marshall Day Acoustics, who specialises in the room acoustic design of venues for the performing arts. Peter presented a '*Demonstration of the new IRIS room acoustics measurement system in Bonython Hall*' at the University of Adelaide. The IRIS measurement system, developed by Marshall Day Acoustics, enables 3D impulse responses to be captured and analysed through a commercially available tetrahedral microphone array and a four channel USB audio interface.

New Products

CadnaA Online Noise Map

Datakustik using CadnaA was commissioned by the King Fahd University of Petroleum & Minerals to produce a noise map of Khobar City located in Saudi Arabia. The objective of the study was to devise a monitoring and assessment procedure for both noise and air pollution of road traffic in major (arterial) roads in urban areas within cities in the kingdom and to formulate a decision-assisting framework employing state-of-the-art technology of computer-aided modeling, prediction and mapping of noise and air pollution levels as a result of road traffic.

The interactive map can be found at
www.benoiseaware.kfupm.edu.sa/interactivemaps/

This noise and air pollution map can be accessed by most mobile devices. By moving the mouse/finger across the map, the respective noise level (dBA) or pollutant ($\mu g/m^3$) is shown.

Information: Rodney Phillips at Renzo Tonin & Associates (02) 8218 0500 or RPhillips@renzotonin.com.au

Logos Technologies Receives DARPA Award to Develop Silent-Capable Hybrid-Electric Military Motorcycle

Logos Technologies has received a small business innovation research (SBIR) grant from DARPA (the Defense Advanced Research Projects Agency) to develop a military-use hybrid-electric motorcycle with near-silent capability. When fully matured, the technology will allow small, distributed military teams to move long distances quickly and stealthily across harsh enemy terrain.

Developed in partnership with San Francisco-based all-electric motorcycle producer BRD, the platform will combine Logos Technologies' quieted, multifuel hybrid-electric power system with a cutting-edge, off-road electric motorcycle platform developed by BRD. This initiative will be the first time that a two-wheel-drive, multifuel hybrid capability has been integrated into a full-size off-road motorcycle.

Beyond the efficiency and mobility improvements the design aims to bring, the hybrid-electric approach also allows for extended periods of near-silent, electric-only-propulsion as well as the generation of supplemental electric power for use by personnel in the field.

Larson Davis - Outdoor Preamp

The Larson Davis PRM2103 Outdoor Preamplifier is designed for use with the Model 831 Sound Level Meter, the 377B02 microphone, and an environmental shroud for portable or permanent outdoor use in a wide range of weather conditions. The PRM2103 combines the cost savings of a standard preamplifier with features for unattended monitoring and an inexpensive shroud to create a product that is ideal for remote sound level measurement.

The new PRM2103 features an automatic calibration that checks five different frequencies all at the same time which enables better detection of any failure. When used with the Model 831 sound level meter, the acoustic response can be selected as free-field, random or

90 degree using a simple setting on the sound level meter. Due to this flexibility, only one preamplifier is necessary and there is no need to use different microphones for different fields.

Because field visits to a remote monitor can be expensive and time consuming, the PRM2103 has been designed to require no routine maintenance. It includes a built-in humidity and temperature sensor and can automatically turn on an internal heater when there is a risk of condensation. All this has been accomplished while keeping power usage low (< 2 mA with heater off) so the PRM2103 is an excellent solution for battery powered applications.

Audiometric Couplers

Larson Davis introduces two new $2cm^3$ couplers and a new '711' coupler for testing insert earphones and other audio transducers. AEC204 is a very affordable occluded-ear simulator that includes a built in microphone and is fully compliant with IEC 60318-4. The AEC202 and AEC203 are IEC 60318-5 compliant, $2cm^3$ couplers designed to be used with ½ in and 1 in microphones respectively.

The addition of these new couplers enhances the extensive line of Larson Davis products for calibrating audiometers which include a sound level meter with firmware features necessary for calibrating audiometers, artificial ears, couplers, an artificial mastoid, and PC software for managing calibrations. Using the AUDit software, you can automatically correct RETSPL measurements for the effects of a microphone and coupler while tracking calibration history in a database.

Information: suheil.khandwalla@thermofisher.com
www.thermofisher.com.au

Mobile Sound Viewer© - Low Cost Hand Held Acoustic Camera

The Mobile Sound Viewer© (www.mobilesoundviewer.eu) is a new product showing in real time the sound sources ‘hot spots’ responsible for overall noise radiation “in the field”. As a thermal camera would do it, it is now possible to locate, quantify (in terms of frequency content and sound radiation) and visualise noise from a colour palette overlaid on a video. The Mobile Sound Viewer© handles traditional analysis methods like A-weighting, one-third octave band analysis, narrow band analysis.

The Mobile Sound Viewer© is the first device to be totally portable. It is very light, standalone and compact in order to analyse noise sources in very tight places. Locating a noisy area can be done just by orientating the unit which has an operating distance between a few centimetres to several meters.

Information www.acousticsbotte.com

Noiselab – Acoustics Engineering Laboratory Online

With Noiselab you can compute flanking transmissions in buildings and simulate new materials to predict sound insulation with a higher precision than traditional methods. Noiselab follows the Australian acoustic standard AS ISO 140.4- Field measurements of airborne sound insulation between rooms (Annex C Measurement of flanking transmission) and is based on the concept of reciprocity (coined by Lord Raleigh, H. von Helmholtz).

The Solid Noise Probe (one axis accelerometer SVANTEK SV80 cased in a patented anti vibration casing avoiding the need for mechanical mounting- sensitivity 100 mV/g) easily measures vibrations on room finishes. The Probe measures the average surface velocity levels of the walls and the flanking surfaces in the source/receiving room with an electro acoustic source as the excitation signal. Once the measurement data are uploaded to Noiselab, you can simulate the addition of multi layers of plasterboard or different cladding, flooring and suspended ceilings. The on-line software www.noiselab.net displays single-number quantity – Rw(C; Ctr), Dn, w (C; Ctr), DnT, w(C; Ctr) and vibrations levels. These ratings of sound insulation are calculated according to AS/NZS 1276.1 (ISO 717.1).

Information: www.acousticsbotte.com

SADiA-3© - Acoustic Monitoring System for Outdoor Music Festivals

SADiA-3 (www.euphonia.fr) displays the sound pressure levels measured in public area together with the sound levels measured in the neighbourhood of the musical event; the sound engineer can then control the emission levels to comply with the noise limits at the nearest residential properties. Thanks to a dedicated screen positioned at the mixing desk, SADiA enables the sound engineer to watch over the time-averaged Leq plotted on a graph. The time span can be set from 1 to 60 minutes. The frequency weighting can be chosen between A, C or Z (lin). In addition, SADiA displays the Leq1s (short time Leqs), the Peak levels and the spectrum measured in 1/3 octaves. It is possible to link short term/long term monitoring tags in the neighbourhood, on strategic locations. These tags will send level information every second over the web, allowing the sound engineer to adjust the level to prevent any noise nuisances.

Information www.acousticsbotte.com

S-Cube Development Kit! Silence It Yourself

Silentium's S-Cube™ Development Kit (www.silentium.com) allows system designers to come up with a custom noise reduction solution for noisy equipment. The kit is the heart of Silentium's Active Noise Control (ANC) range of products that reduce annoying fan noise in air conditioning systems, medical equipment, and more. It can be used in any application where the constant humming of fans causes irritating noise in the surrounding area.

The S-Cube™ Development Kit reduces broadband noise of machinery fans and blowers up to 10dBA on top of the reduction achieved by acoustic material.

- Easily programmable user interface, 2 button operation and LCD display
- Adaptable for both point to point and point to zone ANC applications
- Ideal for acoustic and vibration applications
- Real time adaptive algorithms
- Broad band noise reduction
- Self-calibrating system
- Computer connection not required

Information www.acousticsbotte.com

AAS Website Update

The new website for the AAS is very close to becoming fully live with all the features activated. As well as the pages open to all, AAS members are advised to check that they can access the members-only area. While it has been trialled there are always problems with links in a new site so please let the web manager know if you experience any problems.

Future Conferences

ICSV22

The 22nd International Congress on Sound and Vibration (ICSV22) will be held from 12 to 16 July 2015 in Florence, Italy. The ICSV22 will include invited and contributed papers and the following distinguished plenary lectures by Otto von Estorff, Germany, Kirill Horoshenkov, UK, Dick Botteldooren, Belgium, Semyung Wang, South Korea, Lily M. Wang, USA, Wim Van Keulen, Netherlands and Roberto Pompoli, Italy.

ICSV22 participants will be able to take part not only in a congress with a first-rate scientific programme but will also be able to experience the vibrant culture of Italy. The Congress venue will be the Firenze Fiera Congress & Exhibition Centre, an exclusive area in the centre of Florence. The venue is only ten minute walk from the old city and 1 km from the famous Cathedral of Santa Maria del Fiore. There will be an extensive exhibition of sound and vibration control technology, measurement instrumentation and equipment, and various social activities will be featured.

The deadline for abstract submission is 1 December 2014.
Details www.icsv22.org

Acoustics Australia

Submissions for the journal are welcomed.

Special issue on Room Acoustics planned for early 2015

Details are available from the web page.
Submit your paper or technical note to:
The Editor at AcousticsAustralia@acoustics.asn.au

Online submission via Springer Publishing soon to be available.

Inter-Noise 2014

MELBOURNE AUSTRALIA 16-19 NOVEMBER 2014

The Australian Acoustical Society will be hosting Inter-Noise 2014 in Melbourne, from 16-19 November 2014. The congress venue is the Melbourne Convention and Exhibition Centre which is superbly located on the banks of the Yarra River, just a short stroll from the central business district. Papers will cover all aspects of noise control, with additional workshops and an extensive equipment exhibition to support the technical program. The congress theme is *Improving the world through noise control*.

Registration Fees

The registration fees have been set as:

Delegate	$840
Student	$320
Accompanying person	$140
Congress Banquet	$130pp

The registration fee will cover entrance to the opening and closing ceremonies, distinguished lectures, all technical sessions and the exhibition, a USB containing the full papers and light lunch plus morning and afternoon teas. The Congress Banquet will have a strong Australian theme and feature the opportunity for delegates to take photographs of themselves with native Australian animals, so should prove to be a major attraction.

The social program commences with the welcome reception on the Sunday evening after the opening and first plenary lecture. On each of the following days, the morning and afternoon refreshments and light lunch (all included in the registration fee) will be provided in the exhibition area. The optional banquet (additional charge applies) will be held at the venue and provide, along with great food and wine, an Australiana theme. After the final sessions the closing reception will bring the congress to an end. Additional features are included in the program for accompanying persons.

An exhibition of the latest developments in equipment and acoustic related materials will take place in the foyer of the Conference centre from Monday morning until Wednesday lunch-time. Over 50 out of 60 booths are already booked by international and Australian companies. More details on booking space in the exposition available from www.internoise2014.org.

Technical Program

The opening plenary lecture will be: “Sound Sketch: its Theory and Application using Loudspeaker Arrays” by Prof. Jung-Woo Choi of South Korea.

The closing plenary lecture will be: “Soundscape Planning as a Complement to Environmental Noise Management” by Prof. Lex Brown of Australia.

The four keynote topics, by world authorities on their subject, will complement major areas within the Congress. They cover Aircraft Noise, Active Noise Control, Wind Turbine and LFN as well as the Impact of Building Acoustics on Speech Comprehension and Student Achievement.

Over 900 abstracts have been submitted covering all areas of noise control engineering and within the broad theme of the Congress – “Improving the World through Noise Control”. There will be 10 to 12 parallel sessions plus an area for poster sessions.

In addition to the congress program there will be technical study group meetings and a number of short courses are offered on the Sunday before the opening session of the Congress.

More details on all aspects of the conference at: www.internoise2014.org


Specifier: Bates Smart

SUSTAINING MEMBERS

The following are Sustaining Members of the Australian Acoustical Society.
Full contact details are available from http://www.acoustics.asn.au/sql/sustaining.php

3M AUSTRALIA
www.3m.com

ACOUSTIC RESEARCH LABORATORIES
www.acousticresearch.com.au

ACOUSTICA
www.acoustica.com.au

ACRAN
www.acran.com.au

ACU-VIB ELECTRONICS
www.acu-vib.com.au

ADAMSSON ENGINEERING
www.adamsson.com.au

AERISON PTY LTD
www.aerison.com

ASSOCIATION OF AUSTRALIAN ACOUSTICAL CONSULTANTS
www.aaac.org.au

BARRISOL AUSTRALIA
www.barrisol.com.au

BORAL PLASTERBOARD
www.boral.com.au/plasterboard

BRUEL & KJAER AUSTRALIA
www.bksv.com.au

CSR BRADFORD INSULATION
www.bradfordinsulation.com.au

EMBELTON
www.vibrationisolation.com.au

HOWDEN AUSTRALIA
www.howden.com

IAC ACOUSTICS (AUSTRALIA)
www.iac-acoustics.com/au/

INSTYLE CONTRACT TEXTILES
www.instyle.com.au

JACOBS
www.jacobs.com

KNAUF AMF AUSTRALIA PTY. LTD.
www.amfceilings.com.au

MASON MERCER AUSTRALIA
www.masonmercer.com.au

OFFICE OF ENVIRONMENT AND HERITAGE NSW
www.environment.nsw.gov.au

PEACE ENGINEERING
www.peaceengineering.com

PYROTEK NOISE CONTROL
www.pyroteknc.com

REGUPOL (AUSTRALIA) PTY LTD
www.regupol.com.au

SOUND CONTROL
www.soundcontrol.com.au

SOUNDSCIENCE
www.soundscience.com.au

VIPAC ENGINEERS AND SCIENTISTS
www.vipac.com.au


RISE ABOVE THE NOISE
SoundPLAN® just elevated noise modeling software to new heights!
SoundPLAN 7.3 – now in 64 and 32 bit versions.
The most powerful indoor/outdoor noise simulation tool available. Build models from Google Earth, aerial surveys & other sources.
SoundPLAN Essential 3.0 – now with source library.
Fast simulations for small to medium projects. Features our full calculation core with an easy-to-use interface.
Call or download a FREE demo:
02 8211 1142 or soundplan@marshallday.com
http://soundplan.com.au
Available in multiple languages with expert local support in nearly 50 countries.
Sound PLAN

Diary

2014

7 - 12 September, Krakow, Poland
Forum Acusticum 2014
http://www.fa2014.pl/

8 - 10 September, Fort Lauderdale, Florida
Noise-Con 2014
http://www.inceusa.org/nc14

16 – 19 November, Melbourne, Australia
Inter-Noise 2014
http://www.internoise2014.org/

2015

20 - 23 April, Glasgow, Scotland
Wind Turbine Noise 2015
http://www.windturbinenoise.eu/

11 – 15 May, Metz, France
International Congress on Ultrasonics (ICU 2015)
http://2015-icu-metz.gatech.edu/

31 May - 3 June, Maastricht, Netherlands
Euronoise 2015
https://www.euracoustics.org/events/events-2015/euronoise-2015

12 - 16 July, Florence, Italy
22nd International Congress on Sound and Vibration (ICSV22)
www.icsv22.org

9-12 August, San Francisco, USA
Inter-Noise 2015
http://internoise2015.com

6-10 December Singapore,
Wespac 2015
wsgan@acousticaltechnologies.com

2016

10-14 July, Athens, Greece
23rd International Congress on Sound and Vibration (ICSV23)
http://iiav.org/index.php?va=congresses

21 - 24 August, Hamburg, Germany
INTER-NOISE 2016

5-9 September, Buenos Aires, Argentina
22nd International Congress on Acoustics (ICA 2016)
http://www.ica2016.org.ar/

12-16 September, Terrigal, NSW, Australia
International Workshop on Rail Noise (IWRN)
http://www.acoustics.asn.au/IWRN12

Meeting dates can change so please ensure you check the conference website: http://www.icacommission.org/calendar.html

Inter-Noise 2014

MELBOURNE AUSTRALIA 16-19 NOVEMBER 2014

The Australian Acoustical Society will be hosting Inter-Noise 2014 in Melbourne, from 16-19 November 2014. The congress venue is the Melbourne Convention and Exhibition Centre which is superbly located on the banks of the Yarra River, just a short stroll from the central business district.

The congress theme is *Improving the world through noise control*. Major topics will include community and environmental noise, building acoustics, transport noise and vibration, human response to noise, effects of low frequencies and underwater noise.

Further details are available on the congress website **www.internoise2014.org**

ACOUSTICS AUSTRALIA ADVERTISER INDEX - VOL. 42, No. 2

Pyrotek .... Inside front cover
CSR 74
SAVTek 76
Cadna 137
Thermo Fisher Scientific 138
Regupol 140
ARL 142
Bruel & Kjaer 143
Cliff Lewis Printing. 144
Matrix 144
Sound Level Meters 145
NATacoustic 145
Screenwood 147
Soundplan 149
ACU-VIB .... inside back cover
Bruel & Kjaer .... back cover

SVAN 979 ANALYSER

Sound & Vibration Analyser

Product description

The Most Advanced & Powerful Single Channel Instrument Ever Made!

Features

- Two DSP for superior performance
- 3 parallel independent profiles
- 1/1 & 1/3 octave RT & FFT anallysis
- 1/6 or 1/12 octave real-time analysis
- Advanced logger with audio events
- Reverberation time measurements
- Time-domain signal recording
- Wireless data transfer
- Signal generator
- Micro SD card
- User programmable Band-Pass Filters

TOPSELLER Best Choice

More info: www.acu-vib.com.au

NATA Calibrations

Acoustic and vibration instrument calibration laboratory servicing the whole of Australia and Asia Pacific region. FAST turnaround. NATA registered. All brands

Sales & Service

We supply instruments from several manufacturers, giving you a wider selection to choose from. We also service ALL brands providing you a one-stop solution.

Instrument Hire

We have a wide range of acoustic and vibration instruments for hire. Daily, weekly or monthly rental periods available. Call for advice on the best instrument to suit you

ACU-VIB ELECTRONICS- 14/22 Hudson Ave - Castle Hill 2154 - Phone 02 9680-8133 - eMail info@acu-vib.com.au

REAL-TIME NVH VALIDATION GOES MOBILE

PUT SMART RECORDING IN THE DRIVER´S SEAT

View test information in real-time

Touch NVH data to explore and analyse

Record and play sound with the binaural headset

SONOSCOUT™

- Confidently compare and analyse in the vehicle
- Simple one-person operation
- Grab and go when time is short
- Finish testing as soon as you are certain
- Send highly relevant data onwards
- Mark events and view track later in Google Earth™
- Free demonstration mode

Connect transducers to the wireless LAN-XI battery module

Brüel & Kjær

BEYOND MEASURE

Download on the App Store

Bruel & Kjaer Australia
Sydney: Tel: +61 2 9889 8888 · Fax: +61 2 9889 8866
Melbourne: Tel: +61 3 9508 4907 · Fax: +61 3 9687 5969

auinfo@bksv.com.au

www.bksv.com/sonoscout

BN 1287 – 12