SOUND CHECK

TEXT: TOBIAS HERRMANN

Operatic singing. Birdsong. Loud shouting. An off-pitch violin. We instinctively find some sounds pleasant, others unpleasant. But how do we decide whether something sounds good or bad? And how is sound actually processed within the brain? In an attempt to answer these questions, a team led by David Poeppel at the Max Planck Institute for Empirical Aesthetics in Frankfurt is trying to break down speech and music into their most elementary components. And at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, researchers are investigating the secret of super-hits.

Alarm! Screams, including babies' cries, feature an acoustic peculiarity that we experience as particularly unpleasant. It's what guarantees their social impact.

PHOTO: PLAINPICTURE/BLEND IMAGES/ROBERTO WESTBROOK

PHOTO: JÖRG BAUMANN/MPI FOR EMPIRICAL AETHETICS

Unique testing stage: the Max Planck Institute's ArtLab in Frankfurt is both a concert hall and a laboratory. With it, scientists can record the various physiological data of both the performers – like the vocal ensemble Cut Circle pictured here – and the listeners during performances.

"What is the role of neuroscience?" David Poeppel's response to this question, posed in an interview, was as follows: "Breaking something up into its constituent parts." This observation reflects both his personal approach as a researcher and that of the Max Planck Institute for Empirical Aesthetics in Frankfurt, where Poeppel has been Director since 2014. However, that's not where you would have reached him on the day of this interview in April 2020. Instead, you would have had to dial a telephone number starting with +1 – the country code for the U.S. Since 2009, he has held a part-time professorship in psychology and neuroscience at New York University. At the start of the COVID-19 pandemic, Poeppel and his family left the hotspot New York City for Connecticut, where he continues to work from home. In his words, it has been a "blessing in disguise"; he has had the time to pursue ideas that he had previously put on the back burner in his daily work.

Time is also relevant to his research; one of his research interests is how speech and music are processed in time. Poeppel gives an explanation for the layperson: "A sound wave reaches your ear, is converted into an electrical signal and is then split apart at switching points in your brain. The final result is tiny elementary constituents, which – if processed correctly – carry the appropriate information." His interest, then, is how acoustic signals are processed in the human brain. If he can answer these questions, Poeppel hopes to make advances in linguistic theories and in the aesthetics of speech and music.

Many languages, one tempo

There's no denying that a conversation with David Poeppel is inspiring. Many of the questions he poses are ones you're likely to have already asked yourself, while others are very unexpected. Some sound very complex, and others are almost astonishingly simple. For instance, you don't have to be a linguist to know that words are made up of one or more syllables; surely, the question "what is a syllable?" is, at first glance, banal. But from a scientific and technical standpoint, nothing could be further from the truth. As Poeppel explains, linguists have been discussing for some 70 years whether syllables should be regarded as elementary constituents of speech or whether they are just a type of by-product of smaller acoustic elements, such as phonemes, the individual units of speech sounds.

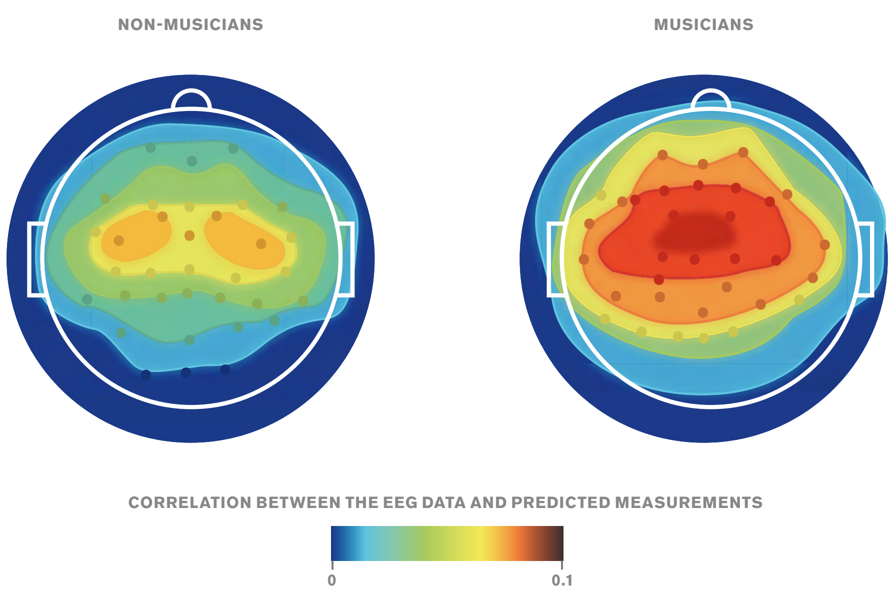
What is beyond dispute is that syllables play a fundamental role in speech perception and speech production. In one long-term project, Poeppel and his colleagues compared the speed of various languages with the number of syllables uttered. They discovered that the average speed of speech corresponds to the rate of successive syllables. The astonishing thing was that this tempo was almost identical for countless languages. “Our sense that some languages are spoken much faster than others is therefore mistaken,” says Poeppel.

A person can easily speak four to five multisyllabic words in the space of one second. To clearly understand each word, the listener needs to perceive every single sound. Moreover, the sequence of the sounds is crucial. The slightest errors can result in chaos, as any child who’s ever played the popular game of “Telephone” will know. “Wall” quickly becomes “ball”, “shoes” becomes “choose”, while “smell”, “sell” and “sail” can hardly be distinguished when whispered.

To perceive such fine differences in ordinary discussions, the brain has to achieve a temporal resolution of between 20 and 80 milliseconds. “The brain needs to be structured in such a way that it can construct very short intervals of time: ‘samples’, as we refer to them,” says Poeppel. But that would only result in a staccato-like sequence of sounds. Speech is also highly dependent on precise stresses, pauses and intonation. Take the example of the German phrase “München wird modern”. Its meaning depends on whether the last syllable of the word “modern” is stressed (“moDERN”, meaning “Munich is modern”) or the first syllable is emphasized, “MOdern”, which changes the meaning to “Munich is rotting.” Speech only becomes interesting and vibrant when it includes such ‘prosodic contours’. The importance of such factors is particularly evident in sarcasm and irony. Whether you are praising someone when you say “well done”, or whether you’re being sarcastic depends completely on your chosen emphasis.

To perceive such nuances, longer time intervals are needed; the brain needs to generate both a temporal and a spectral analysis, and this occurs at frequencies of a few Hertz, corresponding to time intervals of between 200 and 300 milliseconds. “Ultimately, two parallel processes need to take place in the brain,” explains Poeppel. “I can work out the correct order of sounds based on the short time intervals, and the long time intervals indicate the intonation and the speech melody.” If you want to discover how these differing lengths are then precisely analyzed and converted into concrete information, you have to delve deep into neurobiology. That’s where neuronal oscillations play an essential role.

By this, neuroscientists mean the synchronized activity of particular groups of cells. Once a sentence, a melody or a sound reaches the ear as a sound wave and is converted into electrical signals, certain nerve cells in the brain become synchronized, switching on and off in defined cycles. In order to process the aforementioned short time intervals of under



Musical heads: researchers at the Max Planck-NYU Center for Language, Music and Emotion used electroencephalography (EEG) to study how the brain follows a melody. They utilized their findings to make predictions about how the subjects would measurably react to melodies. These predictions were more accurate for musicians’ reactions, and the correlation with the melodies was also better than with the reactions of non-musicians.

GRAPHIC: GCO BASED ON ACOUSTICAL SCIENCE AND TECHNOLOGY, 41, 361-364

100 milliseconds, the relevant cells oscillate at a frequency between 25 and 35 Hertz, known as 'gamma waves'. For the longer intervals, a different type of cell becomes active at a frequency between three and eight Hertz, and these oscillations are known as 'theta waves'.

Neural oscillations don't just play a role in perceiving speech. They also underlie the brain's ability to process music, as David Poeppel discovered with his colleague Keith B. Doelling from New York University. In their study, the two compared active musicians with at least six years of musical training with non-musicians. The test subjects listened several times in succession to 13-second excerpts from various classical pieces by Johann Sebastian Bach, Ludwig van Beethoven and Johannes Brahms. The pieces were played on the piano in varying tempos – from one note every two seconds to eight notes per second.

We process speech and music in similar ways

For the pieces of music with a faster rhythm than one note per second, the researchers were able to record cortical oscillations in musicians and non-musicians, and these oscillations were synchronized with the speed of the notes in the piece being heard. "The findings show that the presence of these oscillations improves our perception of music and pitch changes," explains Keith Doelling.

At the same time, they also observed that the brains of the musicians synchronized more clearly with the rhythm of the music than those of subjects with no musical training. In addition, it was only in the musicians that oscillations were recorded that synchronized with unusually slow pieces. This difference indicates that people without musical training may have difficulty recognizing continuous melodies, instead perceiving music as just a series of tones. In the larger context of their research, the findings also demonstrate that low-frequency oscillations enable the brain to decipher speech or music.

Pauline Larrouy-Maestri, a senior researcher in Poeppel's group, also investigates parallels between speech and music. Given her broad background, she appears predestined to conduct this kind of research. She studied psychology and music, plays the piano and used to work as a clinical speech therapist. In a typical experiment, Larrouy-Maestri asks subjects to listen to music and then rate the performance. She plays them either synthesized or acoustic pieces, both of which have their advantages and disadvantages. "Synthesized music is very easy to manipulate and control, but it's not as natural, so it's not as easy to define how people actually perceive the music," says Larrouy-Maestri.

In one of her experiments, she asked volunteers to listen to famous chorales by Bach that were altered at certain points in the music. She then analyzed how the subjects' brains reacted to the altered passages or notes. Listeners, she discovered, were able to recognize harmonic structures and, therefore, precisely identify the places where the music had been altered. Pauline Larrouy-Maestri und Xiangbin Teng's experiment showed that we analyze music and speech in similar ways. While continuous speech is parsed into linguistic units – sentences, words and syllables – the continuous musical phrases in pieces of music are parsed into musical units – melodies, chords and notes. The more musically trained the subjects were, the better their brains could distinguish the musical units from each other.

Another focus of Larrouy-Maestri's work is the question of whether musically untrained listeners can recognize wrong notes in songs and which cognitive processes are responsible for this. She found that you don't have to be a professional musician to detect wrong notes in a piece of music. Nor is it necessary to have an expert ear to know whether a singer is singing off pitch. Almost anyone can hear what's right or wrong – regardless of the music being played.

Like David Poeppel, the impetus for Larrouy-Maestri's research often comes from observations of everyday life. People turn on the radio and probably switch sta-

"You don't have to be a professional musician to hear wrong notes in a piece of music."

PAULINE LARROUY-MAESTRI

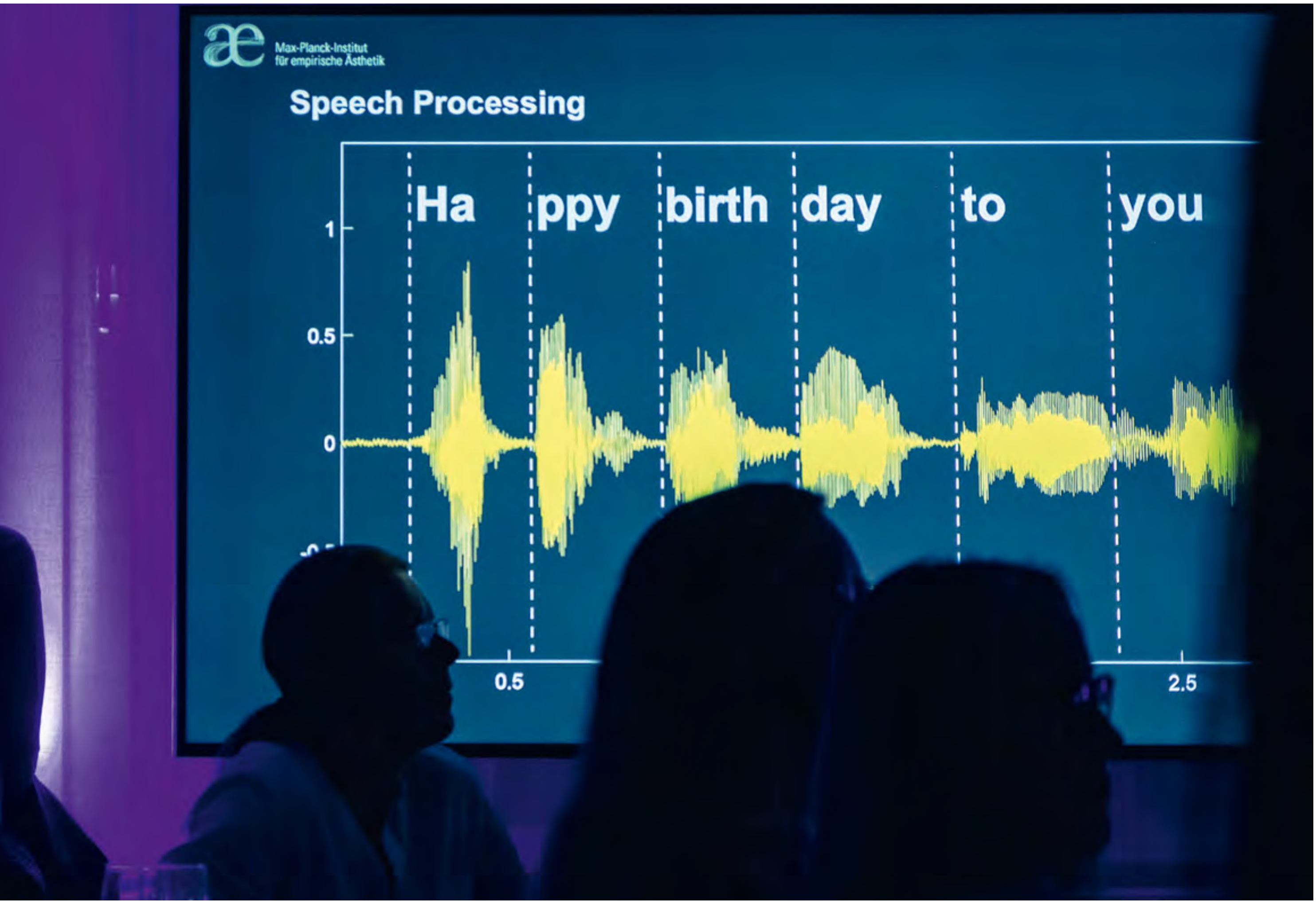


PHOTO: FELIX BERNOULLY/MPI FOR EMPIRICAL AETHETICS

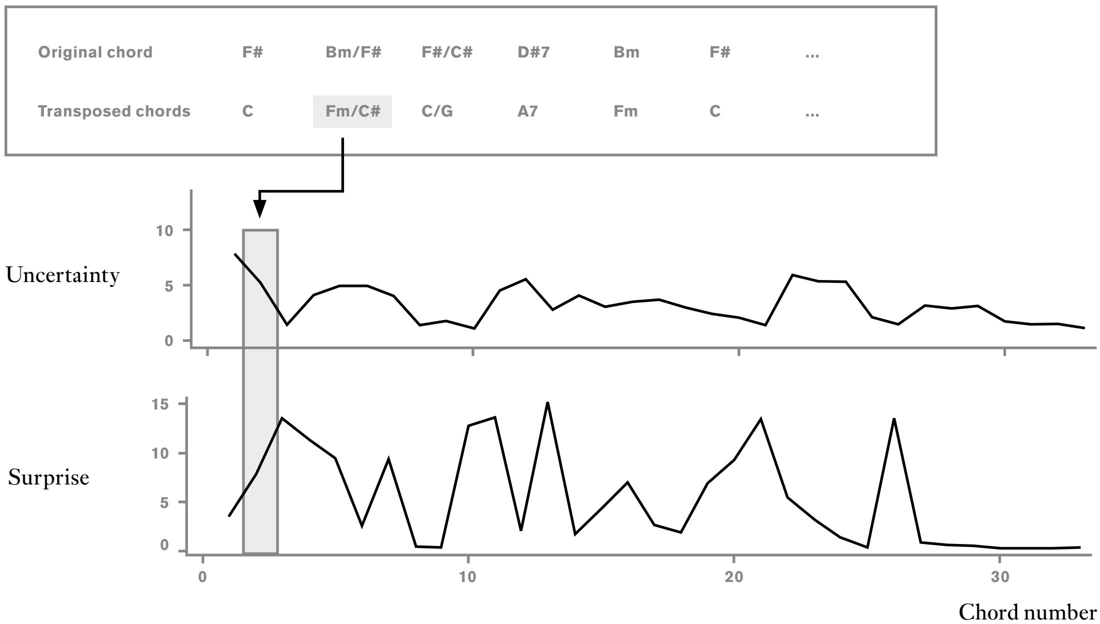
Small-scale processing: in order to understand what someone is saying, our brains must perceive every detail. What we hear is analyzed within time intervals of only 20 to 80 milliseconds. That roughly corresponds to the duration of individual syllables.

tions a couple of times until they find a song they want to listen to. "Irrespective of what kind of sound we perceive, we can immediately say whether we like it or not. Astonishingly, it's something we're all able to do. And so I asked myself, how is that possible?" To answer this question, Larrouy-Maestri has relied on natural acoustic music in her experiments. "We invite volunteers – both trained and untrained musicians – to come into the lab and sing us a song." Then, we ask other volunteers to judge how well they sang.

In another study, Larrouy-Maestri tried to discover which factors influence whether listeners experience a melody as pleasant or less pleasant. In vocal music, these include how accurately the melody is sung, for instance, how on pitch people sing. However, accuracy is not the only criterion. Interestingly, the speed of the music also seems to be a factor. The majority of people experience neither very fast nor very slow music as pleasant. This might have something to do with how the human brain processes music in general. It is these general mechanisms that Larrouy-Maestri is trying to discover, in order to find out what makes people decide whether or not a piece of music is good or bad. "I'm not trying to crack the code for writing the perfect musical hit that everybody loves," she says. "I'm more interested in how listeners reach their decisions regarding whether they like a particular piece or not," says Larrouy-Maestri.

What does a "hit" trigger in the human brain? That's another mystery that a research team at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig is looking into. Vincent Cheung, a doctoral researcher at the Institute, doesn't just have a passion for listening to music; he is also a violinist. He asked himself why certain pieces touched both him and other music lovers so deeply. Together with Stefan Koelsch, he set out to discover the recipe for the success of songs like Yesterday by the Beatles,

⟶



Scientific sound check: using a learning algorithm, researchers from Leipzig analyzed the chord progressions of 745 U.S. super-hits. To make them more comparable, the original chords were transposed into the same key. 'Uncertainty' means that listeners were unable to easily predict the next chord. 'Surprise' denotes how significantly the sound the listeners just heard differed from what they expected to hear.

GRAPHIC: GCO BASED ON VINCENT CHEUNG

Nutbush City Limits by Tina Turner or The Look by Roxette – and he found it. They used machine learning to analyze the 745 super-hits of the U.S. billboard charts from 1958 to 1991. To do this, they removed elements such as text and melody from the pieces, leaving only the chord progressions. Composed of triads and more complicated harmonies, such progressions are familiar to anyone who plays the guitar to accompany songs. Most people in the western world are familiar with their sound; particular sequences of chords have long been the standard building blocks of western music – from simple folk songs to modern pop music.

The scientists then calculated how predictable or surprising the chord progressions in each of the hits were and analyzed the reactions of test subjects to the sound sequences. They found that listening pleasure was greatest when listeners were occasionally surprised, while too much uncertainty was a bad thing. If the listeners were relatively sure which chords would come next, they enjoyed the sensation of being surprised – in other words, if their expectations were not met. If, on the other hand, they were unsure of what was going to come next, they preferred not to be surprised by subsequent chords. These findings were backed up by magnetic resonance imaging studies of the test persons. The nucleus accumbens, the brain structure responsible for anticipating feelings of happiness, only reacted in the test subjects when they were particularly interested in finding out how the music would continue.

But, of course, not everything we hear is pleasant – our auditory system, for example, also exists to warn us of danger. It's a subject the researchers at the Max Planck Institute for Empirical Aesthetics are also examining. David Poeppel caused a stir with a study in which he explored why screams shake us to the core. It's another question that sounds simple until you start looking for the answer. "Everyone can recognize a scream, and everyone has a rough idea of what constitutes a scream – they're loud, high and shrill," says Poeppel, describing the starting point of his analysis. In several studies conducted with his New York colleague Adeen Flinker together with Luc Arnal, Andreas Kleinschmidt and Anne-Lise Giraud from the University of Geneva, he identified an acoustic peculiarity that is unique to screaming.

SUMMARY

The brain analyzes speech at different temporal resolutions to interpret both individual sounds and more complex patterns.

The brain processes music in a similar way. It breaks pieces of music down into individual components, such as melodies, chords and notes.

Professional musicians can often recognize musical structures better than people without musical training. But even lay people can recognize wrong notes or manipulated harmonies.

Popular songs are characterized by a mixture of predictable and surprising chord sequences.

"Screams exhibit a characteristic termed 'roughness'," explains Poeppel. "Roughness occurs when sounds acquire a particular temporal structure due to changing amplitude. If such changes occur extremely quickly, the auditory system can no longer resolve them – they are instead experienced as rough and therefore unpleasant." Normal speech has a modulation frequency of about four to five Hertz, but for roughness that frequency is between 30 and 150 Hertz – the changes are much faster.

In one study, the research team generated a sound database containing a wide variety of human sounds, from screams and sentences to artificial sounds, such as an alarm clock going off. They discovered that both screams and artificial sounds, such as an alarm clock, and dissonant intervals, such as an off-pitch fifth, fall within the frequency range of roughness – a finding that shows that the manufacturers of alarm clocks have done a great job in imitating the modulation of a human scream.

Thus, the sounds that we perceive can be meaningful to us in many different ways. The brain performs an enormous feat in not only distinguishing between different sound sources, but also simultaneously filtering out what is important for us and correctly decoding what we hear. And yet, as David Poeppel points out, even though scientists have made numerous discoveries in recent years, there are still fundamental puzzles that are still unresolved – for instance, the interaction between sound and memory. Finding the answers will require the contribution of many bright minds and visionaries who pose the right questions.

www.mpg.de/podcasts/schall (in German)