Neuromagnetische Repräsentation von Timbre im auditorischen Cortex

Melanie Günther[1], Martin Andermann[1], Christoph Reuter[2], André Rupp[1]

[1] *Sektion Biomagnetismus, Neurologische Klinik, Universität Heidelberg, D-69120 Heidelberg*
[2] *Musikwissenschaftliches Institut, Universität Wien, A-1090 Wien*

Einleitung

Nachdem die Verarbeitung der Tonhöhe in früheren Studien (z.B. [1][2]) bereits intensiv neurophysiologisch untersucht wurde, soll im vorliegendenden Experiment der Fokus besonders auf Aspekte der Klangfarbe gerichtet werden, deren neurophysiologische Verarbeitung bisher kaum untersucht wurde. Dies bezieht sich vor allem auf die Formantstruktur und den Einschwingvorgang: Bzgl. der Formantstruktur stehen hierbei die Veränderungen der Formantposition (Registerwechsel) und Veränderungen der Formantrelation (Änderungen der Instrumentenfamilie) im Vordergrund. In einer früheren Studie unserer Arbeitsgruppe wurde bereits die neuromagnetische Repräsentation von Sprachlauten, bei denen Periodizität, Formantposition und Formantrelation verändert wurden, untersucht [3].

Methoden

An der Studie nahmen 25 erwachsene, normalhörende Probanden teil (13 weiblich; 1 Linkshänder; mittleres Alter: 26.8 ± 7.5 Jahre). Die Vokale /a/ und /o/ wurden von einer deutschen Muttersprachlerin eingesprochen und mithilfe eines Brüel & Kjær Kondensatormikrofons 4193 aufgenommen. Die Aufnahme erfolgte in einem reflexionsarmen Raum, mit einer 24-Bit-AD-Wandlung (RME Hammerfall DSP) und einer Abtastrate von 48.000 Hz. Anschließend wurden die Vokalaufnahmen mittels STRAIGHT [4][5] manipuliert, sodass sich die Position der Obertöne änderte, aber nicht die Form und Position der spektralen Einhüllenden. Die Vokalaufnahmen wurden schließlich auf eine einheitliche Dauer von 666 ms geschnitten und an das Ende der Aufnahme wurde ein 20-ms Hanning window appliziert.

Um Ähnlichkeit zu natürlich gesprochener Sprache herzustellen, wurden die Vokale in zwei aufeinanderfolgenden Triplets angeordnet; die Tonhöhe des jeweils mittleren Vokals jedes Triplets wurde um eine kleine Terz erhöht. Die Position des Spektrums entlang der Frequenzachse wurde mithilfe von STRAIGHT manipuliert. Auf diese Weise ergaben sich vier verschiedene experimentelle Bedingungen (siehe Abbildung 1):

1) Veränderung der Formantrelation 1: Beim Übergang vom ersten zum zweiten Triplet ändert sich der Vokaltyp von /a/ zu /o/. Die Sprechergröße entspricht der Original-Sprechergröße und bleibt über beide Triplets konstant.

2) Veränderung der Formantrelation 2: Beim Übergang vom ersten zum zweiten Triplet ändert sich der Vokaltyp von /a/ zu /o/. Die Sprechergröße wurde mittels STRAIGHT verkleinert und auf beide Triplets appliziert.

3) Veränderung der Formantposition 1: Beim Übergang vom ersten zum zweiten Triplet ändert sich die Sprechergröße vom Original zu einer kleineren Sprechergröße. Der Vokaltyp /a/ bleibt über beide Triplets konstant.

4) Veränderung der Formantposition 2: Beim Übergang vom ersten zum zweiten Triplet ändert sich die Sprechergröße vom Original zu einer kleineren Sprechergröße. Der Vokaltyp /o/ bleibt über beide Triplets konstant.

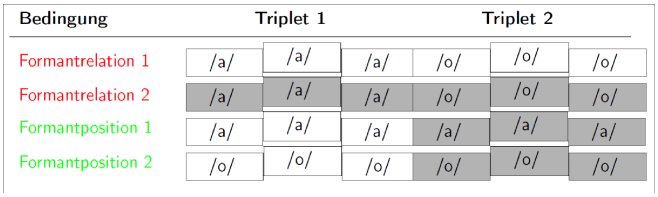

Abbildung 1: Die vier experimentellen Bedingungen der MEG-Studie. Bei den Bedingungen „Formantrelation 1“ und „Formantrelation 2“ ändert sich jeweils der Vokaltyp beim Übergang vom ersten zum zweiten Triplet. f_0 des mittleren Tons jedes Triplets ist um eine kleine Terz erhöht. Bei den Bedingungen „Formantposition 1“ und „Formantposition 2“ ändert sich jeweils die Sprechergröße beim Übergang vom ersten zum zweiten Triplet. Weißer Hintergrund: originale Sprechergröße; grauer Hintergrund: mittels STRAIGHT veränderte Sprechergröße.

Während der MEG-Messung wurde jede experimentelle Bedingung 200 mal in einer pseudorandomisierten Reihenfolge abgespielt. Die Stimuli wurden über Etymotic Research Treiber (ER3) mit einem Pegel von 70 dB SPL präsentiert. Zur Aufrechterhaltung der Vigilanz schauten die Probanden während der Messung einen Stummfilm (mit Untertitel) ihrer Wahl. Nach der Messung wurden die MEG-Daten mithilfe der Software BESA 5.2. [6] analysiert. Die Analyse lieferte Daten zur Lokalisation der neuromagnetischen Quellen (Dipole) im Gehirn sowie zum zeitlichen Verlauf des Dipolmoments dieser Generatoren. Mit jeweils einem Dipolpaar (jeweils ein Dipol im linken und rechten Hörkortex) wurden die Lokalisationen der Dipole ermittelt, die die Änderungen des Magnetfelds durch die Manipulation der Periodizität (P-Dipole), Änderungen der Formantposition (Fp-Dipole) oder Änderungen der Formantrelation (Fr-Dipole) am besten repräsentierten.

Ergebnisse & Diskussion

Die Analyse zeigte, dass sich die P-Dipole im posterioren Heschl’schen Gyrus (HG), in der Nähe zum Planum temporale (PT) befanden, während die Fp- und Fr-Dipole nah beieinander im anterioren PT lokalisiert waren. Die N1-Amplitude war auf den verschiedenen Dipolen dann am größten, wenn eine Änderung in der entsprechenden Vokaleigenschaft vorkam (siehe Abbildung 2): Änderungen der Periodizität führten zu einer großen N1-Amplitude auf den P-Dipolen; Änderungen der Formantposition führten zu einer großen N1-Amplitude auf den Fp-Dipolen, und

Änderungen der Formantrelation führten zu einer großen N1-Amplitude auf den Fr-Dipolen.

Die Ergebnisse deuten darauf hin, dass die anterioren Quellen im HG an der zeitlichen Verarbeitung zur Extraktion der Periodizität beteiligt sind. Die posterioren Quellen im PT sind hingehen an der spektralen Verarbeitung zur Extraktion von Formantposition und -relation beteiligt; diese Quellen liegen räumlich nah beieinander, dennoch führt eine Veränderung von Formantposition und Formantrelation zu einer klar separierbaren, funktionellen Trennung der Quellen (siehe Abbildung 3). Die Lage der verschiedenen Dipole und die Tatsache, dass die neuromagnetische Antwort auf Änderungen von Periodizität, Formantposition und Formantrelation Teil der frühen N1-Komponente ist, stützen die These, dass diese Vokaleigenschaften früh verarbeitet werden [7][8] und dass es im auditorischen Cortex zu einer frühen Trennung von inhaltlichen Informationen („Was?“) und Sprecher-Informationen („Wer?“) kommt [9][10].

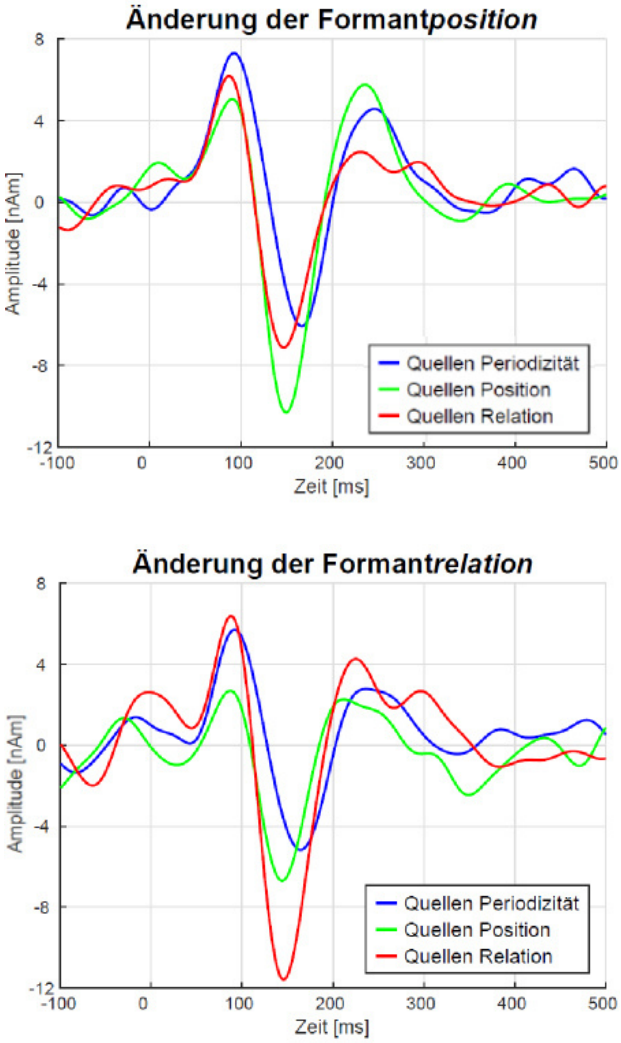

Abbildung 2: Die N1-Amplituden der neuromagnetischen Antworten auf Änderungen der Periodizität, der Formantposition und der Formantrelation sind jeweils auf den entsprechenden Dipolen am größten.

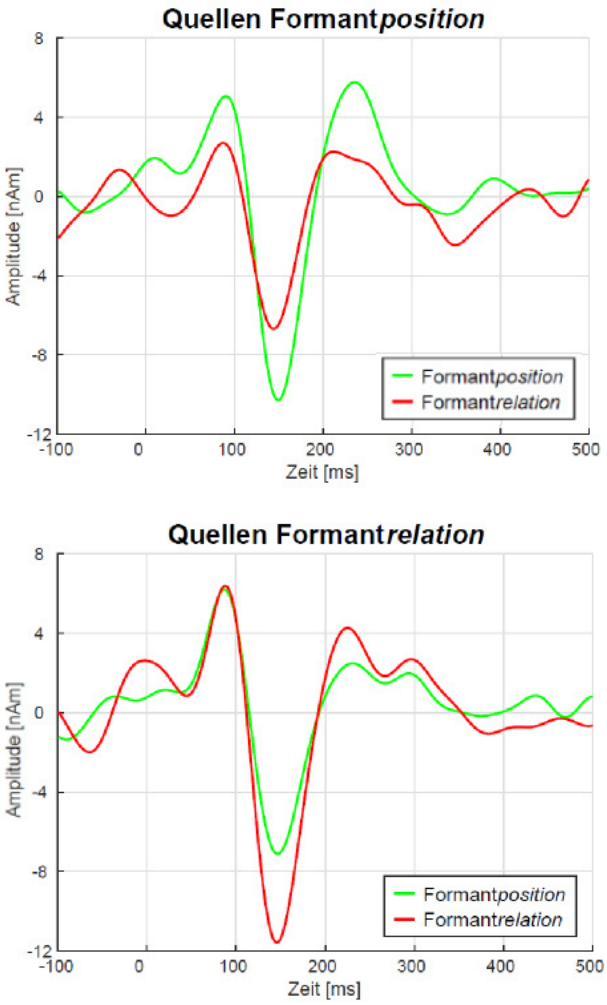

Abbildung 3: Funktionale Trennung der N1-Komponenten in Reaktion auf Veränderungen der Formantposition bzw. Formantrelation.

Ausblick

Die hier präsentierte Studie diente als thematische Grundlage für die Entwicklung einer neuen MEG-Studie zur Klangfarbe von Musikinstrumenten. Ähnlich wie bei Vokalen lassen sich auch die Klänge von Musikinstrumenten anhand ihrer spektralen Struktur, ihrer Formantbereiche, ihrer Einschwingvorgänge und anderer Features beschreiben [11][12][13]. Das Paradigma der im folgenden skizzierten Studie stellt eine Weiterentwicklung und Erweiterung der vorangegangenen Vokalstudie dar, denn es werden nicht nur Änderungen von Periodizität, Formantposition und -relation und ihre Repräsentation im auditorischen Cortex betrachtet, sondern auch Änderungen von Einschwingzeit und der Lokalisation im Raum (entlang der horizontalen Achse).

Ähnlich wie in der vorangegangenen Vokalstudie werden kurze Melodiekonturen präsentiert, die ebenfalls aus zwei aufeinanderfolgenden Triplets aufgebaut sind. Alle Triplets bestehen aus den Tönen d‘, fis‘, a‘ (bzw. zwei Oktaven tiefer), die Reihenfolge der Töne variiert jedoch. Es gibt vier unterschiedliche Tripletkombinationen (siehe Abbildung 4):

1) Änderung der Melodiekontur: Die Melodiekontur besteht aus zwei unterschiedlichen Triplets (Triplet A, Triplet B). Die gesamte Melodiekontur wird jeweils von einer Oboe,

einem Fagott, einer B-Klarinette und einer Bassklarinette gespielt.

2) Änderung des Registers (hoch <-> tief): Die Melodiekontur besteht aus zwei baugleichen Triplets (Triplet C). Beim Übergang vom ersten zum zweiten Triplet ändern sich Instrument und Register (Änderung der Formantposition).

3) Änderung der Instrumentenfamilie (Klarinetten <-> Doppelrohrblattinstrumente): Die Melodiekontur besteht aus zwei baugleichen Triplets (Triplet D). Beim Übergang vom ersten zum zweiten Triplet ändern sich Instrument und Instrumentenfamilie (Änderung der Formantrelation).

4) Änderung der Lokalisation (rechtes Ohr <-> linkes Ohr): Die Melodiekontur besteht aus zwei baugleichen Triplets (Triplet E). Beim Übergang vom ersten zum zweiten Triplet ändert sich die wahrgenommene Lokalisation des Klangs; dies geschieht mittels einer Reduktion des Schallpegels um -6 dB auf nur einer Seite. Die Reduktion des Schallpegels auf der linken Seite führt zu einem stärkeren Höreindruck auf der rechten Seite, und umgekehrt.

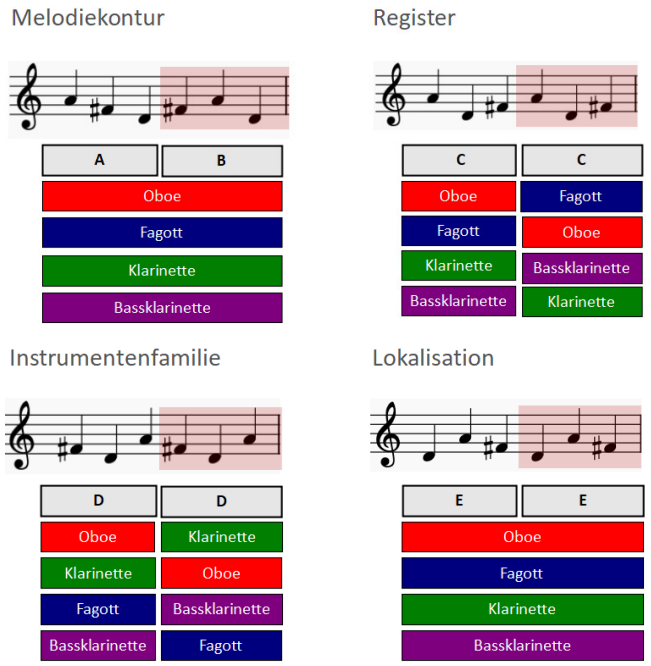

Abbildung 4: Die vier experimentellen Bedingungen der geplanten Studie zum Timbre von Musikinstrumenten. Bei einer Änderung der Melodiekontur zwischen Triplet A und B bleibt das Instrument jeweils gleich. Bei einer Änderung des Registers bzw. der Instrumentenfamilie wird dasselbe Triplet C bzw. D von einem anderen Instrument wiederholt, das sich vom ersten Instrument entweder bzgl. des Registers bzw. der Instrumentenfamilie unterscheidet. Bei einer Änderung der Lokalisation wird das Triplet E zweimal von demselben Instrument gespielt; der Schallpegel ist jeweils auf einer (anderen) Seite um -6 dB reduziert.

Neben der Weiterentwicklung des Paradigmas und der Verwendung von realistischen Instrumentaltönen besteht eine wesentliche Neuerung in der Verbindung von neurowissenschaftlichen und musikwissenschaftlichen Methoden: Die Auswahl der Instrumente erfolgte nicht nur unter Berücksichtigung der Anspieltechnik und des Registers, sondern auch anhand der Lage der Klänge im Formantfeld [14]. Mithilfe der verschiedenen Transitionen zwischen den Triplets wollen wir nachvollziehen, welche Lageänderungen innerhalb des Formantfeldes zu welchen Aktivitätsänderungen im auditorischen Cortex führen.

Literatur

[1] Patterson, R. D. et al.: The Processing of Temporal Pitch and Melody Information in Auditory Cortex. Neuron 36 (2002), 767-776

[2] Andermann, M. et al.: Early cortical processing of pitch height and the role of adaptation and musicality. NeuroImage 225 (2021)

[3] Andermann, M. et al.: Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex. NeuroImage 158 (2017), 79-89

[4] Kawahara, H. et al.: Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. 1999

[5] Kawahara, H.: STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. Acoustical Science and Technology 27, 6 (2006), 349-353

[6] BESA 5.2.: BESA GmbH, Gräfelfing, Deutschland

[7] Irino, T. et al.: Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-Mellin transform. Speech Commun. 36 (2002), 181–203

[8] Patterson, R.D., Irino, T., 2014. Size matters in hearing: how the auditory system normalizes the sounds of speech and music for source size. In: Popper A., Fay R. (eds) Perspectives on Auditory Research. Springer Handbook of Auditory Research, vol 50. Springer, New York, NY

[9] Formisano, E. et al.: "Who" is saying "what"? Brainbased decoding of human voice and speech. Science 322 (2008), 970–973

[10] Nelken, I. et al.: Auditory cortical processing in realworld listening: the auditory system going real. J. Neurosci. 34 (2014), 15135–15138

[11] Siddiq, S. et al.: Timbre Space reloaded: Tonhöhe und Dynamik als Teil der Klangfarbenempfindung. DAGA (2017), 194-197.

[12] Siddiq, S. et al.: Die physikalischen Korrelate von Instrumentalklangfarben. DAGA (2018), 1695-1698.

[13] Meyer, J.: Akustik und musikalische Aufführungspraxis. Bergkirchen: PPV Medien 2015.

[14] Reuter, C. et al.: Formanten als hilfreiche Timbre-Deskriptoren für die Darstellung von Blasinstrumentenklängen.. 43. DAGA (Kiel) 2017, 190–193.