J-029

振幅情報を組み替えた合成音声の脳波を使用した評価と考察

Evaluation of the Synthesized Speech by Using Mismatch Negativity (I)

羽山 雄偉† 吉田 秀樹†
Hayama Yuui Hydecky Yoshda

1．はじめに

音声のスペクトル情報をある程度破壊しても、音声の内容を理解することができる。音声信号に 15 dB 未満の白色雑音を重畳したり[1]、音声や音楽の波形の一部を良く似た正弦波で置き換えれば[2]、元のスペクトル構造は変化するが、内容を理解する上では支障を来さない。同じ音響情報を遜色の無い音質で伝えるのに、様々な時系列波形の変更や修正が許容されており、この意味で、安易に二乗誤差や相関係数を使用した手法では、音質の劣化度を測定するのは難しい。同様に主観評価法[3]を適用する際には、実験を工夫しないと、繰り返される刺激音への慣れや飽きあるいは被験者の意欲の度合いが、評価結果に思わぬ偏りを生み出す可能性がある。そこで我々が着目したのは、脳波を使用した刺激音質の客観的な定量化方法である。ミスマッチ陰性(MMN) 応答[4]は誘発脳波の一成分であり、反復する刺激音列中に低頻度で周波数[5]や強度[6]の異なる音を呈示すると、前頭部から頭頂部を中心にして陰性電位が観察される[7]。MMN 応答は、先行する刺激の特徴を一時的に短期記憶中に蓄えて、後続する刺激音の変化を自動的に検出する精神作用であると考えられている[8]。このことから合成音声の音質の違いを明らかにする上で、MMN 応答の振幅は、被験者の意欲や関心に左右されない指標になることが期待される。

先行研究の中で、音響波形の極大値と極小値の情報が、様々な高さ、大きさ、それに音色の音を奏でる上で、本質的な働きをしていることが提唱された[9]。しかし複雑な形状をした複合音から極値を抽出することは困難であるので、複数の帯域通過フィルターを使用してうなり様の時系列波形にまで簡略化した後に、極値をサンプリングする必要がある。当該周波数帯域を合成するには、例えば隣り合う極値間が正弦波状に変化する様に、欠損したデータ列を補間すれば良い。原波形に比べ、合成波形には特有の歪みを伴うことになるが、歪み成分が音色の変化として知覚されることはない。本研究ではスペクトル成分の代わりに、極値を使用して音響波形の音色構造について理解を深めることを意図している。許容誤差内で極値の計測が出来ない場合には、音質は劣化してざらついたものとして知覚される。極値の振幅情報を操作した合成音声を使用して MMN 応答を計測し、極値の振幅の許容誤差を見積もることを目的とする。

2．実験方法

女性 1 名を含む健常被験者 10 名（年齢 19.1±0.3 歳、全員右利き）がボランティアとして実験に参加した。実験は非侵襲的に実施され、計測脳波は公表とし、計測中でも被験者の意思で実験は随時終了できる旨の充分な説明がなされた。被験者は簡易の静電対策を施した防音室 MC-3（155 x 255 x 210 cm, Music Cabin Co., Ltd.）の中で安静に椅子に座して読書をしながら、ヘッドホンを着用して刺激音を聞き流した。計測中は刺激音を無視させ、考え事も自由とした。刺激音は高頻度刺激(frequent)と低頻度刺激(rare)から成るランダムな刺激列であり、両者の呈示割合は 6:1、音の長さは共に 144 ms、刺激間間隔 600 ms とした。高頻度刺激は音節/ki/であり、44.1 kHz でサンプリングして、80-5,120 Hz に帯域制限したのに対し、低頻度刺激には振幅情報をランダムに変化させた合成音声とした。先行研究[9]によれば、入力音声の有する周波数情報を狭帯域に制限することにより、濾波された波形から極大値と極小値の情報が抽出できる様になる。隣り合う極値間を正弦波様に後から補間することで、音声が合成できることが報告された。同手法に従って、80-5,120 Hz の帯域を 1 オクターブ毎に 6 分割して 6 個のチャンネルとし、それぞれのチャンネルから極値を抽出した。尚、当該帯域中には、音声信号のピッチおよび第 1 、第 2 フォルマント成分を網羅するのに、充分な帯域を有している。

図 1 に任意のチャンネルでの濾波波形の模式図を破線で示す。白丸で示された極値は振幅方向に一定量だけ移動させ、操作後の極値を黒丸で表した。この時、振幅を増大させるか減少させるかはランダムとした。同様の操作を 6 個のチャンネルに含まれる全ての極値について実施した。これにより振幅誤差は abs(A-a)/A で記述する。低頻度刺激としては振幅誤差が 0%の時と、20 %から 50 %まで 5 %刻みで計 8 種類の刺激音を用意して、同一被験者には 8 個の実験課題を課した。この様に波形操作は極値の情報のみで実現し、図 1 の実線に沿って欠落した情報を補った後、6 個のチャンネルを重ね合わせて合成波形とした。刺激音の音圧は約 65 dB SPL で、被験者毎に適宜微調整した。

脳波は国際式 10-20 法に基づく C3（左側頭頂）と C4（右側頭頂）の部位で両耳朶を基準電極として計測し (Digital Bio-Amplifier System 5202, ノイズ < 0.5 μVrms, NR 社製)、0.53-30 Hz の帯域通過フィルター(3 dB down, 12 dB octave/slope)と 50 Hz のノッチフィルターを適用した。高頻度刺激と低頻度刺激のそれぞれに同期させて、刺激前 100 ms から刺激後 400 ms の脳波を 80 回、加算平均処理した。同時に計測した眼電図が 150 μV を超えた場合には、当該期間の脳波を加算平均から除外した。脳波は非線形システムである脳が生成する電気信号であるが、MMN 振幅を測定する目的で慣習に従って、低頻度刺激に対する応答波形から、高頻度刺激に対する応答波形を引き算して算出した[7]。MMN 振幅について 2×8 ANOVA（分散分析法、部位{左側、右側}×実験課題 8 種類）を適用して有意水準 $p<0.01$ で下位検定（Tukey の方法）を実施した。

3．結果

図 2 に低頻度刺激と高頻度刺激についての加算平均波形をそれぞれ実線と薄線で重ねて示す。両波形の振幅は、潜

†北見工業大学大学院 情報システム工学専攻

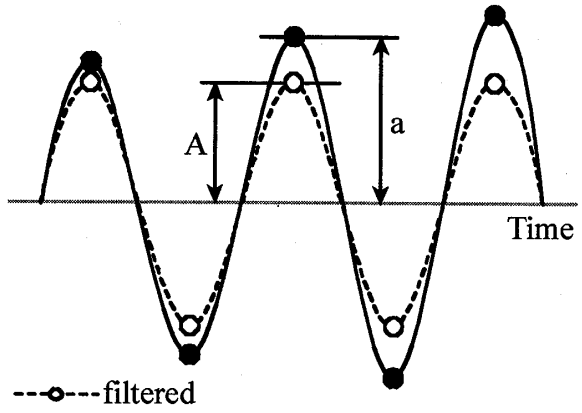

Fig.1 Quantization error, abs(A-a)/A

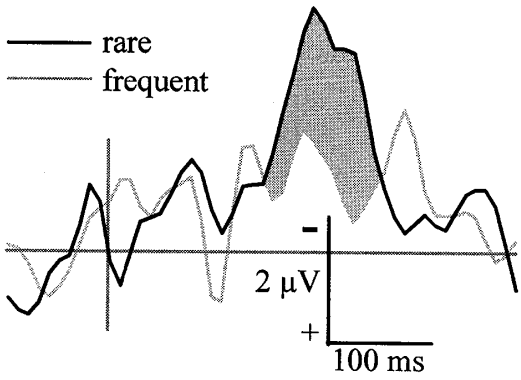

Fig.2 MMN component

時 200 ミリ秒前後を中心に差異が観察されており、着色した部分が MMN 成分と考えられる。刺激音として純音の代わりに音声を使用した実験では、MMN の潜時が遅れる傾向にある。

図 3 には合計 8 課題について観察された MMN 振幅の平均値と標準偏差を示す。MMN 振幅は実験課題により有意に変化しており[F(7, 63)=10.2]、下位検定により MMN 振幅は振幅誤差が 35 %以上の時、有意に増大していることが示されている[課題間の振幅誤差が 0 %から 35 %の時 t_0=1.7、0-40 %時 t_0=2.2、0-45 %時 t_0=2.0、0-50 %時 t_0=2.2]。計測部位による差異と[F(1, 9)=2.0]、課題と計測部位の交互作用[F(7, 63)=0.37]については、影響が観察されていない。

4. 考察

高頻度刺激に原音を使用して、低頻度刺激には振幅を変化させなかった極値を使用した場合には、高振幅の MMN 応答は観察されなかった。これは刺激音の差異が、脳自身が定めた弁別閾値未満であったことを意味している。本結果は主観評価（Scheffe の一対評価法）を使用した先行報告[10]と一致しており、極値を使用した音声合成法が、原音の音韻と等価な合成音を生成できることを示唆している。振幅誤差は 35 %以上あると MMN 振幅（絶対値）が有意に増大した。これも振幅の許容誤差は 20 %から 40 %の間にあるとする先行報告[10]を支持する結果であり、指標に MMN を導入することで、許容誤差の閾値をより詳細に見積れる様になった。MMN 応答は、純音ならば僅か 1,000 Hz と 1,016 Hz の違いを自動識別するだけの検出力[11]があることからも、観察結果は裏付けられると云えよう。音声信号の重要な成分は包絡線（振幅包絡）の形状にあると考えられている[12, 13]。そこで極値の振幅誤差を増大する操作は、この包絡線形状の破壊に他ならない。興味深い点は、音声信号は振幅方向の外乱に頑健であり、振幅情報が 3 割変化してもその変化に気付かぬ程、聴覚は鈍感にできている。そうして極値の振幅を閾値を超えてランダムに増減させると、音色自体が消失するのではなく、ただざらざらとした聴き取り辛さが急激に耳に残る様になる。

MMN 応答を音質評価の指標とする利点は、被験者に刺激音への注意を課さないことから、被験者の精神的負担を軽減することにある。ただし脳波計測全般に云えることで

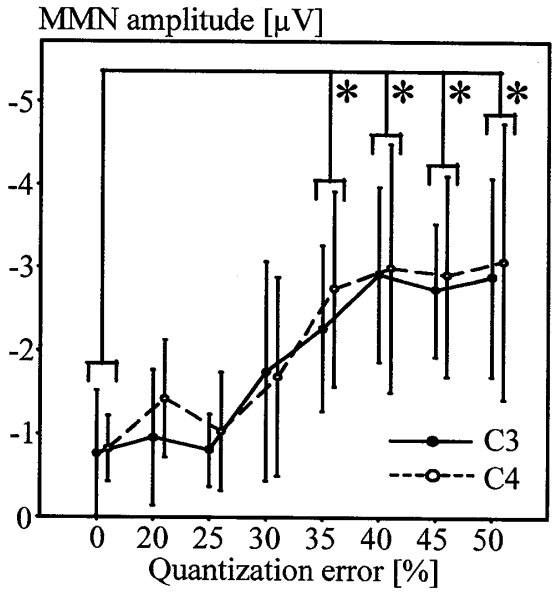

Fig.3 MMN amplitude, *p<.01

あるが、被験者が瞬目や眼球の動きをなるべく堪えて、安静状態で計測に協力する姿勢や、使用できる刺激にも制約があることは否めない。MMN より後潜時には注意、驚き、記憶の更新に関連した誘発脳波を計測することもできる。この様な高次の精神活動を反映すればする程、脳波は意識や意欲と云った精神状態によっても影響を受け、脳内の活動源も聴覚野を離れた複数の部位が複雑に連絡を取る様になる[14-29]。これに対し MMN の電源活動の中心は、聴覚野(Brodmann's area 41)前方約 10 mm に局在していると考えられている[30]。当該部位は一次聴覚野（ヘッシュル頭回および Planum temporale）との連絡が密で、先行して呈示された刺激の物理的な属性と、入力刺激との比較照合処理を自動的に行っていることになる。MMN 応答は聴覚野近傍での情報処理を観察しており、被験者の意欲や意識の影響を受けることが少ないので、客観的な音質評価の指標としての利用が期待される。

5. まとめ

極値の振幅の許容誤差は 35 %未満であり、閾値は 30 %から 35 %の間に見積もられた。

参 考 文 献

[1] Makhoul, J. and Berouti, M. (1979): "Adaptive Noise Spectral Shaping and Entropy Coding in Predictive Coding of Speech", IEEE Trans. on ASSP, Vol.27, No.1, pp. 63-73.

[2] McAulay, R. J. & Quatieri, T. F. (1986): "Speech Analysis/Synthesis Based on a Sinusoidal Representation", IEEE Tran. on ASSP, Vol.34, No.4, pp. 744-754.

[3] Thurstone, L. L. (1947): "Multiple-Factor Analysis -A Development and Expansion of the Vectors of Mind, The University of Chicago Press, Chicago Illinois.

[4] Naatanen, R., Gaillard, A. W. K. and Mantysalo, S. (1978): "Early Selective Attention Effect on Evoked Potential Reinterpreted", Acta Psychologica, Vol.42, pp. 313-329.

[5] Fitzgerald, P. G. and Picton, T. W. (1983): "Event-Related Potentials Recorded during the Discrimination of Improbable Stimuli", Biological Psychology, Vol.17, pp. 241-276.

[6] Naatanen, R., Paavilainen, P., Alho, K., Reinikainen, K. and Sams, M. (1989): "Do Event-related Potentials Reveal the Mechanism of the Auditory Sensory Memory in the Human Brain?", Neuroscience Letters, Vol.98, pp. 217-221.

[7] Naatanen, R. and Winkler, I. (1999): "The Concept of Auditory Stimulus Representation in Cognitive Neuroscience", Psychological Bulletin, Vol.125, No.6, pp. 826-859.

[8] Naatanen, R. (1990): "The Role of Attention in Auditory Information Processing as Revealed by Event-Related Potentials and Other Measurements of Cognitive Functions", Behav. Brain Sci., Vol.13, pp. 201-288.

[9] 入場健仁、吉田秀樹、藤原祥隆、岡田信一郎(2004):リアルタイム音響構造配信システムの開発 情報技術レターズ, LK-009, pp. 267-268.

[10] 吉田秀樹、角井健二、前田康成、藤原祥隆(2008): 極値サンプリング技術と許容誤差– wavファイルからの情報抽出– バイオメディカル・ファジイ・システム学会誌 Vol.10(2), pp. 123-131.

[11] Sams, M., Paavilainen, P., Alho, K. (1985): "Auditory Frequency Discrimination and Event-Related Potentials", Electroencephalography and Clinical Neurophysiology, Vol.62, pp. 437-448.

[12] Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M. (1995): "Speech Recognition with Primarily Temporal Cues", Science, Vol.270, pp. 303-304.

[13] Zachary, M. S., Bertrand, D. and Andrew, J. O. (2002): "Chimaeric sounds reveal dichotomies in auditory perception", Nature, Vol.416, pp. 87-90.

[14] Okada, Y. W., Kaufman, L. and Williamson, S. J. (1993): "The Hippocampal Formation as a Source of the Slow Endogeneous Potentials", Electroencephalography and Clinical Neurophysiology, Vol.55, pp. 417-426.

[15] Richer, F., Johnson, R. A. and Beatty, J. (1993): "Sources of Late Components of the Brain Magnetic Response", Soc. Neurosci. Abstr., Vol.9 p. 656.

[16] Wood., C. C., McCarthy, G., Squires, N. K., Vaughan, H. G., Woods, D. L. and McCallum, W. C. (1984): "Anatomical and Physiological Substrates of Event-Related Potentials: Two Case Studies", Ann. N.Y. Acad. Sci., Vol.425, pp. 681-721.

[17] Kikuchi, Y., Endo, H., Yoshizawa, S., Kait, M., Nishimura, C., Tanaka, M., Kumagai, T. and Takeda, T., (1997): "Human Cortico-hippocampal Activity Related to Auditory Discrimination Revealed by Neuromagnetic Field", NeuroReport, Vol.8, No.7, pp. 1657-1661.

[18] Knight, R. T., (1996): "Contribution of Human Hippocampal Region to Novelty Detection", Nature, Vol.383, pp. 256-259.

[19] Mecklinger, A. and Ullsperger, P., (1995): "The P300 to Novel and Target Events: A Spatio-temporal Dipole Model Analysis", NeuroReport, Vol.7, No.1, pp. 241-245.

[20] Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P. and Musolino, A., (1995): "Intracerebral Potentials to Rare Target and Distractor Auditory and Visual Stimuli. I. Superior Temporal Plane and Parietal Lobe", Electroencephalography and Clinical Neurophysiology, Vol. 94, pp. 191-220.

[21] Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., Vignal, J-P. and Biraben, A., (1995): "Intracerebral Potentials to Rare Target and Distractor Auditory and Visual Stimuli. II. Medial, Lateral and Posterior Temporal Lobe", Electroencephalography and Clinical Neurophysiology, Vol. 94, pp. 229-250.

[22] Baudena, P., Halgren, E., Heit, G. and Clarke, J. M., (1995): "Intracerebral Potentials to Rare Target and Distractor Auditory and Visual Stimuli. III. Frontal Cortex", Electroencephalography and Clinical Neurophysiology, Vol. 94, pp. 251-264.

[23] Tarkka, I. M., Stokic, D. S., Basile, L. F. H. and Papanicolaou, A. C. (1995): "Electric Source Localization of the Auditory P300 Agrees with Magnetic Source Localization", Electroencephalography and Clinical Neurophysiology, Vol. 96, pp. 538-545.

[24] O'Donnell, B. F., Cohen, R. A., Hokama, H., Cuffin, B. N., Lippa, C., Shenton, M. E. and Drachman, D. A. (1993): "Electrical Source Analysis of Auditory ERPs in Medial Temporal Lobe Amnestic Syndrome", Electroencephalography and Clinical Neurophysiology, Vol. 87, pp. 394-402.

[25] Polich, J. and Squire, L. R. (1993): "P300 from Amnesic Patients with Bilateral Hippocampal Lesions", Electroencephalography and Clinical Neurophysiology, Vol. 86, pp. 408-417.

[26] Rogers, R. L., Baumann, S. B., Papanicolaou, A. C., Bourbon, T. W., Alagarsamy, S. and Eisenberg, H. M. (1991): "Localization of the P3 Sources Using Magnetoencephalography and Magnetic Resonance Imaging", Electroencephalography and Clinical Neurophysiology, Vol. 79, pp. 308-321.

[27] Knight, R. T., Scabini, D., Woods, D. L. and Clayworth, C. C. (1989): "Contributions of Temporal-parietal Junction to the Human Auditory P3", Brain Research, Vol. 502, pp. 109-116.

[28] McCarthy, G., Wood, C. C., Williamson, P. D. and Spencer, D. D. (1989): "Task-Dependent Field Potentials in Human Hippocampal Formation", J. NeuroSci., Vol.9, No.12, pp. 4253-4268.

[29] Stapleton, J. M. and Halgren, E. (1987): "Endogenous Potentials Evoked in Simple Cognitive Tasks: Depth Components and Task Correlates", Electroencephalography and Clinical Neurophysiology, Vol. 67, pp. 44-52.

[30] Hari, R., Hamalainen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., Alho, K., Naatanen, R. and Sams, M. (1984): "Responses of the Primary Auditory

Cortex to Pitch Changes of Tone Pips: Neuromagnetic Recordings in Man", Evoked by Short Auditory Stimuli", Neuroscience Letters, Vol.50, pp. 127-132.