令和３年度研究助成【サウンド技術振興部門】より

音声理解における脳内時間処理：時間劣化音声を用いた統合失調症患者と健常者との比較

九州大学大学院芸術工学研究院
デザイン人間科学部門
准教授
上田　和夫

１．研究の背景

１．１　音声の冗長性と音声知覚の頑健性

音声は、時間とともに周波数スペクトルが変化する信号であると見なすことができる［図１(a)］。したがって、音声知覚の仕組みを調べるためには、そのような信号からどのようにして適切な知覚手がかりが取り出されているのか、また取り出された知覚手がかりがどのように使われているのかを調べる必要がある。

通常音声には、極めて豊富な知覚手がかりが重複して含まれていると考えられている。すなわち、音声は非常に冗長性の高い信号であると考えられている。そのように考えられる根拠として、音声が様々な劣化に対して耐性を持ち、かなりな程度の削除や変形が行われても、音声知覚が可能であることがあげられる。たとえば、フィルターによる周波数帯域の制限や（French & Steinberg, 1947; Fletcher & Galt, 1950; Miller & Nicely, 1955; Studebaker, Pavlovic, & Sherbecoe, 1987; Warren, Bashford, & Lenz, 2005; Humes & Kidd, 2016）、時間・周波数領域における変調の劣化（Drullman, Festen, & Plomp, 1994b, 1994a; ter Keurs, Festen, & Plomp, 1992, 1993; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Smith, Delgutte, & Oxenham, 2002; Zeng et al., 2004; Kidd, Streeter, Ihlefeld, Maddox, & Mason, 2009; Jørgensen & Dau, 2011; Venezia, Hickok, & Richards, 2016; Flinker, Doyle, Mehta, Devinsky, & Poeppel, 2019; Nakajima, Matsuda, Ueda, & Remijn, 2018; Schlittenlacher, Staab, Çelebi, Samel, & Ellermeier, 2019; Santi, Nakajima, Ueda, & Remijn, 2020）といった操作に対して、音声知覚は頑健性を示す。身近な例としては、固定電話システムが300 Hzから3,400 Hzまでのごく限られた周波数帯域の信号しか通過させないにもかかわらず、支障なく音声通話が成り立っていることがあげられる。

１．２　時間劣化音声の知覚

このような音声知覚の頑健性に、時間領域における手がかりの劣化がどのような影響を与えるのかについても、古くから調べられている。たとえば、周期的断続（Miller & Licklider, 1950; Powers & Wilcox, 1977; Kidd & Humes, 2012; Shafiro, Sheft, & Risley, 2016; Shafiro, Fogerty, Smith, & Sheft, 2018; Ueda, Kawakami, & Takeichi, 2021）と局部時間反転（Steffen & Werani, 1994; Saberi & Perrott, 1999; Greenberg & Arai, 2004; Ishida, Arai, & Kashino, 2018; Stilp, Kiefte, Alexander, &

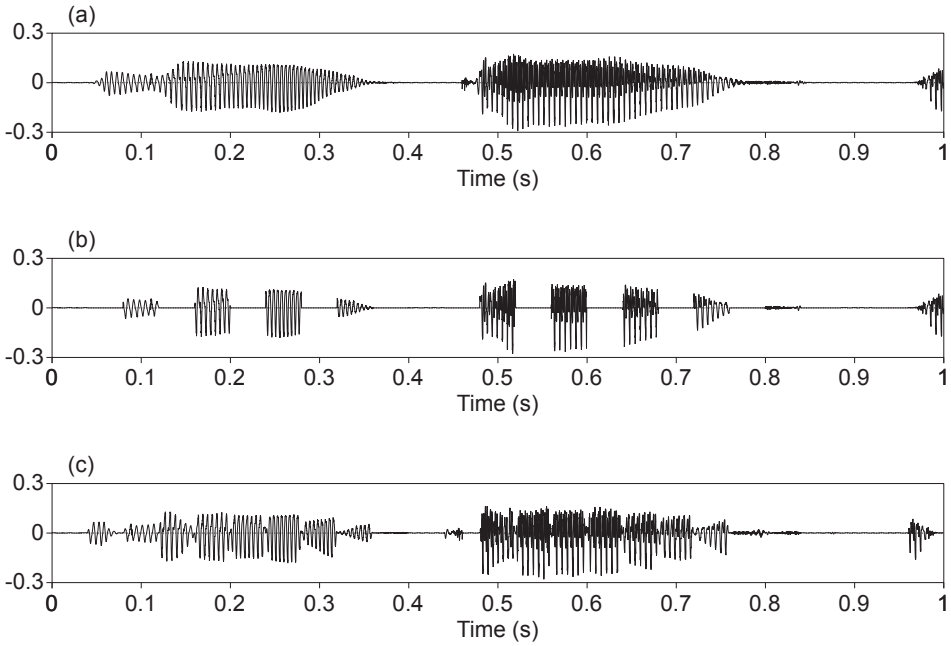

図１　音声波形の例。横軸は時間、縦軸は瞬時音圧（単位は任意）を表す。(a)女性話者が発話した英語原音声、(b)区間長40ミリ秒の断続音声、(c)区間長40ミリ秒の局部時間反転音声。原音声(a)は、時間とともに成分の周波数および振幅（スペクトル）が変化する信号である。断続音声(b)は、音声を出力する区間と、音声を出力せず空白に置換した区間とを交代させたもの。局部時間反転音声(c)は、区間内の波形の時間軸を逆転させて作った断片をつなぎ合わせたもの。断続音声、局部時間反転音声とも、各区間の始まりと終わりの部分に５ミリ秒の立ち上がりと立ち下がりをコサイン関数で付けている。

Kluender, 2010; Teng, Cogan, & Poeppel, 2019; Ueda, Nakajima, Ellermeier, & Kattner, 2017; Ueda, Nakajima, Kattner, & Ellermeier, 2019; Matsuo, Ueda, & Nakajima, 2020; Ueda & Ciocca, 2021）は、時間領域における音声の劣化手法の例である。

周期的断続は、音声を元のまま出力する区間と、音声を出力せず空白に置換する区間とを交互に繰り返すものである［図１(b)］。音声を出力する区間と空白に置換する区間とが同じ長さの場合、原音声の50%が残ることになる。区間長がおよそ100ミリ秒以下の場合は、80%以上の明瞭度あるいは了解度で音声を聴きとることができる（たとえば、Miller & Licklider, 1950）。区間長が長くなるにしたがって、了解度は低下していくが、健聴な実験参加者に対して有意味文の音声を断続した場合、区間長650ミリ秒でも了解度は50%を上回ることがわかっている（たとえばPowers & Wilcox, 1977）。

局部時間反転音声［図１(c)］は、音声を一定の区間長で区切った後、それぞれの区間ごとに時間軸を反転して部分的に逆再生の状態を作り、反転した区間をもとの順番通りにつなぎ合わせた劣化音声である。当然であるが、話速が了解度に影響するので、以下に示す区間長と了解度との関係は、話速を正規化した場合になり

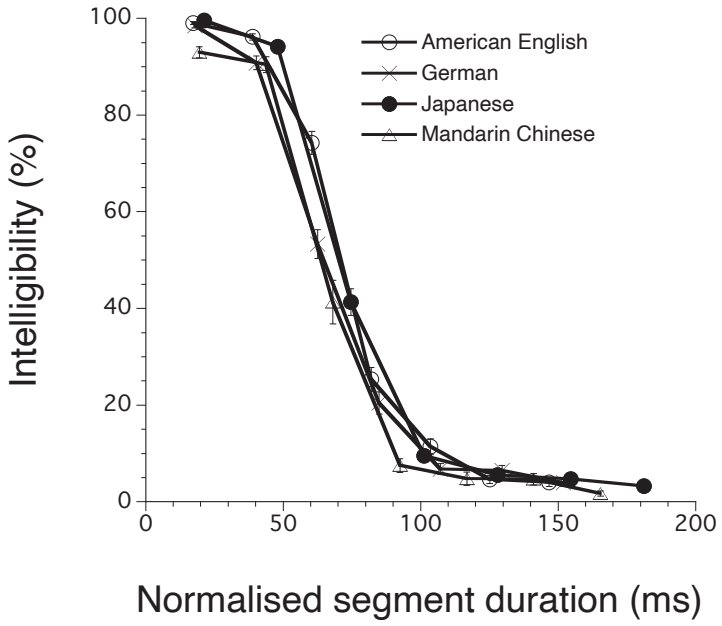

図２　英独日中の４言語における局部時間反転音声の了解度。それぞれの言語ごとに男女各１名の話者の音声を刺激として使い、話者ごとに14名の実験参加者からデータを得て、それらを平均している（中国語の男性話者のデータのみ13名の実験参加者のデータ）。横軸の区間長は、それぞれの言語における話者全員（男女それぞれ５名または10名）の平均話速で正規化されている。区間長と了解度との関係は、言語によらず、極めて類似した傾向を示す。エラーバーは標準誤差。Ueda et al.(2017）による。

立つものであることをご理解いただきたい。有意味文を用いて、区間長が40ミリ秒以下の場合、局部時間反転音声でも100%近い了解度が得られるが、区間長が40ミリ秒を越えると、言語にかかわらず、急激に了解度が低下し、区間長65ミリ秒で了解度がおよそ50%程度、区間長が100ミリ秒を越えると了解度は10%以下にまで低下する（図２; Ueda et al., 2017）。局部時間反転音声の区間長が40ミリ秒を超えると了解度が急激に低下するのは、周波数スペクトル全体にわたる変調スペクトルの強さと位相とが共に減少していくことと関係すると考えられている（Greenberg & Arai, 2004）。

では、局部時間反転音声を断続したらどうなるのかというと、区間長20ミリ秒の場合を除いて、著しい了解度の低下が生ずる（図３; Ueda & Ciocca, 2021）。たとえば、区間長40ミリ秒の場合、断続なしの局部時間反転音声なら了解度はほぼ100%であるが、区間を一つおきに空白と置き換えると、了解度は約45%まで低下する。区間長60ミリ秒の場合、断続なしの局部時間反転音声なら了解度はおよそ80%になるが断続すれば約10%にまで低下する。断続により、もとの局部時間反転音声と比較すれば、刺激の半分が失われただけであることから考えると、それ以上に了解度が低下したことになる。ここで、局部時間反転断続音声の空白を強い雑音と置き換えれば、音声の情報は何も増えていない

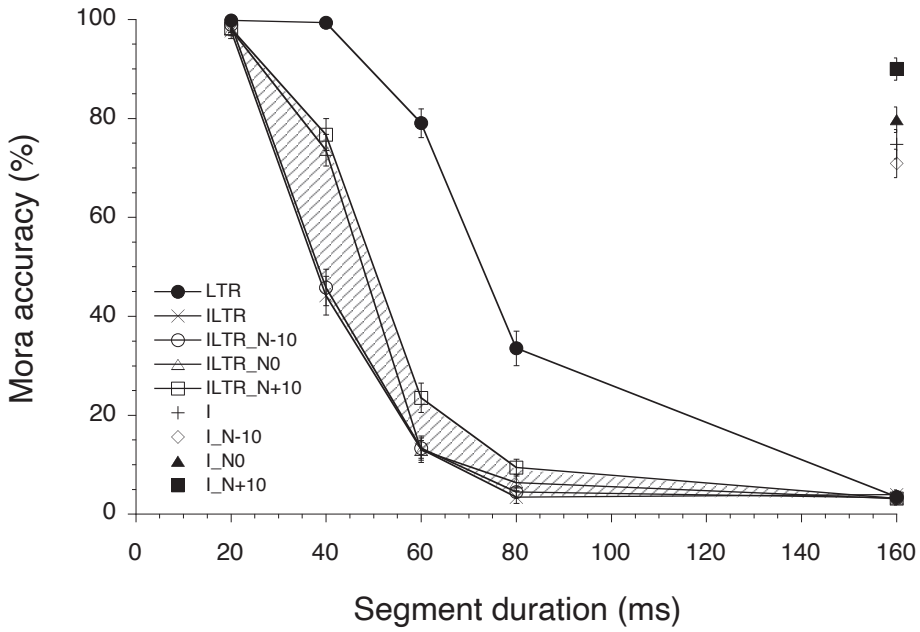

図３　局部時間反転音声（locally time-reversed speech, LTR）、局部時間反転断続音声（interrupted locally time-reversed speech, ILTR）、および断続音声（interrupted speech, I）の了解度（20名の実験参加者について求めたモーラ正答率）。横軸は区間長を表す。Nは雑音（noise）で空白を置換したことを意味する。−10, 0, +10は音声パワーの実効値を基準（0）としたときの雑音の強さ（dB）。斜線部は、空白を+10 dBの雑音で置き換えることによる了解度の増加分を示す。エラーバーは標準誤差。Ueda & Ciocca（2021）による。

にもかかわらず、了解度は回復し（図３の斜線部分）、その増分は最大40%程度に達する（区間長40ミリ秒の場合）。

断続音声の場合、区間長40ミリ秒では了解度は90%以上あり（Ueda et al., 2021）、区間長160ミリ秒でも70%前後である（Ueda & Ciocca, 2021; Ueda et al., 2021）。断続音声でも、空白部分を強い雑音で置換することにより了解度が向上するが、区間長160ミリ秒の場合で10%程度の増加にしかならない（Ueda & Ciocca, 2021）。断続音声の場合は、もちろん、もともとの了解度が高いので、何らかの操作によって了解度がさらに向上する余地が少ないとも言える。いずれにせよ、局部時間反転音声の場合、断続による了解度の低下が顕著であり、空白部分を雑音で置換することの効果も大きいと言える。このような知覚的修復による了解度の向上は、音声知覚においてほぼ自動的に行われる処理過程を反映していると考えられる。

１．３　多重時間窓仮説

区間長が短い場合、断続音声や局部時間反転音声が高い了解度で知覚される理由として、脳内における音声の処理に一定の長さの時間窓が用いられ、時間窓単位で処理が行われるためと考えることが可能である。

Poeppelらの提唱する多重時間窓モデルでは、短い時間窓（～20－30ミリ秒、音素の長さ

とほぼ対応する）と長い時間窓（～200ミリ秒、音節の長さとほぼ対応する）とが脳内に存在し、並行して音声（あるいは非音声）の処理を進めると考えられている（Poeppel, 2003; Giraud & Poeppel, 2012; Sanders & Poeppel, 2007; Chait, Greenberg, Arai, Simon, & Poeppel, 2015; Teng, Tian, & Poeppel, 2016; Teng & Poeppel, 2020）。このモデルをあてはめて考えてみると、区間長20ミリ秒の断続音声や局部時間反転音声で（図２、３）、ほとんど100%近い正答率が得られることをうまく説明できるように見える。このような刺激は、どちらの時間窓でもうまく処理できると考えられるからである。また、断続音声や局部時間反転音声の正答率が、区間長が長くなるにつれて低下していくことを説明することもできる。

一方で、局部時間反転断続音声の実験結果は（図３）、40ミリ秒以上の区間長による断続が局部時間反転音声に生じた場合には、断続音声や局部時間反転音声に対して働いていた知覚的修復がうまくいかなくなること、また強い雑音で空白を置換することによって多少なりとも修復が起こることを示した。すなわち、多重時間窓仮説においても、このような知覚的修復の過程をモデルに組み込む必要がある。さらに、これらの実験結果は、短い時間窓における処理結果が長い時間窓による処理に影響を与える可能性があることも示している。

１．４　統合失調症患者の脳波および脳磁図

統合失調症患者においては、知覚から認知にいたる幅広い神経活動に関して、脳活動の周期性に異常が見られるとの報告がなされている（Hirano & Uhlhaas, 2021）。特に聴覚や音声知覚に関しては、統合失調症患者では聴覚信号に対する40Hz帯（ガンマ帯の一部）の脳波あるいは脳磁図に異常が見られるとの報告がある。周期に換算すると40Hzは25ミリ秒であるので、上記の多重時間窓仮説における短い時間窓における処理に異常が生じている可能性があると考えられる。さらに、音節の長さに対応する長い時間窓に関しては、同様の周期で動作する発話運動系の発振回路が知覚処理系からのフィードバックを受けるというモデルも提案されているので（Poeppel & Assaneo, 2020）、このような異常を、劣化音声を用いた知覚実験で検出できる可能性があると考えた。

２．研究予定

本研究計画では、統合失調症患者が幻聴と現実の音声とを区別できず、著しく生活の質が低下しているのは、現実の音声を短い時間窓でうまく処理できず、幻聴を制御できないためであるという仮説を立てた。そこで、短い時間窓による処理が難しい、時間劣化音声を用いて統合失調症患者と健常者との知覚の違いについて調べる予定である。

文献

Chait, M., Greenberg, S., Arai, T., Simon, J. Z., & Poeppel, D. (2015). Multi-time resolution analysis of speech: Evidence from psychophysics. *Frontiers in Neuroscience*, 9(214), 1-10. doi:10.3389/fnins.2015.00214

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, 95(2), 1053-1064. doi:10.1121/1.408467

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, 95(5), 2670-2680. doi:10.1121/1.409836

Fletcher, H., & Galt, R. H. (1950). The perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America*, 22(2), 89-151. doi:10.1121/1.1906605

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour*, 3(4), 393-405. doi:10.1038/s41562-019-0548-z

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America*, 19, 90-119.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emergning computational principles and operations. *Nature Neuroscience*, 15(4), 511-517. doi:10.1038/nn.3063

Greenberg, S., & Arai, T. (2004). What are the essential cues for understanding spoken language? *IEICE Transactions on Information and Systems*, E87-D (5), 1059-1070.

Hirano, Y., & Uhlhaas, P. J. (2021). Current findings and perspectives on aberrant neural oscillations in schizophrenia. *Psychiatry and Clinical Neurosciences*, n/a (n/a). doi: 10.1111/pcn.13300

Humes, L. E., & Kidd, G. R. (2016). Speech recognition for multiple bands: Implications for the Speech Intelligibility Index. *The Journal of the Acoustical Society of America*, 140(3), 2019-2026. doi:10.1121/1.4962539

Ishida, M., Arai, T., & Kashino, M. (2018). Perceptual restoration of temporally distort-

ed speech in L1 vs. L2: Local time reversal and modulation filtering. *Frontiers in Psychology*, 9(1749), 1-16. doi:10.3389/fpsyg.2018.01749

Jørgensen, S., & Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, 130(3), 1475-1487. doi:10.1121/1.3621502

Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *The Journal of the Acoustical Society of America*, 131(2), 1434-1448. doi:10.1121/1.3675975

Kidd, G., Streeter, T. M., Ihlefeld, A., Maddox, R. K., & Mason, C. R. (2009). The intelligibility of pointillistic speech. *The Journal of the Acoustical Society of America*, 126(6), EL196-EL201. doi:10.1121/1.3258062

Matsuo, I., Ueda, K., & Nakajima, Y. (2020). Intelligibility of chimeric locally time-reversed speech. *The Journal of the Acoustical Society of America*, 147(6), EL523-EL528. doi:10.1121/10.0001414

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22(2), 167-173. doi:10.1121/1.1906584

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27, 338-352.

Nakajima, Y., Matsuda, M., Ueda, K., & Remijn, G. B. (2018). Temporal resolution needed for auditory communication: Measurement with mosaic speech. *Frontiers in Human Neuroscience,* 12(149), 1-8. doi:10.3389/fnhum.2018.00149

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41, 245-255.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 322-334. doi:10.1038/s41583-020-0304-4

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *The Journal of the Acoustical Society*

of America, 61(1), 195-199. doi:10.1121/1.381255

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, 398(29 APRIL 1999), 760.

Sanders, L. D., & Poeppel, D. (2007). Local and global auditory processing: Behavioral and ERP evidence. *Neuropsychologia*, 45(6), 1172-1186. doi:10.1016/j.neuropsychologia.2006.10.010

Santi, Nakajima, Y., Ueda, K., & Remijn, G. B. (2020). Intelligibility of English mosaic speech: Comparison between native and non-native speakers of English. *Applied Sciences*, 10(19), 1-13. doi:10.3390/app10196920

Schlittenlacher, J., Staab, K., Çelebi, Ö., Samel, A., & Ellermeier, W. (2019). Determinants of the irrelevant speech effect: Changes in spectrum and envelope. *The Journal of the Acoustical Society of America*, 145(6), 3625-3632. doi:10.1121/1.5111749

Shafiro, V., Fogerty, D., Smith, K., & Sheft, S. (2018). Perceptual organization of interrupted speech and text. *Journal of Speech, Language, and Hearing Research*, 61(10), 2578-2588. doi:10.1044/2018_JSLHR-H-17-0477

Shafiro, V., Sheft, S., & Risley, R. (2016). The intelligibility of interrupted and temporally altered speech: Effects of context, age, and hearing loss. *The Journal of the Acoustical Society of America*, 139, 455-465. doi: 10.1121/1.4939891

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303-304. doi:10.1126/science.270.5234.303

Steffen, A., & Werani, A. (1994). Ein Experiment zur Zeitverarbeitung bei der Sprach wahrnehmung (An experiment on temporal processing in speech perception). In G. Kegel, T. Arnhold, K. Dahlmeier, G. Schmid, & B. Tischer (Eds.), *Sprechwissenschaft & Psycholinguistik* (Speech Science and Psycholinguistics) (Vol. 6, pp.189-205). Opladen: Westdeutscher Verlag.

Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *The Journal of the Acoustical Society of America*, 128, 2112-2126.

doi:10.1121/1.3483719

Studebaker, G. A., Pavlovic, C. V., & Sherbecoe, R. L. (1987). A frequency importance function for continuous discourse. *The Journal of the Acoustical Society of America*, 81(4), 1130-1138.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416 (7 March 2002), 87-90.

Teng, X., Cogan, G. B., & Poeppel, D. (2019). Speech fine structure contains critical temporal cues to support speech segmentation. *NeuroImage*, 202(116152), 1-12. doi: 10.1016/j.neuroimage.2019.116152

Teng, X., Tian, X., & Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Scientific Reports*, 6(34390), 1-13. doi:10.1038/srep34390

Teng, X., & Poeppel, D. (2020). Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cerebral Cortex*, 30(4), 2600-2614. doi:10.1093/cercor/bhz263

ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America*, 91, 2872-2880.

ter Keurs, M., Festen, J. M., & Plomp, R. (1993). Effect of spectral envelope smearing on speech reception. II. *The Journal of the Acoustical Society of America*, 93, 1547-1552.

Ueda, K., & Ciocca, V. (2021). Phonemic restoration of interrupted locally time-reversed speech: Effects of segment duration and noise levels. *Attention, Perception, & Psychophysics*, 83(5), 1928-1934. doi:10.3758/s13414-021-02292-3

Ueda, K., Kawakami, R., & Takeichi, H. (2021). Checkerboard speech vs interrupted speech: Effects of spectrotemporal segmentation on intelligibility. *JASA Express Letters*, 1(7), 075204. doi:10.1121/10.0005600

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports*, 7(1782), 1-8. doi:10.1038/s41598-017-01831-z

Ueda, K., Nakajima, Y., Kattner, F., & Ellermeier, W. (2019). Irrelevant speech effects with locally time-reversed speech: Native vs non-native language. *The Journal of the Acoustical Society of America*,

145(6), 3686-3694. doi:10.1121/1.5112774

Ueda, K., & Matsuo, I. (2021). Intelligibility of chimeric locally time-reversed speech: Relative contribution of four frequency bands. *JASA Express Letters*, 1(6), 065201. doi:10.1121/10.0005439

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory "bubbles": Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America*, 140(2), 1072-1088. doi:10.1121/1.4960544

Warren, R. M., James A. Bashford, J., & Lenz, P. W. (2005). Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired. *The Journal of the Acoustical Society of America*, 118(5), 3261-3266.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y.-Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). *The Journal of the Acoustical Society of America*, 116, 1351-1354.