IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 18, 2021, accepted December 12, 2021, date of publication January 18, 2022, date of current version January 28, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3143882*

# Auditory Spatial Saliency and Its Effects on Perceptual Noisiness

**YUKI NAKATANI, MASAYUKI WATANABE, AND NAOKO YOROZU**
Technical Research Center, Mazda Motor Corporation, Hiroshima 730-8670, Japan

Corresponding authors: Yuki Nakatani (nakatani.yu@mazda.co.jp), Masayuki Watanabe (watanabe.masay@mazda.co.jp), and Naoko Yorozu (yorozu.n@mazda.co.jp)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Mazda Motor Corporation under Approval No. TRC-150-3.

**ABSTRACT** Environmental noise affects significantly our health and quality of life. Simple techniques, such as A-weighted decibels, have been applied commonly to the real issues of environmental noise. More elaborate techniques, considering mechanisms in the central nervous system, have also been developed continuously and approved by the international organization of standardization (e.g., ISO 532-3; hereafter, ISO loudness model). These techniques have advanced our knowledge of perceptual noisiness, but still have some limitations to account for a variety of psychophysical phenomena and our empirical experiences in acoustic engineering. Here, we propose that perceptual noisiness can be explained better by considering auditory attention. Attention driven by sensory input has been modeled originally as "saliency" in vision. This algorithm has also been applied to capture spectral-temporal dynamics of auditory attention (hereafter, spectral saliency). It has been suggested that the central auditory system contains two pathways identifying what and where a sound source is. The above spectral saliency corresponds only to the what-pathway. We therefore created a new auditory spatial saliency model to capture attentional effects along the where-pathway based on an algorithm of horizontal sound localization. We found that our spatial saliency model accounted for perceptual phenomena that cannot be explained by the ISO loudness model. Furthermore, the prediction of perceptual noisiness of environmental sounds (driving sounds of passenger cars) was improved significantly by integrating spatial saliency with ISO loudness. We conclude that spatial saliency can be used to capture sound features affecting perceptual noisiness in everyday life.

**INDEX TERMS** Acoustic signal processing, attention, automotive engineering, biomedical acoustics, loudness perception, spatial filters.

## I. INTRODUCTION

Environmental noise is a critical issue in our industrialized life. Continuous exposures to loud noises (e.g., airports and traffics) could induce a variety of health issues [1]. Operating noise from home appliances (e.g., refrigerators and computers) also influences our quality of life. Accordingly, evaluation of perceptual noisiness has been an important issue in medicine and engineering.

Simple techniques have been used for the convenient evaluation of perceptual noisiness (e.g., A-weighted decibels [2]). A more elaborate model has also been developed (ISO 532-3; hereafter ISO loudness model) [3], taking into account auditory periphery and some operations in the central nervous system [4]–[6]. Indeed, its predictions agree with the temporal dynamics of neural signals in the auditory cortex [7].

The ISO loudness model has advanced our knowledge of perceptual loudness, but still has some limitations [4]. We have also experienced empirically through the development of passenger cars that the model does not necessarily generate predictions matched fully with the impressions of our expert engineers and customers. We therefore speculated that there would be missing factors that could improve the ISO loudness model.

Auditory attention is a candidate of such missing factors [8], [9]. There are several features capturing our attention automatically. One of the well-known features is the temporal

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

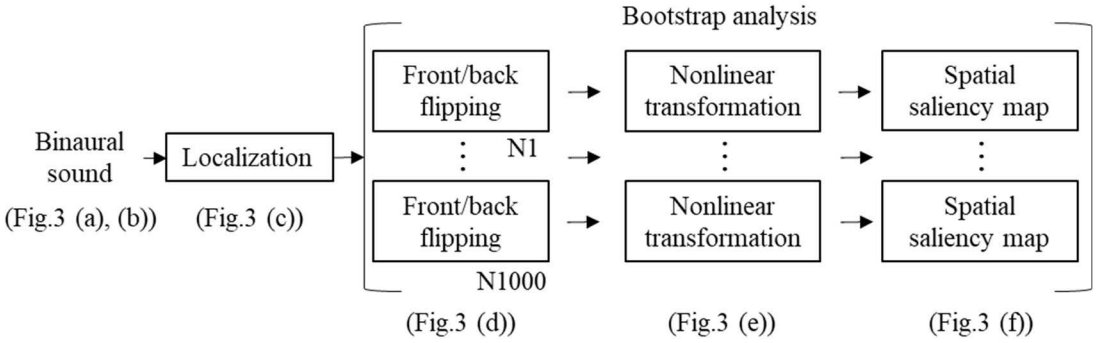



**FIGURE 1.** Overview of algorithms transforming a binaural sound into a spatial saliency map. The role of each algorithm is as follows. Binaural sounds were first fed to the localization model to create a linear localization map. The front/back random flipping was then repeated 1000 times to take into account a cone of confusion. Then, each map was transformed into a nonlinear localization map reflecting the nonlinear property of the superior colliculus. Spatial-temporal contrasts were calculated to create a spatial saliency map.

dynamics of auditory signals [10], [11], considered already in the ISO loudness model. We extend this viewpoint of auditory attention further by the concept of ''saliency'' developed originally in vision [12], [13].

Visual saliency calculates contrasts in individual features (e.g., color and orientation) and integrates them to form a *saliency* map capturing where visual attention is directed automatically. This saliency model not only accounts for oculomotor behavior [14], [15], but also neural activities controlling spatial attention and gaze directions [16].

The above algorithm has been extended to sounds [9], which characterizes spectral contrasts in addition to temporal and intensity contrasts incorporated already in the ISO loudness model [3]. The *spectral* saliency model has now been updated taking into account auditory specific spectral-temporal features [8].

However, the spectral saliency model still has a major limitation from the viewpoint of the global architecture of the auditory system consisting of the following two major pathways: *what* and *where* [17], [18]. The *what* pathway analyzes the spectral-temporal features of auditory objects. The *where* pathway, in contrast, localizes spatially auditory objects [19]. The concept of the what-pathway corresponds to the above spectral saliency model. In contrast, auditory spatial saliency has not been modeled yet based on auditory physiology.

Here, we propose a new framework that integrates auditory spatial saliency model and the ISO loudness model to explain perceptual noisiness. We first suggest a new model of auditory spatial saliency based on a sound localization algorithm implemented in the midbrain [20], [21]. It captures psychophysical phenomena for which the ISO loudness model fails to account [3]. Furthermore, the application of the spatial saliency model along with the ISO loudness model to driving sounds in passenger cars explained perceptual noisiness better than the ISO loudness model alone. These results suggest that the auditory spatial saliency model can upgrade the ISO loudness model, and contribute to improving environmental noise issues in everyday life.

## II. SPATIAL SALIENCY MODEL

We propose a new saliency model that quantifies the spatial features of sounds received by the left and right ears. The spatial saliency model is based on the functions of the superior colliculus, a midbrain structure crucial for directing spatial attention [22], [23]. It has been shown that the superior colliculus detects and localizes salient events in the visual space [24]. We believe that the mechanism of visual saliency in the superior colliculus can be extended easily to the auditory space because the superior colliculus has an auditory spatial map in accordance with its visual spatial map, and its neural circuits are shared between the auditory and visual maps [25].

In the following sections, we describe the three processes that consist of the spatial saliency model (Figs.1 and 2): (A) sound localization on a linear spatial map, (B) nonlinear transformation to the superior colliculus map, and (C) transformation from nonlinear localization map to spatial saliency map.

### A. SOUND LOCALIZATION ON A LINEAR SPATIAL MAP

The spatial localization in the auditory system is computationally intensive compared to that in the visual system. The spatial locations of visual objects are extracted directly by photoreceptors arranged orderly on the retina. In the auditory system, on the other hand, sound directions are not represented explicitly in the spatial arrangements of sensory receptors, but need to be reconstructed indirectly by the central nervous system [19].

The reconstruction of sound directions is carried out separately along the horizontal (azimuth) and vertical (elevation) axes in the central auditory system. We focused on horizontal localization because elevation localization usually requires spectral features whose frequencies are higher than those included in sounds recorded and analyzed mainly in this study ($<$2 kHz) [26].

To describe the algorithms of horizontal sound localization and spatial saliency, we use a driving sound of a car recorded at the front passenger seat (see chapter IV for detail

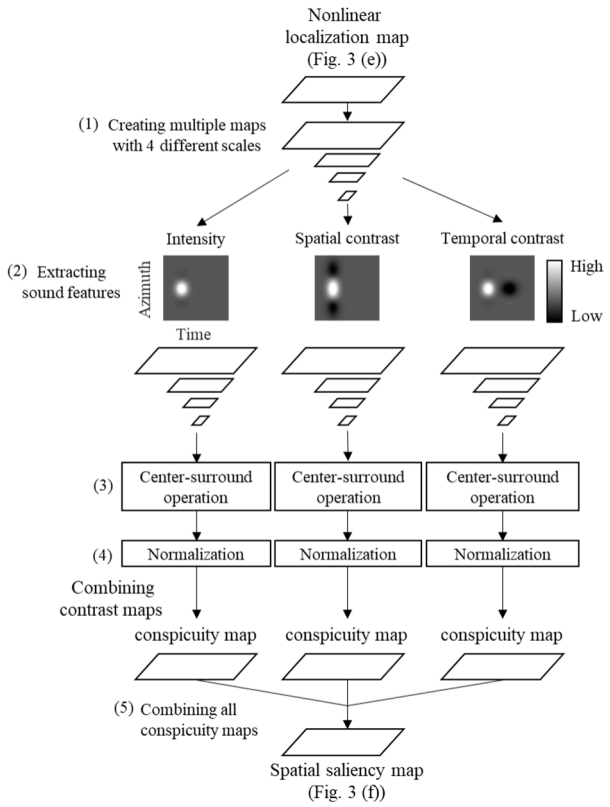



**FIGURE 2.** Schematic diagram of the spatial saliency model. The algorithm consists of the following five steps: (1) creating multiple maps with 4 different spatial/temporal scales (Gaussian pyramids), (2) extracting sound features (intensity, spatial contrast, and temporal contrast) from a sound localization map using Gabor filters, (3) calculating contrasts (center-surround operation) between the four maps with different scales, (4) normalizing the contrast maps and combining them across scales to obtain a conspicuity map in each sound feature, and (5) obtaining a saliency map by combining all conspicuity maps. See the main text for details.cy map.

of measurement conditions). Horizontal localization is based mainly on the following two cues that compare sounds arriving differently at the left and right ears: interaural time/level differences [19] (Fig. 3 (a), (b). These cues are detected mainly by the medial and lateral superior olive, respectively, in the brainstem [19], [27], [28]. We adopted a sound localization algorithm containing processes corresponding to these structures [21], [29].

The above algorithm calculates sound directions separately for each frequency channel with equivalent rectangular bandwidth decomposed by a cochlear model [30]. That is, it localizes multiple sound objects with different frequencies at the same time (e.g., male and female vocalization) [21]. However, we collapsed the output of all frequency channels and created a unified auditory spatial map to focus on the spatial features of sounds for our spatial saliency model (Fig. 3 (c)). This simplification is supported indirectly by the fact that signal frequencies in sounds and light (i.e., color) are collapsed in the superior colliculus at the level of population neurons [16], [31].

A majority of algorithms for horizontal sound localization, including the one adopted in this study, have a common limitation called a cone of confusion; they cannot determine whether sounds are derived from front or back based only on the interaural time/level differences. Two cues have been suggested to overcome this limitation: the head related transfer function and head movements [32], [33]. However, neither of these cues cannot be used in this study because binaural sounds were delivered by headphones. To detour this problem, we flipped sound directions front and back randomly (i.e., $\theta$ or $360 - \theta$) and repeated this procedure 1000 times (bootstrap method in Fig. 3 (d)). That is, a thousand localization maps were created for each sound and transformed into spatial saliency maps. A temporal window for horizontal sound localization was set to 5 ms [34], and shifted by 1 ms to capture the spatial-temporal dynamics of sounds.

### B. NONLINEAR TRANSFORMATION TO THE SUPERIOR COLLICULUS MAP

The above localization algorithm estimates sound directions along the azimuth linearly. However, spatial representation on the superior colliculus map is not linear [35]; spatial resolution is the highest at the fovea (central vision) while it degrades gradually in the periphery. This nonlinear transformation on the superior colliculus map has been modeled by the following formula [35]:

$$Azimuth_{Nonlinear} = P\ ln\left(\frac{1}{Q}\sqrt{Azimuth_{Linear}^2 + 2Q \times Azimuth_{Linear}}\right)$$

where $Azimuth_{Linear}$ and $Azimuth_{Nonlinear}$ correspond to sound directions before and after the nonlinear transformation, respectively, and two constants (P, Q) are equal to (1.4 mm, 3 deg) to map the linear external space on the anatomical superior colliculus surface. This transformation is modeled based on vision, but consistent with the following auditory findings; the spatial resolution of sound localization is the highest at the frontal midline [36], [37].

Fig. 3 (e) shows a nonlinear localization map obtained by the nonlinear transformation from Fig. 3 (d). The azimuth axis is represented as being wider on the front side and narrower on the rear side, reflecting the difference in spatial resolution (x-axis in Fig. 3 (e)).

### C. TRANSFORMATION FROM NONLINEAR LOCALIZATION MAP TO SPATIAL SALIENCY MAP

A variety of saliency models have been suggested in vision [12], [38] as well as spectral features in sounds [8], [9]. Here, we adopted the basic algorithm of the original saliency model [9], [12] and applied it to the spatial-temporal dynamics of auditory signals on the nonlinear localization map derived from the above procedures. Briefly, the spatial saliency algorithm for auditory signals consists of the following five steps (Fig. 2): (1) creating multiple maps with 4 different spatial/temporal

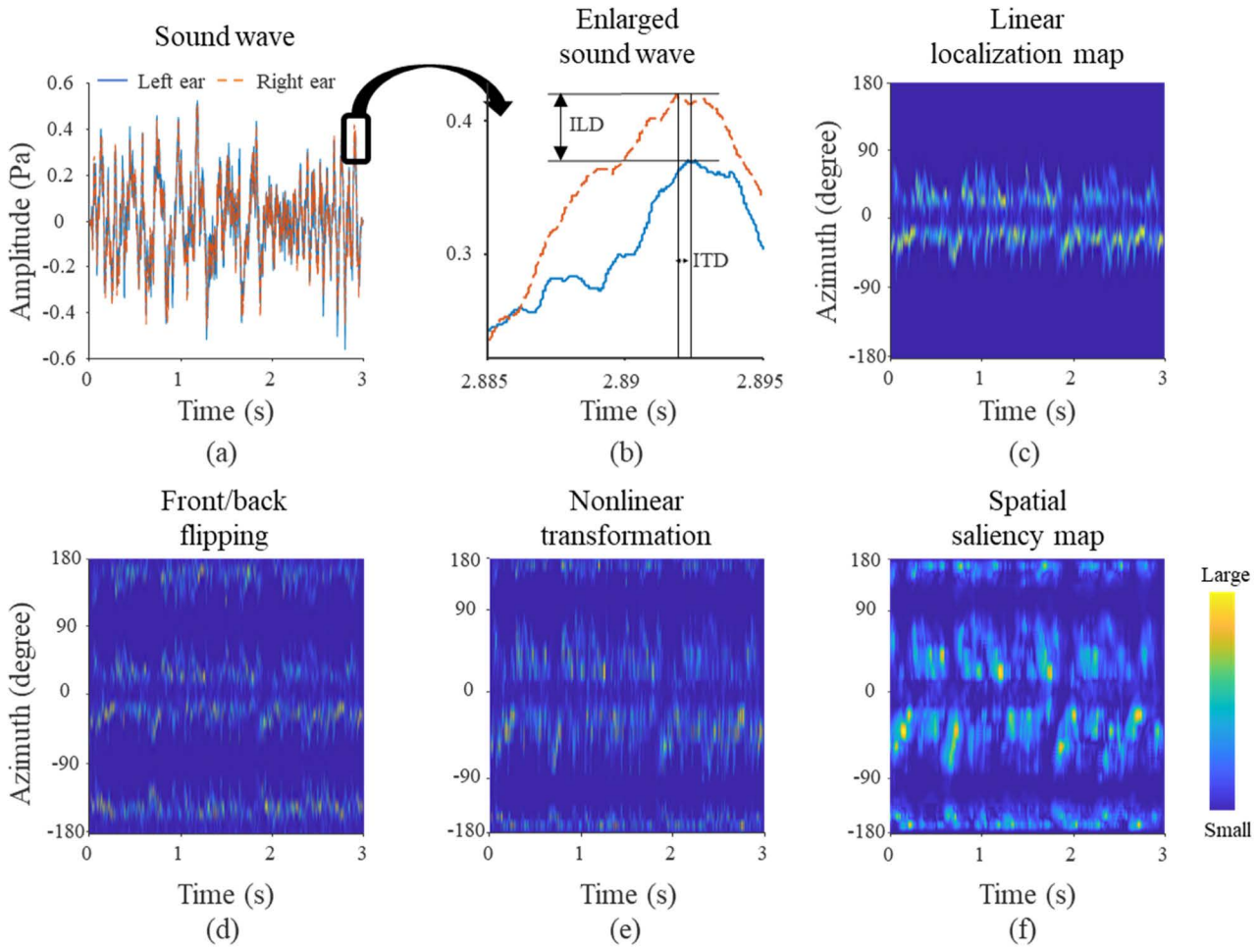



**FIGURE 3.** Input and output images of the spatial saliency model. (a) Input sound wave (binaural sound). The x/y-axes indicate time and amplitude, respectively. (b) Enlarged view of (a). The ITD/ILD are shown conceptually in the panel. (c) Linear localization map by the localization model based on ITD/ILD [21]. (d) Front/back random flipping was repeated 1000 times to take into account a cone of confusion. (e) Nonlinear localization map on the superior colliculus transformed from (d). (f) Spatial saliency map is calculated based on the spatial-temporal contrasts of (e). The x/y-axes in the panels (c)-(f) indicate time and azimuth, respectively. The maps (c)-(f) are obtained by analyzing a car driving sound used in Chapter IV.

scales (Gaussian pyramids), (2) extracting sound features (intensity, spatial contrast, and temporal contrast) from a nonlinear localization map on the superior colliculus using Gabor filters [spatial width (at the top layer of the Gaussian pyramids in the following second step): 0.16 mm (corresponding to 3 degrees in the central visual field); temporal width: 35 ms; all results reported in this study were similar using Gabor filters whose spatial-temporal width was half or twice the above values], (3) calculating contrasts (center-surround operation) between the four maps with different scales, except for the combination of the highest and lowest scales, resulting in 5 contrast maps, (4) normalizing the contrast maps and combining them across scales to obtain a *conspicuity map* in each sound feature, and (5) obtaining a spatial saliency map by combining all conspicuity maps linearly.

We calculated a thousand nonlinear localization maps from each sound by a boot-strap method to take into account front/back confusion as mentioned above (see *A. SOUND LOCALIZATION ON A LINEAR SPATIAL MAP*).

We applied the above spatial saliency algorithm to a spatial map derived from each boot-strap iteration and calculated its summed saliency value. Then, we obtained a thousand summed saliency values from each sound. Their grand average was regarded as a representative saliency value of the corresponding sound (hereafter, spatial saliency).

As with the spatial saliency, we used the time-averaged short term loudness calculated by the ISO loudness model (hereafter, ISO loudness). This is valid for the following two reasons. First, it has been shown that short term loudness is correlated with EEG signals originating from the auditory cortex [7]. Second, we confirmed that short term loudness was also correlated significantly with long term loudness calculated from the same data and known to explain subjective loudness judgements [6] (Pearson's $r > 0.99$, $p < 3.6 \times 10^{-33}$).

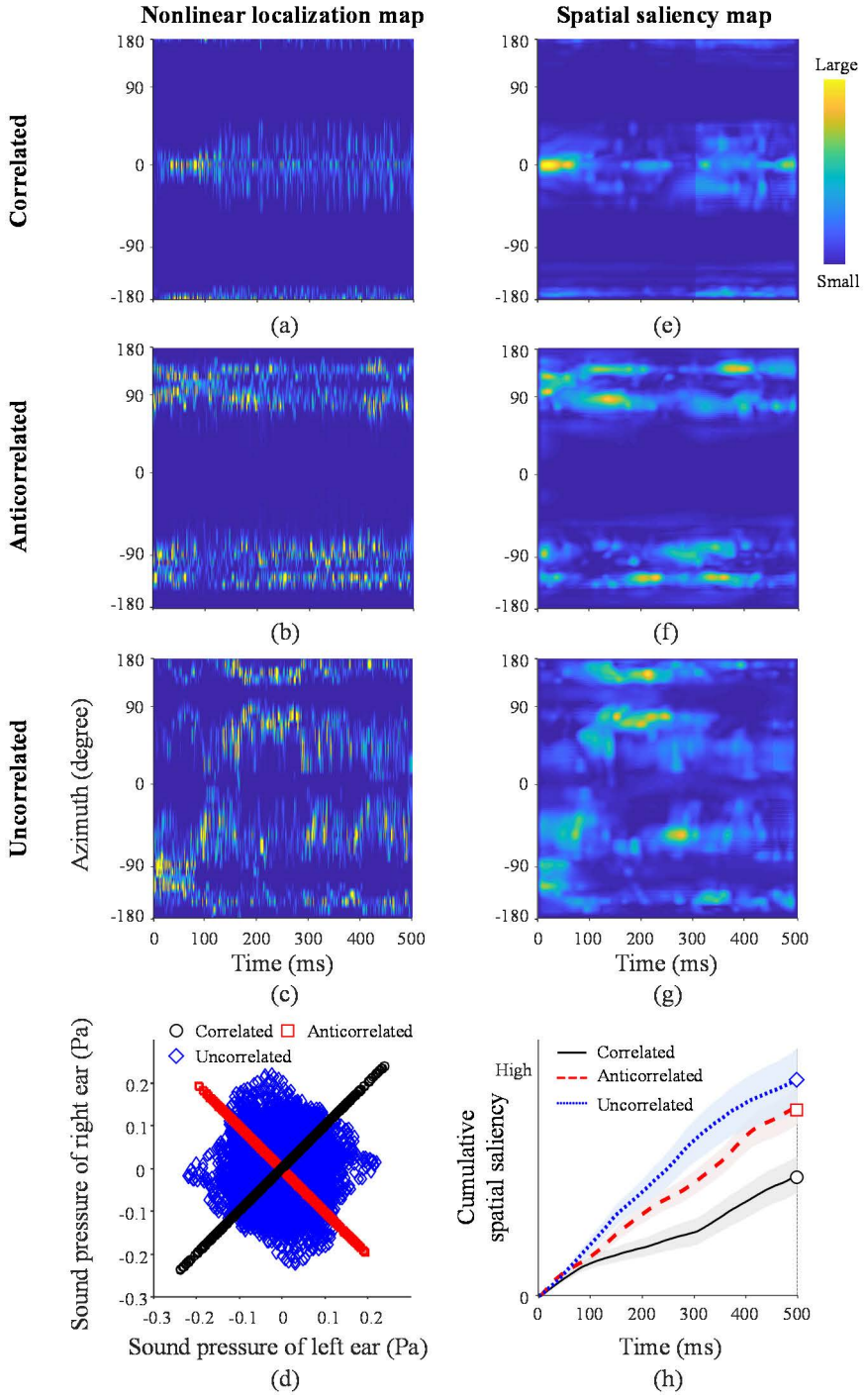



**FIGURE 4.** Examples of nonlinear localization/saliency maps for sounds with different interaural correlations. Left column (a)-(c): Nonlinear localization maps. Right column (e)-(g): Spatial saliency maps. Top raw (a), (e): Correlated sound. Middle raw (b), (f): Anticorrelated sound. Bottom raw (c), (g): Uncorrelated sounds. (d) Scatter plots between sound pressure of left and right ears with different interaural correlation. (h) Cumulative distributions of spatial saliencies for the three sounds with different interaural correlation. Thick lines and corresponding pale bands are averages and 95% intervals, respectively, calculated after the bootstrap process (Fig. 1). The markers at 500 ms correspond to the values of spatial saliency used in the following analyses. Intuitively, correlated sounds should be localized exactly at the midline, but they had a limited distribution (a) because of an anatomical delay of contralateral input relative to the corresponding ipsilateral input, implemented in the localization model [21].

## III. EFFECTS OF SPATIAL SALIENCY ON PERCEPTUAL NOISINESS

It has been shown previously that auditory perceptual noisiness is influenced by the spatial features of sounds [39]–[42]. The current ISO loudness model has been extended to take into account binaural input, but cannot fully explain the spatial features of perceptual noisiness because it does not calculate spatial cues explicitly (interaural time/level differences) [4], [5]. We therefore examined whether our auditory spatial saliency model could complement the shortage of the current ISO loudness model. Specifically, we hypothesized that sounds with higher spatial saliency are experienced louder than those with lower spatial saliency.

We focused on behavioral phenomena reported previously by Edmonds and Culling [39] where variation in sound directions affects perceptual noisiness. Using binaural sounds with different interaural phase correlation, the following two phenomena have been reported; (1) perceptual noisiness depends on interaural correlation in the following order: correlated (the same phase) < anticorrelated (the opposite phase) < uncorrelated (independent phases)(Fig. 4 (d)), and (2) the dependence of perceptual noisiness on interaural correlation is more significant in sounds with stronger low-frequency intensity (below 1.5 kHz) [39]. An index of interaural cross-correlation has been proposed previously to account for binaural perceptual noisiness [43]. However, it cannot explain the above behavioral phenomena simply because it predicts identical perceptual noisiness for anticorrelated and correlated sounds. Here, we describe how the auditory spatial saliency model has overcome the limitation of interaural cross-correlation and could explain the above behavioral phenomena reported in the original study.

### A. DEPENDENCE OF PERCEPTUAL NOISINESS ON INTERAURAL CORRELATIONS

Fig. 4 shows the nonlinear localization maps and the corresponding spatial saliency maps of correlated, anticorrelated and uncorrelated binaural sounds with identical monoaural spectral features (central frequency: 1 kHz; band range: 937 ∼ 1065 Hz; note that artificial sounds used in the original report [39] were also adopted in this chapter instead of natural driving sounds shown in Fig. 3). The correlated sound (Fig. 4 (a)) was localized at the center because of the limited range of interaural time differences (see figure legend for the distribution around the midline). The anticorrelated and uncorrelated sounds had wider distributions than the correlated sound, but there were several important differences between them (see below).

The anticorrelated sound had a characteristic spatial feature that localized sounds were distributed widely except for the center (Fig. 4 (b)). This phenomenon is explained easily by the localization algorithm based on interaural time differences. The anticorrelated sound had a fixed interaural phase difference (180 deg) regardless of frequency. This means equally that each frequency signal had a unique absolute value of an interaural time difference. In other words, each frequency signal was localized at a unique absolute direction in anticorrelated sounds. Interaural time differences were longer for low-frequency signals compared to high-frequency signals in anticorrelated sounds. Accordingly, low-frequency signals were plotted peripherally while high-frequency sounds were localized centrally. The lack of central directions in Fig. 4 (b) is explained by the fact that the highest frequency included in the anticorrelated sound used in this example was limited to 1065 Hz.

The uncorrelated sound had a distribution that was more uniform compared to the anticorrelated sound (Fig. 4 (c)). This is because each frequency signal changed its interaural phase randomly, which resulted in localized directions distributed within the range determined by the maximum possible interaural time difference of each frequency.

Based on the distributions of localized directions in correlated, anticorrelated and uncorrelated sounds, it was expected intuitively that their spatial saliencies would be in the following order: correlated < anticorrelated < uncorrelated. This prediction was confirmed quantitatively for sounds used in this example (Fig. 4 (h)). We further examined this prediction in other sounds with a variety of spectral features used in the original study by the following linear regression model (Table 1; see also Fig. 5 (a)):

$$\textit{Spatial saliency} \sim \textit{Correlated} + \textit{Anticorrelated} + \textit{Uncorrelated} \quad (1)$$

where *Spatial saliency* was normalized by z-score transformation, *Anticorrelated* is a constant, and *Correlated* and *Uncorrelated* are binary dummy variables. As shown in Table 1, *Correlated* had a significant negative regression coefficient, indicating that correlated sounds had lower spatial saliency than anticorrelated sounds. Similarly, *Uncorrelated* had a significant positive regression coefficient, indicating that uncorrelated sounds had higher spatial saliency than anticorrelated sounds. Accordingly, the result was consistent with the predicted order of spatial saliency (correlated < anticorrelated < uncorrelated). More importantly,

**TABLE 1. Effect of interaural correlations on spatial saliency.**

| *Variables* | *Coeff.* | *S.E.* | *T* | *D.O.F* | *P* |
|---|---|---|---|---|---|
| *Correlated* | -1.45 | 0.08 | -18.64 | 237 | $2.36 \times 10^{-48}$ |
| *Anticorrelated (Constant)* | 0.28 | 0.05 | 5.02 | 237 | $1.01 \times 10^{-6}$ |
| *Uncorrelated* | 0.62 | 0.08 | 7.99 | 237 | $5.87 \times 10^{-14}$ |

See (1) for the corresponding regression model. Anticorrelated corresponds to the constant (= intercept), and Correlated and Uncorrelated are binary dummy variables. We designed this regression model as simply as possible to test if the predicted order (correlated < anticorrelated < uncorrelated) is correct (i.e., without interactions).Coeff. : regression coefficient, S.E. : standard error, T : t value, D.O.F : degree of freedom, P : p value.

**TABLE 2. Effect of spatial saliency on perceptual noisiness for sounds with different interaural correlations.**

| *Variables* | *Coeff.* | *S.E.* | *T* | *D.O.F* | *P* | $\sigma_{Spectrum}$ | $\sigma_{Condition}$ |
|---|---|---|---|---|---|---|---|
| *Constant* | 70.13 | 0.41 | 172.05 | 717 | $< 1.00 \times 10^{-324}$ | 0.23 | 0.69 |
| *Spatial saliency* | 0.64 | 0.08 | 7.98 | 717 | $5.86 \times 10^{-15}$ | 0.21 | 0.04 |
| *ISO loudness* | -0.12 | 0.09 | -1.34 | 717 | 0.18 | 0.12 | 0.13 |

See (2) for the corresponding mixed effect model. Spatial saliency and ISO loudness were normalized by z-score transformation. Constant corresponds to the intercept. $\sigma_{Spectrum}$ and $\sigma_{Condition}$ are the standard deviation of random effects for groups with different spectral features and experimental conditions, respectively [39]. In addition to the three levels of interaural correlations, the original study also examined monaural sounds not included in the above analysis. The regression coefficient of ISO loudness became significant by including the monaural sounds [regression coefficient ± standard error = 8.77 ± 1.15, t(1277) = 7.63, p < 4.55×10-14], while that of spatial saliency remained significant [0.67 ± 0.09, t(1277) = 7.13, p < 1.69×10-12]. See Table 1 for abbreviations.

the order of spatial saliency agreed with the perceptual noisiness reported in the original study.

We further confirmed the relationship between perceptual noisiness and spatial saliency by the following mixed effect model:

$$\begin{aligned} &Perceptual\ noisiness \\ &\sim Features + (Features|Spectrum) \\ &+ (Features|Condition) \\ &Features = Spatial\ saliency + ISO\ loudness \end{aligned} \quad (2)$$

where *Perceptual noisiness* corresponds to subjective evaluation of noisiness for band noise sounds with three levels of interaural correlation (correlated, anticorrelated, and uncorrelated) reported originally. The parentheses in the second and third terms of the regression formula are random effects for experimental conditions used in the original study. Specifically, the first random effect of *Spectrum* in (2) indicates groups with different spectral features (i.e., differences in central frequency and bandwidth); each group contains sounds with three levels of interaural correlation whose spectral features are identical. The second random effect of *Condition* indicates experimental conditions used in the original study; the baseline of perceptual noisiness was set to either correlated, anticorrelated, or uncorrelated sounds (see [39] for details). This is not critical, but included here to account for potential variation across conditions. *Spatial saliency* and *ISO loudness* included in *Features* were normalized by z-score transformation.

The above mixed effect model showed that spatial saliency contributed significantly to perceptual noisiness (Table 2). In contrast, ISO loudness had no effect on the perceptual noisiness of binaural sounds analyzed here [see the legend of Table 2 for the contribution of ISO loudness to monoaural/binaural sounds [4]].

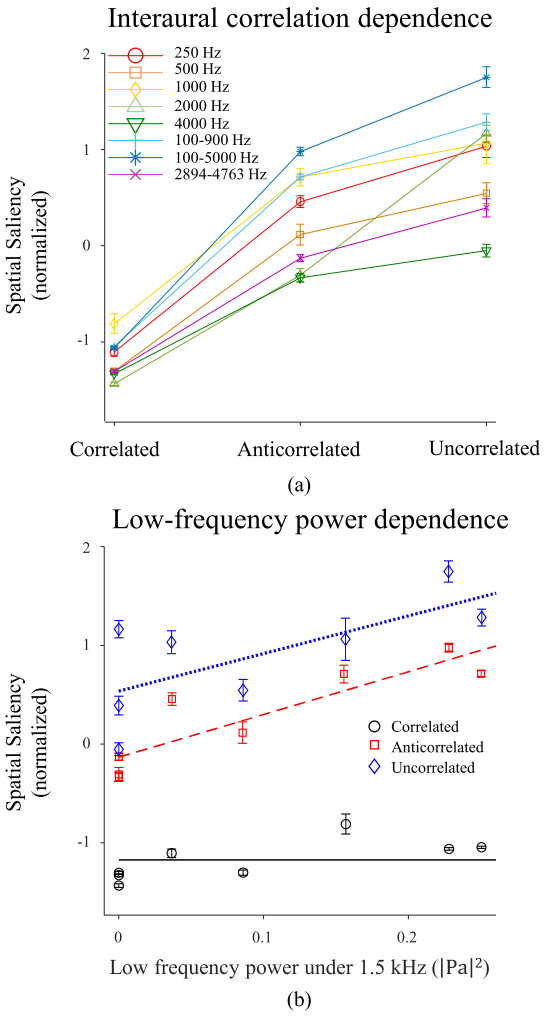



**FIGURE 5. Relationships between interaural correlations and spatial saliencies. (a) Effects of interaural correlations on spatial saliencies from sounds with a variety of spectral contents. (b) Effects of low-frequency power under 1.5 kHz on spatial saliencies with different interaural correlations. The results in the panels (a) and (b) were derived from the same data. The duration and sound pressure level of each test sound were fixed at 500 ms and 70 dB, respectively. Ten test sounds were generated for each category (e.g., correlated-250 Hz). Their spatial saliencies were normalized by z-score transformation. Each data point and its error bar correspond to the average and standard error of normalized spatial saliency for each category. The three regression lines in (b) (solid: correlated; dash: anticorrelated; dotted: uncorrelated) were derived from the same model shown in Table 3.**

These results support our hypothesis that auditory spatial saliency reflects some independent aspects of perceptual noisiness that could not be captured by ISO loudness.

### B. STRONGER IMPACT OF LOW FREQUENCY INTERAURAL CORRELATION ON SPATIAL SALIENCY

In addition to the dependence of perceptual noisiness on interaural correlation, the original report has also shown that the behavioral phenomenon was observed more significantly in sounds with stronger low-frequency power (below 1.5 kHz) [39]. It is intuitive to predict that the behavioral observation could be mediated by auditory spatial

**TABLE 3.** Effects of low frequency power on spatial saliency for sounds with different interaural correlations.

| *Variables* | *Coeff.* | *S.E.* | *T* | *D.O.F* | *P* |
|---|---|---|---|---|---|
| *Correlated (Constant)* | -1.17 | 0.04 | -28.44 | 235 | $4.94 \times 10^{-78}$ |
| *Anticorrelated* | 1.04 | 0.07 | 14.69 | 235 | $4.23 \times 10^{-35}$ |
| *Uncorrelated* | 1.71 | 0.07 | 24.16 | 235 | $1.25 \times 10^{-65}$ |
| $Power_{Low}$ *:Anticorrelated* | 4.34 | 0.42 | 10.24 | 235 | $1.41 \times 10^{-20}$ |
| $Power_{Low}$ *:Uncorrelated* | 3.81 | 0.42 | 9.02 | 235 | $6.95 \times 10^{-17}$ |

See (3) for the corresponding regression model. Correlated corresponds to the constant (= intercept), Anticorrelated and Uncorrelated are binary dummy variables indicating anticorrelated and uncorrelated sounds, and $Power_{Low}$:Anticorrelated and $Power_{Low}$:Uncorrelated are interactions between the power of low-frequency signals (< 1.5 kHz) and anticorrelated and uncorrelated sounds respectively, meaning specific contribution of low-frequency signals in anticorrelated and uncorrelated sounds to spatial saliency. See Table 1 for abbreviations.

saliency because of its dependence on the spectral features of binaural sounds as described above (Fig. 5 (a)). We tested this prediction by the following regression model:

$$\begin{aligned} &Spatial\ saliency \\ &\sim Correlated + Anticorrelated + Uncorrelated \\ &+ (Anticorrelated + Uncorrelated) : Power_{Low} \end{aligned} \quad (3)$$

where *Correlated* corresponds to a constant (i.e., baseline spatial saliency), *Anticorrelated* and *Uncorrelated* are binary dummy variables indicating anticorrelated and uncorrelated sounds, and $Power_{Low}$ is power obtained by integration of A-weighted power spectrum below 1.5 kHz according to the original report [39]. The interactions between *Anti/Uncorrelated* and $Power_{Low}$ should capture the enhanced contribution of low-frequency signals to spatial saliency. An interaction between $Power_{Low}$ and *Correlated* was not included because correlated sounds should be localized around the midline regardless of the power of low frequency signals. *Spatial saliency* was normalized by z-score transformation. The result was consistent with our prediction; the spatial saliency of uncorrelated sounds was higher for sounds with stronger low-frequency power (regression coefficients of PowerLow:Uncorrelted in Table 3, Fig. 5 (b)). A similar result was also observed in anticorrelated sounds (regression coefficients of PowerLow:Anticorrelted in Table 3).

These results indicate that behavioral phenomena reported in the original study [39] could be explained, at least in part, by auditory spatial saliency.

## IV. APPLICATIONS OF SPATIAL SALIENCY MODEL TO DRIVING SOUNDS

The results described in the section III are consistent with our hypothesis that auditory spatial attention, captured possibly by the algorithm of spatial saliency, could account for perceptual noisiness, at least under limited experimental conditions. Here, we tested the hypothesis further in situations people encounter in everyday life. Single category sounds, driving sounds in passenger cars, were selected as our first approach. Participants were asked to judge the noisiness of each driving sound. We then examined whether spatial saliency could complement the current ISO loudness model to account better for perceptual noisiness.

**TABLE 4.** Six cars used for driving sound recordings.

| *Cars* | *Sizes* | *Engine displacements* | *Styles* |
|---|---|---|---|
| Car 1 | Compact | 660 cc | Wagon |
| Car 2 | Compact | 1300 cc | Wagon |
| Car 3 | Middle | 1500 cc | Wagon |
| Car 4 | Middle | 1600 cc | Sedan |
| Car 5 | Middle | 2000 cc | Wagon |
| Car 6 | Large | 3000 cc | Sedan |

### A. METHODS

#### 1) PARTICIPANTS

Fourteen adults with driver's licenses (3 women; age ranged from 27 to 55 years old: average $\pm$ s.d. $= 37 \pm 9$ years old) participated in this experiment. They were informed of the nature of the study and written consented to be part of the study approved by the ethics committee of Mazda Motor Corporation.

#### 2) DRIVING SOUNDS

To collect a variety of driving sounds with a wide range of spatial saliency as well as ISO loudness, we adopted six cars with different sizes (compact, middle, and large), engine displacements (660 to 3,000 cc), styles (sedan and wagon), and brands (see Table 4). To further enhance the variety of driving sounds, the cars were driven on coarse and smooth road surfaces at the Mazda Miyoshi Proving Ground. The driving speeds in all condition were constant approximately at 100 km/h. All driving sounds were recorded using a head and torso simulator (HMS III.0, HEAD acoustics, Herzogenlart, Germany; 48 kHz and 24 bits binaural sampling). The simulator was placed on the front passenger seat (Fig. 6 (a)).

Test sounds (3 seconds) used for the judgements of perceptual noisiness were extracted from the recorded driving sounds to include mainly sounds caused by interactions between tires and road surfaces. Sounds caused by other factors, such as stone-pitching, were excluded as much as possible. Because of the limited durations of the driving

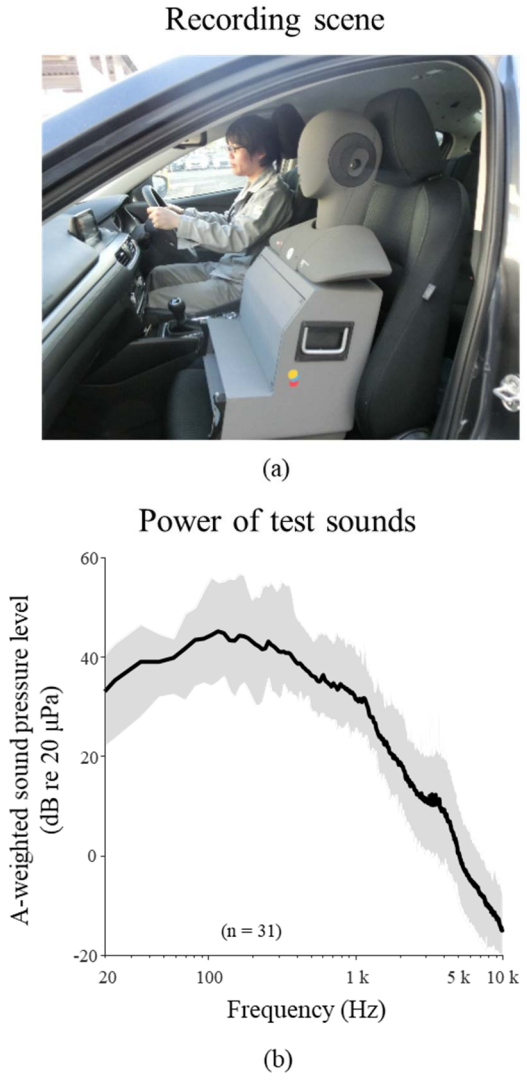



**FIGURE 6.** Recording of driving sounds. (a) Binaural recording by a dummy head and torso simulator placed on the passenger seat. (b) Sound pressure levels of 31 test sounds used in our experiment. Thick line: average, Gray band: 95 % confidence interval.

sounds (10 seconds for each recorded audio file), the temporal periods of test sounds derived from the same driving sounds were overlapped partially with each other for 1 s at most. This procedure resulted in 31 test sounds in total [1 to 4 sounds (2.6 sounds on average) from each car under each road surface condition]. The edges of extracted test sounds were processed by a Hanning window (50 ms window length).

The sound pressure levels of test sounds were different depending on the car and road surface conditions. However, to focus on our hypothesis that sound features other than sound pressure levels contribute to perceptual noisiness, we adjusted the average sound pressure levels of all test sounds within a limited range around the grand average of the original test sounds (88 ± 0.7 dB; see Fig. 6 (b) for the power spectra of the test sounds).

The test sounds were filtered using a finite impulse response (FIR) low-pass filter [function ''lowpass'' with 0.85 steepness and 60 dB stopband attenuation in MATLAB (R2018b, The MathWorks, United States)] with the cutoff frequency of 5 kHz before data analyses because their sound pressure levels above 5 kHz were less than the minimum audible pressure curve [44] (see also Fig. 6 (b)).

In addition to spatial saliency, we also tested whether spectral saliency affects perceptual noisiness [9]. To calculate spectral saliency, we used a high-pass filter with a cutoff frequency of 20 Hz [function ''highpass'' with 0.85 steepness and 60 dB stopband attenuation] to limit its calculation within the audible frequency range [similar results were confirmed without this filtering (data not shown)].

### 3) EXPERIMENTAL SYSTEM

Test sounds were played by a pair of open headphones (HD IV.1, HEAD acoustics, Herzogenlart, Germany, frequency ranges: 12 Hz to 40.5 kHz) through a digital equalizer (PEQ V, HEAD acoustics, Herzogenlart, Germany) in a car parked in an anechoic chamber. A PC monitor was placed on the hood of the car in front of the driver's seat where participants sat to present task instructions visually (see *Behavioral task*). Noisiness judgments were reported by left/right shift keys in a keyboard placed on the lap of participants. This system was controlled by Psychtoolbox-3 (v 3.0.14) [45] in MATLAB.

### 4) BEHAVIORAL PARADIGM

A two-interval (alternative) forced choice task was used to evaluate the relative noisiness of two test sounds played sequentially (Fig. 7 (a)). After displaying a cross mark for 1 second at the center of the screen, the first sound was played for 3 seconds with the word ''$1^{st}$'' presented at the center of the screen. Subsequently, the second sound was played after an inter-stimulus interval of 1 second with the word ''$2^{nd}$'' on the screen. Participants judged which test sound was noisier by pressing the left or right shift key on the keyboard. Mappings between left/right keys and nosiness judgments (e.g., $1^{st}$ stimulus was noisier than $2^{nd}$ stimulus) were counterbalanced across participants. A selected mapping rule for each subject was presented on the screen as a reminder on each trial (e.g., ''$1^{st}$ left side'' and ''$2^{nd}$ right side'').

The first and second sounds were selected from categories with different combinations of cars and road surface conditions [e.g. (car A, coarse) vs. (car B, smooth)]. One of the test sounds in a category, containing 1 to 4 test sounds, was chosen randomly for this comparison.

All participants performed 132 trials in total, corresponding to the number of permutations of two test sounds selected from all 12 categories (6 cars × 2 surfaces). Trials were divided into two blocks, each of which contained approximately a half of total trials. Participants had a short break (about 5 minutes) between the blocks of trials. The above procedure was preceded by a practice block of 20 trials.

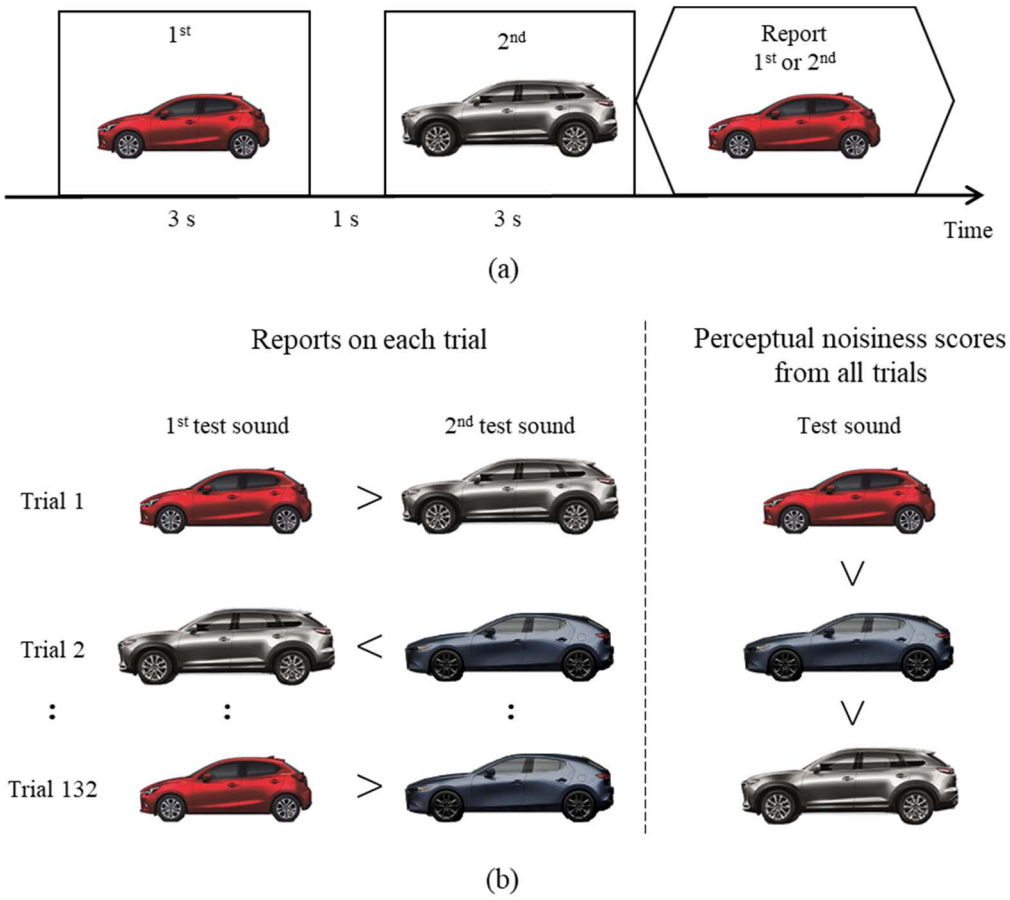



**FIGURE 7.** Behavioral paradigm and analysis. (a) Behavioral paradigm. On each trial, participants listened to two consecutive test sounds (1st and 2nd; 3 seconds each) and reported which sound was noisier (1st or 2nd). An inter-stimulus interval was 1 second. (b) Behavioral analysis. Left: Examples of reports on individual trials (132 in total). Right: Perceptual noisiness scores estimated from the whole reports of a participant using the Bradley-Terry model (see *METHODS* in chapter IV).

White noises with 2 different sound pressure levels (80.7, 74.7dB) were used as training stimuli.

### 5) BRADLEY-TERRY MODEL FOR PERCEPTUAL NOISINESS

We estimated the perceptual noisiness of individual test sounds based on noisiness judgements during the above two-interval forced choice task using a Bradley-Terry model [46]. The model estimates the strength of each test sound based on their wins (i.e., selected as noisier) and losses (i.e., not selected) during the task (Fig. 7 (b)).

The perceptual noisiness of individual test sound was modeled as follow:

$$\begin{aligned} \pi_i^s &= exp\left(\beta_i + \gamma_i^s\right) \\ \beta_i &\sim Normal\,(0, 1) \\ \gamma_i^s &\sim Normal\left(0, \sigma^s\right) \\ \sigma^s &\sim Cauchy\,(0, \sigma) \\ \sigma &\sim Cauchy\,(0, 1) \end{aligned}$$

where $\pi_i^s$ corresponds to the estimated noisiness of $i^{\text{th}}$ test sound for $s^{\text{th}}$ subject, $\beta_i$ is a fixed effect of $i^{\text{th}}$ test sound across all subjects, and $\gamma_i^s$ is a random effect of $i^{\text{th}}$ test sound for $s^{\text{th}}$ subject to account for individual differences across subjects. The random effects of $s^{\text{th}}$ subject for each stimulus $(\gamma_1^s, \gamma_2^s, \ldots, \gamma_{31}^s)$ form a normal distribution with its standard deviation of $\sigma^s$ specified for each subject. Individual differences in $\sigma^s$ are taken into account by a Cauchy distribution with its scale $\sigma$.

The probability of the 1st test sound ($i$) judged as noisier than the 2nd test sound ($j$) in the two-interval forced choice task was modeled by the following formula:

$$\begin{aligned} P_{(i>j)} &\sim \frac{\tau_{Decay}^s \cdot \pi_i^s}{\tau_{Decay}^s \cdot \pi_i^s + \pi_j^s} \\ \tau_{Decay}^s &= exp\left(\beta_{Decay} + \gamma_{Decay}^s\right) \\ \beta_{Decay} &\sim Normal\,(0, 1) \\ \gamma_{Decay}^s &\sim Normal\left(0, \sigma_{Decay}\right) \\ \sigma_{Decay} &\sim Cauchy\,(0, 1) \end{aligned}$$

where $\tau_{Decay}^s$ accounts for the possible temporal attenuation of the 1st test sound in memory when participants compared 1st and 2nd test stimuli at the end of the trial for each subject. $\beta_{Decay}$ is a constant, while its random effect of

subject, $\gamma^s_{Decay}$, corresponds to individual differences across subjects. $\gamma^s_{Decay}$ (s = 1, 2, …, 14) forms a normal distribution with its standard deviation of $\sigma_{Decay}$ derived from a Cauchy distribution whose scale is equal to 1. An exponential is taken for $\tau^s_{Decay}$ because it should be a positive value ($\tau^s_{Decay} = 0.84$ was obtained in this experiment, meaning that the estimated noisiness of the 1st stimulus (e.g., $\pi^s_i$) was attenuated approximately 16 % before it was compared to that of the 2nd stimulus (e.g., $\pi^s_j$) during the two-interval forced choice task).

#### 6) MODEL FITTINGS

We adopted a Bayesian approach to fit the above model to the binary judgments of relative perceptual noisiness [47]. The posterior distributions of all parameters were generated by Markov Chain Monte Carlo with 4 chains each of which sampled 10000 times (after the warm-up samplings of 1000 times). The model fitting was carried out by RStan (v2.17.3, Stan Development Team 2018) and R (v3.5.0, R Core Team 2018).

### B. RESULTS

As described in the introduction, we hypothesized that perceptual noisiness is explained by spatial saliency as well as ISO loudness [3], [4]. Along with spatial saliency, spectral saliency [9] might also affect perceptual noisiness. Here, we examined these hypotheses using driving sounds.

#### 1) SPATIAL SALIENCY VS. OTHER SOUND FEATURES

Fig. 8 shows two examples of test sounds whose spatial saliencies were the maximum and the minimum among the 31 test sounds used in this study. Their cumulative distributions of spatial saliencies, along with that of the test sound shown in Fig. 4, were different between each other, suggesting that the algorithm of spatial saliency could capture the variety of spatial features across the test sounds.

We first examined relationships between spatial saliency and the other sound features (ISO loudness and spectral saliency). We expected that spatial saliency should be independent from ISO loudness because it should capture features that cannot be accounted for by ISO loudness (see chapter III). We also expected that spatial and spectral saliencies would also be independent because they calculate contrasts along two different physical dimensions (i.e., horizontal direction and spectral frequency). These predictions were confirmed in our test sounds (Fig. 9 (a), (b)). That is, there were not significant correlations between spatial saliency and both ISO loudness and spectral saliency ($p > 0.3$).

At the same time, there was a significant correlation between ISO loudness and spectral saliencies (Fig. 9 (c)), presumably reflecting a common feature between them (i.e., temporal variation [6], [9]).

These results support our hypothesis that spatial saliency captures features not explained by ISO loudness and spectral saliency and could contribute independently to perceptual noisiness.

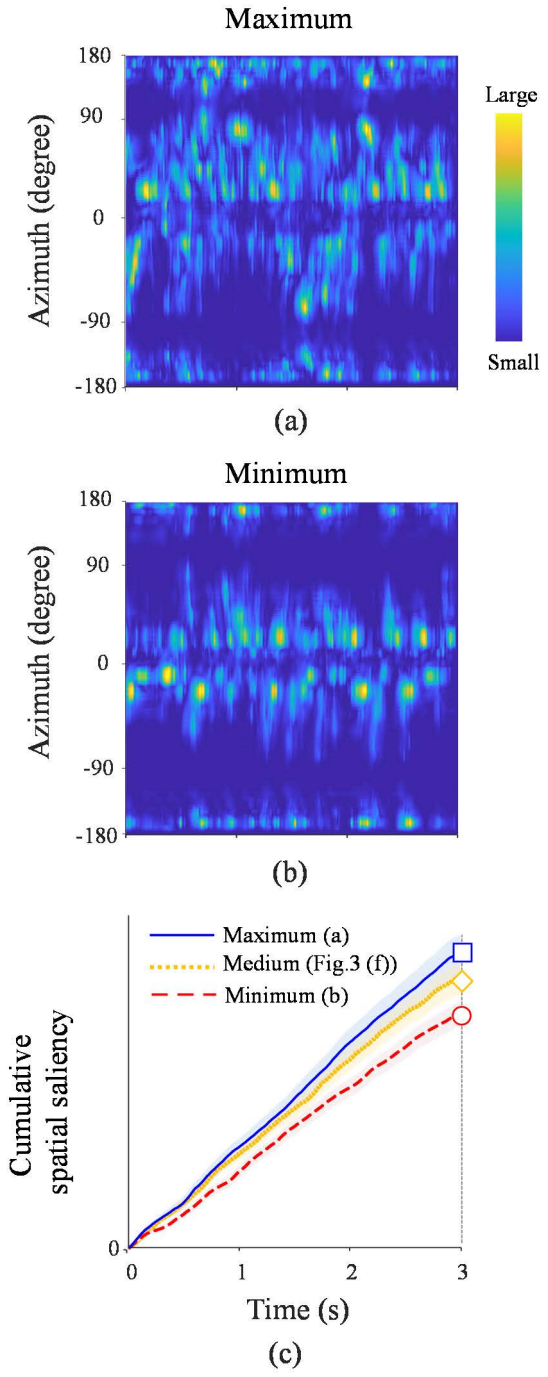



**FIGURE 8. Examples of spatial saliency maps for test sounds. (a) A spatial saliency map of a test sound with the maximum saliency level. (b) Another spatial saliency map of a different test sound with the minimum saliency level. (c) Cumulative distributions of spatial saliencies for test sounds in (a), (b) and Fig. 3 (f). See Fig. 4 for details.**

#### 2) SPATIAL SALIENCY VS. PERCEPTUAL NOISINESS

We then examined directly whether spatial saliency was indeed related to perceptual noisiness. The perceptual noisiness of test sound, $\pi_i$ ($=\beta^i$; constant term of $^i$th test

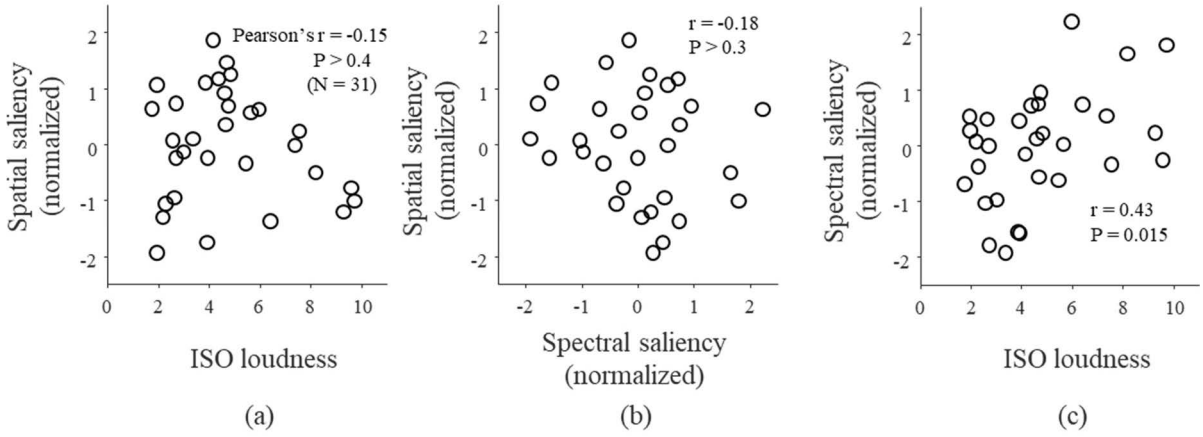



**FIGURE 9.** Relationships between three sound features. (a) Spatial saliency vs. ISO loudness. (b) Spatial saliency vs. spectral saliency. (c) Spectral saliency vs. ISO loudness.

sound regardless of the subject), was quantified using the Bradley-Terry model (see *METHODS* for details). We confirmed that the ISO loudness of test sounds was correlated significantly with their perceptual noisiness (Fig. 10 (a)). We also found that the spectral saliency of test sounds was correlated with their perceptual noisiness (Fig. 10 (b)). Counterintuitively, however, such correlation with perceptual noisiness was not observed in spatial saliency (Fig. 10 (c)).

The above results are apparently inconsistent with our hypothesis that spatial saliency contributes to perceptual noisiness. However, the inconsistency was resolved after removing the contributions of ISO loudness and spectral saliency from perceptual noisiness; spatial saliency was correlated with the residuals obtained from a multiple regression analysis in which perceptual noisiness was defined as a dependent variable, and ISO loudness and spectral saliency as independent variables (Fig. 10 (d)).

To quantify the relative contribution of spatial saliency, ISO loudness and spectral saliency to the perceptual noisiness, we fit the following multiple regression model to our data:

$$\textit{Perceptual noisiness} \sim \textit{ISO Loudness} + \textit{Spatial saliency} + \textit{Spectral saliency} \quad (4)$$

We first found that the regression coefficient of ISO loudness had a significant positive value (Table 5), confirming that ISO loudness affected perceptual noisiness. More importantly, the regression coefficient of spatial saliency also had a significant positive value, suggesting that spatial saliency contributed to perceptual noisiness independently from ISO loudness. The importance of ISO loudness and spatial saliency was also confirmed by likelihood ratio tests comparing the full model represented by (4) with reduced models eliminating either ISO loudness or spatial saliency (Table 5). We also found a similar trend in spectral saliency, although it did not reach statistical significance (Table 5).

The above results support our hypothesis that auditory spatial saliency complements ISO loudness to account for perceptual noisiness.

## V. DISCUSSION

We have proposed a new model of auditory spatial saliency (Figs. 1 and 2). The model complements the conventional ISO loudness by taking into account the new features of spatial saliency which explained perceptual noisiness better for driving sounds in passenger cars than the ISO loudness alone (Table 5). This strategy of integrating the multiple models for perceptual noisiness could be utilized not only for the acoustic design of engineering products, including passenger cars, but also for environmental medicine. In this section, we discuss the following three points: (1) how auditory saliency complements ISO loudness; (2) neurophysiological basis of auditory spatial saliency; and (3) potential applications of auditory spatial saliency to automotive engineering.

### A. AUDITORY SALIENCY COMPLEMENTS ISO LOUDNESS

The ISO loudness model was proposed originally by Zwicker [48], [49], and has been upgraded extensively by Moore and his colleagues [4] based on new findings from auditory psychophysics and periphery. The original loudness model was only for static monoaural sounds, but the current ISO loudness model has been extended to evaluate dynamic binaural sounds [4]–[6]. It has now become the world standard [3] and allows us to evaluate perceptual noisiness quantitatively.

However, the current ISO loudness model still has room for improvements at least in the following two points. First, it does not consider interaural time/level differences, although it explains how to integrate adaptively the levels of binaural sounds to match perceptual loudness. This limitation makes the model fail to account for the effects of interaural correlations on the loudness [4], [39]. We have overcome this limitation by introducing spatial saliency that captures acoustic contrasts in space (Figs. 4 and 5 (a)). Indeed,

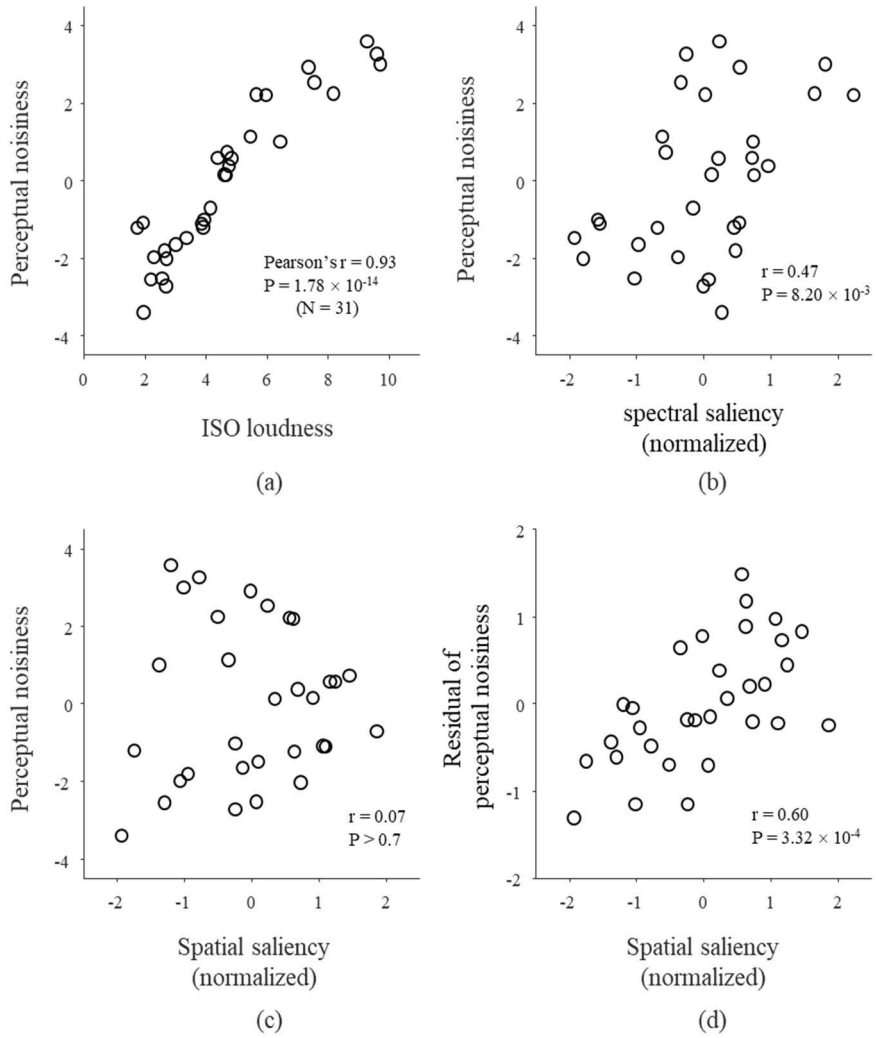



**FIGURE 10.** Relationships between perceptual noisiness and three sound features. (a) Perceptual noisiness vs. ISO loudness. (b) Perceptual noisiness vs. spectral saliency. (c) Perceptual noisiness vs. spatial saliency. (d) Spatial saliency vs. the residuals of perceptual noisiness after removing the influences of ISO loudness and spectral saliency by multiple regression. The y-axis in (d) was expanded compared to those in (a)-(c) to focus on the range of the residuals.

the integration of spatial saliency with ISO loudness better explained perceptual noisiness for driving sounds than the ISO loudness alone (Table 5). Moreover, spatial saliency also accounted for the perceptual phenomena of interaural correlations (Fig. 5) which the ISO loudness and the interaural cross-correlation [43] fail to account for.

Second, the ISO loudness model integrates the intensities of signals across frequency channels [i.e. equal rectangular bands] to calculate the overall level of loudness, but it does not take into account a variation across frequency channels. Although evidence is limited regarding the potential impact of the variation across frequency channels on the loudness, we found a trend that the spectral-temporal variation, quantified by spectral saliency, might possibly reflect some aspects of perceptual noisiness independently from ISO loudness, although it did not reach statistical significance (Table 5). We adopted the original spectral saliency model [9] only for the sake of simplicity despite the fact that it does not take into account the structure of the peripheral auditory system. This limitation could be overcome easily by simply incorporating a peripheral cochlear model [50] and/or adopting more sophisticated spectral saliency models reported recently [8].

### B. NEUROPHYSIOLOGICAL BASIS OF AUDITORY SPATIAL SALIENCY

The word "saliency" has been used previously to model auditory spatial attention in the field of robotics engineering [51], [52]. Their algorithms are designed mainly for localization (interaural correlation), but ignore spatial contrast, a key concept of our spatial saliency model. This limitation was not critical in the previous studies because their purpose was not to account for human perception/behavior. On the other hand, our purpose was to account for perceptual

**TABLE 5.** Multiple regression for perceptual noisiness in driving sounds.

| | *Regression* | | | | | *Likelihood ratio test* | | |
|---|---|---|---|---|---|---|---|---|
| *Variables* | *Coeff.* | *S.E.* | *T* | *D.O.F* | *P* | *D.O.F* | *ΔLL* | *P* |
| *Constant* | -3.74 | 0.25 | -15.23 | 27 | $8.91 \times 10^{-15}$ | *4* | - | - |
| *ISO loudness* | 0.79 | 0.05 | 16.60 | 27 | $1.08 \times 10^{-15}$ | *3* | 35.5 | $< 1.00 \times 10^{-324}$ |
| *Spatial saliency* | 0.44 | 0.10 | 4.34 | 27 | $1.81 \times 10^{-5}$ | *3* | 7.4 | $1.26 \times 10^{-4}$ |
| *Spectral saliency* | 0.21 | 0.11 | 1.90 | 27 | 0.07 | *3* | 1.7 | 0.07 |

See (4) for the corresponding full regression model. Likelihood ratio tests were carried out by comparing between the full model (4) and each reduced model excluding the corresponding variable in each row. ISO loudness, Spatial saliency, and Spectral saliency were normalized by z-score transformation. Constant corresponds to the intercept. See Table 1 for other abbreviations. $\Delta LL$: log likelihood difference between full and each reduced model.

noisiness by auditory attention. We therefore built our spatial saliency model based on a biologically plausible algorithm.

We adopted an algorithm of sound localization based on the functions of neural circuits in the brainstem that localize sounds along the horizontal axis (i.e., medial/lateral superior olive) [21]. Moreover, we also considered the nonlinear map of auditory space on which the spatial resolution is the highest at the frontal midline and degrades gradually as sound directions deviate from the midline [35]. Although it is unclear whether an auditory spatial map exists in the cerebral cortex, assuming such a spatial map allows us to adopt an algorithm used for visual saliency. This approach is supported physiologically by the structures and functions of the superior colliculus (see below).

We based our concept of spatial saliency on the structure and function of the superior colliculus in the midbrain, a major structure of eye movement control in the central nervous system [53], because of the following three points. First, the superior colliculus receives auditory signals from the inferior colliculus and integrates them with other sensory information, especially vision, on its retinotopic map to control spatial attention multimodally [25]. Second, auditory spatial attention depends on gaze directions, indicating its dependence on neural circuits with retinotopic coordinates, such as those in the superior colliculus, instead of head centered coordinates in which cues for horizontal sound localization (i.e., interaural time/level differences) are encoded [54]. Third, neural activity in the superior colliculus is correlated dynamically with spatial saliency, at least in vision [23]. We focused on the superior colliculus just for the sake of simplicity. However, it is highly likely that auditory spatial attention is controlled by neural circuits integrating the superior colliculus and the auditory *where* pathway from the core of the auditory cortex to the parietal cortex [18].

### C. POTENTIAL APPLICATIONS OF AUDITORY SPATIAL SALIENCY TO AUTOMOTIVE ENGINEERING

We believe that the current spatial saliency model as well as the ISO loudness model could be applied to real environmental and engineering issues. Here, we discuss how our model could guide our future research toward the model-based development of acoustic design for passenger cars based on the mechanisms of the human auditory system.

Driving sounds are generated by the following two mechanisms: air transmission and solid propagation of vibration. Sounds originated from frictions between road surfaces and tires are transmitted by the air through the floor panel. Sounds are also radiated from various parts of the body by solid propagation of vibration so that driving sounds could be heard from unexpected directions, such as above the head (i.e., caused by vibration of the roof).

Sounds generated by the above two mechanisms have different contributions to driving sounds depending on frequencies; those below 500 Hz are influenced more strongly by the solid propagation of vibration compared to air transmission, while the relationship becomes the opposite above 500 Hz [55]. Because spatial saliency is influenced more strongly by low frequency sounds compared to high frequency sounds (Fig. 5 (b); a similar result was obtained when the threshold was lowered from 1.5 kHz to 500 Hz), we speculate that the solid propagation of vibration has stronger impact on spatial saliency compared to air transmission.

The solid propagation of vibration depends on the architecture of the car body because it defines the transmission path of vibration. The architecture of the car body varies widely across cars. Accordingly, we expected that spatial saliency caused partly by the solid propagation of vibration should also vary across cars. We confirmed this expectation in our test sounds (Fig. 11; two way ANOVA with the main factors of Cars and road surface (coarse/smooth); the main factor of

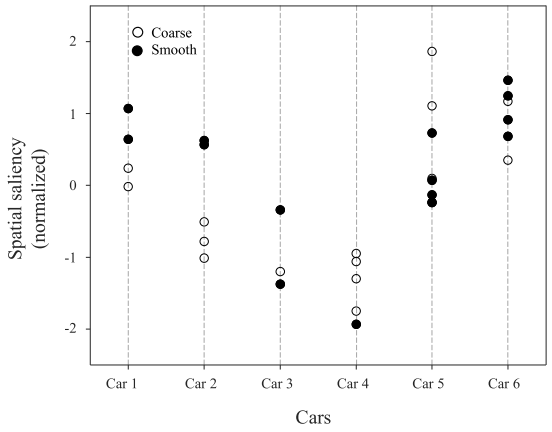



**FIGURE 11. Dependence of spatial saliency on cars and road surfaces (coarse and smooth). Each data point corresponds to each test sound (n = 31). See table 4 for car specifics.**

Cars: $F(5) = 12.94$, $p = 2.0 \times 10^{-5}$). We therefore speculate that spatial saliency could be reduced by controlling the solid propagation of vibration appropriately through the design of the architecture of the car body.

The solid propagation of vibration is influenced by both global and local body architectures. Asymmetric rigidities in the global body architecture create a path that propagates vibrations from tires to remote parts, such as the roof and doors. In contrast, rigidities in local body architecture create sound sources because of the resonance of local components. Accordingly, spatial saliency could be reduced by balancing spatially the effects of the global and local body architectures on the distribution of sound directions at the ears.

The above speculation still needs to be tested in individual cars experimentally. Such experiments are challenging, but, will help guide our research towards the model-based development of the car body architectures based on the human auditory system.

## VI. FUTURE DIRECTION

We have proposed a new model of auditory spatial saliency and integrated it to the conventional ISO loudness to account better for perceptual noisiness for binaural sounds. The spatial saliency model still needs to be extended by integrating other sound features (e.g., spectral features and head related transfer function for vertical localization) and to resolve remaining issues (e.g., cone of confusion) to capture fully the spatial features of perceptual noisiness. Nevertheless, it allows us to start creating a new technology of model-based development for acoustic space in automotive interiors based on human auditory attention.

Automotive engineering faces with a variety of tradeoff problems, including those between perceptual noisiness and other functions/values (e.g., adjusting suspension rigidity to attenuate solid propagation of vibrations vs. driving stability; adding noise insulators vs. their cost/weight to be added). Our new approach could possibly break through such tradeoff problems by replacing the current ''symptomatic treatments'' of noise reductions (e.g., suspensions and noise insulators) with the new ''causal treatments'' of designing the global/local body architectures based on auditory spatial attention. We also believe that our model could be applied to a variety of environmental noise issues other than automotive engineering, and contribute to improving our quality of life.

## ACKNOWLEDGMENT

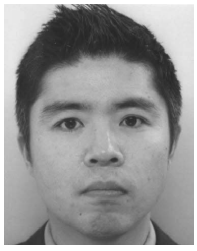
The authors would like to thank Prof. Takahide Nouzawa and Kentaro Ono at Hiroshima University for helpful discussions and comments on the manuscript. They also thank Kazuo Sakamoto and Masanori Honda at Mazda Motor Corporation for their valuable guidance.

## REFERENCES

[1] M. Florentine, A. N. Popper, and R. R. Fay, *Loudness—Springer Handbook of Auditory Research*. New York, USA: Springer, 2011.
[2] *Electroacoustics—Sound Level Meters—Part 1: Specifications*, Standard IEC 61672-1, 2013.
[3] *Acoustics—Methods for Calculating Loudness—Part 3: Moore-Glasberg-Schlittenlacher Method for Time Varying Sound*, Standard ISO/CD 532-3, 2017.
[4] B. C. J. Moore, B. R. Glasberg, A. Varathanathan, and J. Schlittenlacher, ''A loudness model for time-varying sounds incorporating binaural inhibition,'' *Trends Hearing*, vol. 20, Jan. 2016, Art. no. 233121651668269, doi: 10.1177/2331216516682698.
[5] B. C. J. Moore and B. R. Glasberg, ''Modeling binaural loudness,'' *J. Acoust. Soc. Amer.*, vol. 121, no. 3, pp. 1604–1612, Mar. 2007, doi: 10.1121/1.2431331.
[6] B. R. Glasberg and B. C. Moore, ''A model of loudness applicable to time-varying sounds,'' *J. Audio Eng. Soc.*, vol. 50, no. 5, pp. 331–342, 2002.
[7] A. Thwaites, B. R. Glasberg, I. Nimmo-Smith, W. D. Marslen-Wilson, and B. C. J. Moore, ''Representation of instantaneous and short-term loudness in the human cortex,'' *Frontiers Neurosci.*, vol. 10, p. 183, Apr. 2016, doi: 10.3389/fnins.2016.00183.
[8] E. M. Kaya and M. Elhilali, ''Modelling auditory attention,'' *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 372, no. 1714, Feb. 2017, Art. no. 20160101, doi: 10.1098/rstb.2016.0101.
[9] C. Kayser, C. I. Petkov, M. Lippert, and N. K. Logothetis, ''Mechanisms for allocating auditory attention: An auditory saliency map,'' *Current Biol.*, vol. 15, no. 21, pp. 1943–1947, Nov. 2005, doi: 10.1016/j.cub.2005.09.040.
[10] Y. Nakajima, S. Kuwano, and S. Namba, ''The effect of temporal patterns of sound energy on the loudness of intensity increment sounds,'' *Psychol. Res.*, vol. 45, no. 2, pp. 157–175, Oct. 1983.
[11] D. Oberfeld, ''Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both?'' *J. Acoust. Soc. Amer.*, vol. 121, no. 4, pp. 2137–2148, Apr. 2007, doi: 10.1121/1.2710433.
[12] L. Itti, C. Koch, and E. Niebur, ''A model of saliency-based visual attention for rapid scene analysis,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[13] L. Itti and C. Koch, ''A saliency-based search mechanism for overt and covert shifts of visual attention,'' *Vis. Res.*, vol. 40, nos. 10–12, pp. 1489–1506, Jun. 2000, doi: 10.1016/S0042-6989(99)00163-7.
[14] J. R. Bergen and B. Julesz, ''Parallel versus serial processing in rapid pattern discrimination,'' *Nature*, vol. 303, no. 5919, pp. 696–698, Jun. 1983.
[15] A. Treisman, ''Features and objects: The fourteenth Bartlett memorial lecture,'' *Quart. J. Exp. Psychol. A*, vol. 40, no. 2, pp. 201–237, May 1988, doi: 10.1080/02724988843000104.
[16] B. J. White, S. E. Boehnke, R. A. Marino, L. Itti, and D. P. Munoz, ''Color-related signals in the primate superior colliculus,'' *J. Neurosci.*, vol. 29, no. 39, pp. 12159–12166, Sep. 2009, doi: 10.1523/JNEUROSCI.1986-09.2009.
[17] L. M. Romanski, B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker, ''Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex,'' *Nature Neurosci.*, vol. 2, no. 12, pp. 1131–1136, Dec. 1999.
[18] J. P. Rauschecker and B. Tian, ''Mechanisms and streams for processing of 'what' and 'where' in auditory cortex,'' *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 22, pp. 11800–11806, Oct. 2000, doi: 10.1073/pnas.97.22.11800.

[19] B. Grothe, M. Pecka, and D. McAlpine, "Mechanisms of sound localization in mammals," *Physiol. Rev.*, vol. 90, no. 3, pp. 983–1012, Jul. 2010, doi: 10.1152/physrev.00026.2009.
[20] V. Pulkki and T. Hirvonen, "Functional count-comparison model for binaural decoding," *Acta Acustica United Acustica*, vol. 95, no. 5, pp. 883–900, Sep. 2009, doi: 10.3813/aaa.918220.
[21] M. Takanen, O. Santala, and V. Pulkki, "Visualization of functional count-comparison-based binaural auditory model output," *Hearing Res.*, vol. 309, pp. 147–163, Mar. 2014, doi: 10.1016/j.heares.2013.10.004.
[22] B. J. White, J. Y. Kan, R. Levy, L. Itti, and D. P. Munoz, "Superior colliculus encodes visual saliency before the primary visual cortex," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 35, pp. 9451–9456, Aug. 2017, doi: 10.1073/pnas.1701003114.
[23] B. J. White, D. J. Berg, J. Y. Kan, R. A. Marino, L. Itti, and D. P. Munoz, "Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video," *Nature Commun.*, vol. 8, pp. 1–9, Jan. 2017, doi: 10.1038/ncomms14263.
[24] J. Fecteau and D. Munoz, "Salience, relevance, and firing: A priority map for target selection," *Trends Cogn. Sci.*, vol. 10, no. 8, pp. 382–390, Aug. 2006, doi: 10.1016/j.tics.2006.06.011.
[25] B. E. Stein and T. R. Stanford, "Multisensory integration: Current issues from the perspective of the single neuron," *Nature Rev. Neurosci.*, vol. 9, no. 4, pp. 255–266, Apr. 2008, doi: 10.1038/nrn2331.
[26] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Amer.*, vol. 56, no. 6, pp. 1829–1834, Dec. 1974.
[27] B. Masterton, J. A. Jane, and I. T. Diamond, "Role of brainstem auditory structures in sound localization. I. trapezoid body, superior olive, and lateral lemniscus," *J. Neurophysiol.*, vol. 30, no. 2, pp. 341–359, Mar. 1967.
[28] D. Sanes, "An *in vitro* analysis of sound localization mechanisms in the gerbil lateral superior olive," *J. Neurosci.*, vol. 10, no. 11, pp. 3494–3506, Nov. 1990.
[29] P. L. Søndergaard and P. Majdak, "The auditory modeling toolbox," in *The Technology of Binaural Listening* (Modern Acoustics and Signal Processing). Berlin, Germany: Springer, 2013, pp. 33–56, doi: 10.1007/978-3-642-37762-4_2.
[30] S. Verhulst, T. Dau, and C. A. Shera, "Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission," *J. Acoust. Soc. Amer.*, vol. 132, no. 6, pp. 3842–3848, 2012, doi: 10.1121/1.4763989.
[31] T. Takahashi and M. Konishi, "Selectivity for interaural time difference in the owl's midbrain," *J. Neurosci.*, vol. 6, no. 12, pp. 3413–3422, Dec. 1986.
[32] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, 1993.
[33] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Amer.*, vol. 105, no. 5, pp. 2841–2853, May 1999.
[34] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Commun.*, vol. 53, no. 5, pp. 592–605, 2011, doi: 10.1016/j.specom.2010.05.006.
[35] J. A. M. Van Gisbergen, A. J. Van Opstal, and A. A. M. Tax, "Collicular ensemble coding of saccades based on vector summation," *Neuroscience*, vol. 21, no. 2, pp. 541–555, May 1987.
[36] D. R. Perrott and K. Saberi, "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *J. Acoust. Soc. Amer.*, vol. 87, no. 4, pp. 1728–1731, Apr. 1990.
[37] B. J. Fischer and J. L. Peña, "Owl's behavior and neural representation predicted by Bayesian inference," *Nature Neurosci.*, vol. 14, no. 8, pp. 1061–1066, Aug. 2011, doi: 10.1038/nn.2872.
[38] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vis. Res.*, vol. 49, no. 10, pp. 1295–1306, Jun. 2009, doi: 10.1016/j.visres.2008.09.007.
[39] B. A. Edmonds and J. F. Culling, "Interaural correlation and the binaural summation of loudness," *J. Acoust. Soc. Amer.*, vol. 125, no. 6, pp. 3865–3870, Jun. 2009, doi: 10.1121/1.3120412.
[40] V. P. Sivonen, "Directional loudness perception: The effect of sound incidence angle on loudness and the underlying binaural summation," Aalborg, Denmark: Afdeling for Akustik, Aalborg Universitet, 2006, pp. 84–99.
[41] V. P. Sivonen, "Directional loudness and binaural summation for wideband and reverberant sounds," *J. Acoust. Soc. Amer.*, vol. 121, no. 5, pp. 2852–2861, May 2007, doi: 10.1121/1.2717497.
[42] B. Scharf, "Loudness summation between tones from two loudspeakers," *J. Acoust. Soc. Amer.*, vol. 56, no. 2, pp. 589–593, Aug. 1974, doi: 10.1121/1.1903295.
[43] S. Sato, T. Kitamura, H. Sakai, and Y. Ando, "The loudness of 'complex noise' in relation to the factors extracted from the auto-correlation function," *J. Sound Vib.*, vol. 241, no. 1, pp. 97–103, 2001, doi: 10.1006/jsvi.2000.3281.
[44] *Acoustics—Normal Equal-Loudness-Level Contours*, Standard ISO 226:2003, 2003.
[45] M. Kleiner, D. Brainard, and D. Pelli, "What's new in Psychtoolbox-3?" *Perception*, vol. 36, no. 14, pp. 1–16, 2007.
[46] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, nos. 3–4, pp. 324–345, 1952.
[47] M. D. Lee and E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course*. Cambridge Univ. Press, 2014.
[48] *Acoustics—Method for Calculating Loudness Level*, Standard ISO 532B, 1975.
[49] E. Zwicker and B. Scharf, "A model of loudness summation," *Psychol. Rev.*, vol. 72, no. 1, p. 3, 1965.
[50] A. Saremi, R. Beutelmann, M. Dietz, G. Ashida, J. Kretzberg, and S. Verhulst, "A comparative study of seven human cochlear filter models," *J. Acoust. Soc. Amer.*, vol. 140, no. 3, pp. 1618–1634, 2016, doi: 10.1121/1.4960486.
[51] J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. Santos-Victor, and R. Pfeifer, "Multimodal saliency-based bottom-up attention a framework for the humanoid robot iCub," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2008, pp. 962–967.
[52] M. Mosadeghzad, F. Rea, M. S. Tata, L. Brayda, and G. Sandini, "Saliency based sensor fusion of broadband sound localizer for humanoids," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Sep. 2015, pp. 362–367.
[53] D. L. Sparks, "The brainstem control of saccadic eye movements," *Nature Rev. Neurosci.*, vol. 3, no. 12, pp. 952–964, Dec. 2002.
[54] R. K. Maddox, D. A. Pospisil, G. C. Stecker, and A. K. C. Lee, "Directing eye gaze enhances auditory spatial cue discrimination," *Current Biol.*, vol. 24, no. 7, pp. 748–752, Mar. 2014, doi: 10.1016/j.cub.2014.02.021.
[55] G. Sheng, *Vehicle Noise, Vibration, and Sound Quality*. Warrendale, PA, USA: SAE, 2012, pp. 347–362.

**YUKI NAKATANI** received the M.E. degree from Osaka University, Japan, in 2012. Since 2012, he has been working with Mazda Motor Corporation, Hiroshima, Japan. His current research interest includes signal processing of automotive sounds.



**MASAYUKI WATANABE** received the Ph.D. degree from the Graduate University for Advanced Studies, Japan, in 2004. He currently works with Mazda Motor Corporation, Hiroshima, Japan. His current research interest includes applied neuroscience in automotive industries.

**NAOKO YOROZU** received the M.E. degree from Hiroshima University, Japan, in 2003. Since 2003, she has been working with Mazda Motor Corporation, Hiroshima, Japan. Her current research interest includes automotive acoustic engineering.